Reprinted from

SANKHYĀ: THE INDIAN JOURNAL OF STATISTICS

VOL. 1. PART 1. JUNE, 1933

EDITED BY P. C. MAHALANOBIS

STUDIES IN EDUCATIONAL TESTS. No. 1. THE RELIABILITY OF A GROUP TEST OF INTELLIGENCE IN BENGALI

BY

PROF. P. C. MAHALANOBIS.

ART PRESS
20, BRITISH INDIAN
STREET, CALCUTTA

SANKHYĀ: THE INDIAN JOURNAL OF STATISTICS.

Vol. 1. Part 1. June, 1933.

Sankhyā is published with the help of the Indian Statistical Institute and in consultation with the following Journal Committee appointed by the Institute.

Each volume of *Sankhyā* will contain about 400 pages, with plates and tables, and will be issued in approximately four equal parts.

The Journal is supplied free to each member of the Indian Statistical Institute. The subscription price including packing and postage for non-members, payable in advance, is Rs. 20/- per volume ; single issues Rs. 6/- including postage. Rates for foreign countries to be calculated on the basis of 32 shillings for each volume, and 9 shillings for each number. Subscriptions should be sent to the Art Press, 20, British Indian Street, Calcutta.

Papers for publication and books and reprints for review should be sent to the Editor, Prof. P. C. Mahalanobis, Presidency College, Calcutta.

It is a condition of publication in *Sankhyā* that the paper shall not already have been published elsewhere, and will not be reprinted without leave of the Editor. Manuscripts should be written on one side of the paper, and should be preferably typewritten. It is necessary that the raw data or measurements on which the paper is based, and detailed information of their sources, should be furnished in every case. It is desirable that the distributions or correlation charts from which statistical constants are calculated should accompany the paper. Tables should be numbered consecutively with Arabic figures without brackets (Table 1, Table 45, etc.), and should have short explanatory headings. Diagrams and drawings should also be numbered consecutively with Arabic figures, and should be sent in a finished state suitable for direct reproduction with all lettering in pencil. References to tables and diagrams should be given within brackets in the text. The list of literature cited should be arranged alphabetically according to authors' names. In the case of Journals the volume number, year and page references, and in the case of printed books the name of the author and the publisher, the date of publication, and the number of the edition or impression should always be stated clearly. Abbreviations, local names, vernacular measures, etc., should be written very legibly (preferably in block letters) and should be explained where they first occur in the text. A detailed Note for Contributors has been prepared and may be obtained on application to the Editor.

Contributors receive 25 copies of their papers free. The cost of additional reprints will depend upon the number required ; for 50 additional copies the rate is Rs. 6/- for each sheet of eight pages or Rs. 2/- per page for part of a sheet, with an extra charge for plates printed on special paper. Additional copies should be ordered at the time of returning the final proof.

STUDIES IN EDUCATIONAL TESTS. No. 1. THE RELIABILITY OF A GROUP-TEST OF INTELLIGENCE IN BENGALI.*

By P. C. MAHALANOBIS.

General Introduction.

The pioneer period in intelligence testing may be considered to have ended with the publication of the final revision of the Binet scale in 1911 shortly before his death. The work was taken up by a number of competent psychologists like Goddard, Terman, Thorndike and others in America, by Stern, Bobertag, Meumann and others in Germany, and by De Sanctis (on independent lines), Saffiotti and his followers in Italy. While in England, Spearman had already started his work on the two-factors theory of general intelligence which was later developed with great vigour by his followers.

The advent of group testing gave a new turn to the movement. Surveys on a scale were made possible which could not have been dreamt of in the era of individual testing. A great impetus was given by the use of group tests in the American army during the War in 1917-18. The Army Alpha test, for example, was administered to more than one million unselected recruits. The next few years which saw an amazing development in group testing may be characterized as the boom period in mental testing.[1]

The subject of vocational guidance and industrial psychology gained peculiar importance in the post-war period of economic distress and gave additional impetus to the testing movement. It received official recognition in various countries of the world. In India, for example, it was discussed by the Central Advisory Board of Education in October, 1921, and experiments with the Binet-Simon scale and its later revisions were started in certain provinces. A comprehensive report on "Psychological Tests of Mental Abilities" was prepared by Dr. A. S. Woodburne and was published by the University of Madras in 1924. The author quoted with approval at the end of his report the following sentence from Terman:—"Individual psychology has achieved its greatest success in the field of intelligence testing, and the development of the last two decades in this line constitutes the most notable event in the history of modern psychology."

In Europe and America, however, the over-production of uncritical researches had begun to produce a reaction, and a kind of depression in mental testing had set in. Although very little positive or constructive work had been done in India, the negative phase appears to have strongly influenced educational opinion in this country. Very recently an important educational officer in Bengal (who had never handled a test himself) confidently remarked that the whole thing was useless, and it was not worth while wasting money on such experiments.

Such periods of depression are useful in directing attention to basic problems. The technique of mental testing may be broadly divided into three parts: (1) the selection of

*Read before the Psychological Section of the Indian Science Congress, Nagpur, January, 1931.

[1] T. R. Kelly, Professor of Education and Psychology at the Stanford University (8, Preface) wrote in 1927: "The claims put forward for standardized intelligence and educational tests extend from the cradle to the grave. They have been mentioned seriously in connexion with the selection of children for adoption, and in choosing life partners. They have been charged with undermining democracy, and have been hailed as of the greatest aid in solving the complex social problems of the present times."

The figure within brackets refers to the Bibliographical List given at the end of the paper.

suitable test material, (2) the administration of the test, and (3) the analysis and interpretation of the results.

The last part is essentially statistical, and I believe a good deal of the conflicting results reached in certain fields may be definitely traced to the use of faulty or inadequate methods of analysis. Here the co-operation of a student of statistics may prove useful. My chief purpose in these studies has been to examine methods and results independently from the point of view of the theory of statistics. I have endeavoured to use recent developments in statistical technique, and have taken care to explain them in some detail in the form of Statistical Notes.[2]

I may perhaps be permitted to state in anticipation that my experiments have convinced me of the value of mental tests when they are carefully applied, and when the results are analysed with the help of adequate statistical methods.[3]

I should like to emphasize the fact that my work is essentially statistical and empirical. I have no desire to enter into a psychological discussion regarding the meaning of intelligence.[4] I accept it as an established fact that in other parts of the world educational psychologists have succeeded in measuring what they usually refer to as "general intelligence" by the use of tests of a certain type. I have attempted to construct and administer similar tests through the medium of the Bengali language, and interpret the results from a statistical point of view. For my purpose it is sufficient to assume that there exists an entity (simple, or itself an aggregate of other elements) which different subjects may possess in different degrees, and which may possibly be correlated with success in school

---

[2] Such notes are superfluous in a Statistical Journal, but I have retained them in the hope that they may prove useful to psychologists and other workers in the field of intelligence testing in India.

[3] T. R. Kelley in discussing common sources of error in the use of statistical methods says: (8, Preface) "It is my thesis that these instruments are potent for good if intelligently used by honest, capable, and socially minded counsellors and it is the purpose of this book to offer certain guides in the interpretation of test scores and to make explicit the errors involved—all with a view to a more penetrating use of such measures."

In this connexion the following extracts from the presidential address of Dr. C. W. Kimmins (9) to the Educational Section of the British Association in 1929 may also prove interesting:—

"The most important movement in education since 1905, the date of the last meeting of the British Association in South Africa, is the coming of the intelligence test and its incorporation as an essential element in the general scheme of education. * * * * * We are as yet very far from having reached the ideal form of intelligence test, but sufficient has already been done to show by actual experience in a variety of ways, the remarkable value of individual and group tests."

"In the Begabten Schulen in Germany, where, in the final selection for admission, the results are based almost entirely on the intelligence quotients of the candidates, one cannot fail to be greatly impressed by the ease with which these children, without any undue pressure can successfully cover as much ground in one year as normal children would require at least two years to accomplish. In the days to come we shall give far more attention to the super-normal child than we do at present."

"A very promising direction in which intelligence tests may render invaluable assistance is to be found in vocational guidance. Indeed it is not unreasonable to hope that in days to come, every boy (and girl) on leaving school will have reliable information as to the kind of work in which he can most effectively use the ability he possesses, with pleasure and satisfaction to himself and to his employer. In this case the hopeless situation involved in 'the square peg in the round hole' will tend to disappear."

E. R. Hamilton says (5, p. 23):—"Nothing in the history of psychology is more remarkable than the success with which 'mental tests' have been employed by some psychologists; nothing more lamentable than the triviality to which others have been led by the same methods. The far-reaching significance of the more judicious application of mental tests, and the dangerous popularity of the more futile, alike call for a review of the assumptions that undertake their use."

[4] I should also confess that neither am I competent to do so. My knowledge of 'mental tests' is confined to a few standard treatises, and my acquaintance with the literature of the subject is limited in range.

examinations or with other pursuits in life. As a working hypothesis I must, of course, also assume that scores in the present test are at least partially determined by this entity. Thus (without committing myself to any particular theory of intelligence) I may speak of one student being more intelligent than another when the former obtains a higher score than the latter. The same understanding will also extend to the use of the word 'intelligence' in such phrases as 'variation of intelligence' among students of a certain age, 'difference in intelligence' between students of different ages, etc.

In order to avoid delay I am publishing my results as a series of studies instead of a consolidated report. In this first paper I have discussed the statistical reliability of the measurements obtained by the use of the present test as based on the records of 1212 school pupils.

In other papers I intend to discuss such topics as the growth of intelligence with age, the change of variability of intelligence with age, and the correlation between test-scores and school-marks (in annual and half-yearly examinations). Finally I shall give a detailed comparative study of different statistical measures of intelligence.

DESCRIPTION OF THE TEST.

Form of the Test. So far as individual tests are concerned, a certain amount of work had been already done in Bengal with a modified version of the Binet-Simon test.[5] In spite of the great advantages of an individual test, difficulties of standardization make such a test entirely unsuitable for a rapid and wide survey. The choice of a group test was therefore inevitable.

On enquiry I found that very little had been done in this direction in Bengal. In 1926-27 my esteemed friend Mr. Jitendramohan Sen, M.Ed. (Leeds), of the Bengal Education Department, had devised a non-verbal (pictorial) group test suitable for young children and had administered it in different schools in Bengal. It is extremely unfortunate that owing to the pressure of his routine duties he has not been able to analyse or publish the results of his experiment. He, however, very kindly showed me a portion of his records, discussed some of the difficulties he had encountered in his work, and helped me with valuable suggestions. Mr. Sen's form could be administered to a number of pupils at the same time, and all the entries were made by the subjects themselves. But it required the services of a person who could give specific directions in accordance with a fixed time-schedule. This introduced the question of standardization of the procedure, including strict adherence to a time-table scheduled in seconds. I felt that under existing conditions in Bengal, such a degree of standardization in the administration of a test would be difficult to achieve, and I decided to adopt a procedure which would be practically self-administering, and would thus dispense with the services of an "administrator" of the test.

I adopted a linguistic form, and decided to use the Bengali language as my medium. It may therefore be justifiably claimed for the present experiment that it represents the first attempt at conducting a group test of intelligence through the Bengali language. The use of the vernacular language avoids all the complications unavoidably associated with the use of English (or any other foreign) language as the medium.

Preliminary Experiments. Further, as I wanted to avoid difficulties connected with the exact timing of the replies, I decided to allow sufficient time to enable all the questions being answered comfortably. In November, 1927, I prepared a preliminary form, and

[5] Chiefly at the Dacca Training College under the direction of Dr. M. West, but the results have not been published so far.

administered it to a group of about 200 students of the Brahmo Boys' School of Calcutta. The subjects were asked to finish the exercise as quickly as possible and were told that credit would be given for quickness; and the supervisors were requested to note on each book the exact time when it was given up. This preliminary experiment showed that the average time required was about 33 (thirty-three) minutes. I decided that an allotment of one hour for the test would be amply sufficient. Making an allowance of full 10 minutes for the students to come in and take their seats and other preliminaries, this leaves a period of at least 50 minutes to finish the exercise. I am confident therefore that questions of speed do not materially affect the results. In other words, the present test is a test of power, and not of speed.

The preliminary experiment was also useful in revealing ambiguities in the framing of the questions and other defects. These were remedied and the final form, designated *Manan Parīkshā Ka (Intelligence Test A)** was printed in February, 1928.

Description of the Test. The test material consists of two Parts I and II, each comprising 6 questions which were of the usual type found in many English and American forms. The questions in Part II were of the same type as the corresponding questions in Part I, but were more difficult in nature. Such a duplication of questions was arranged intentionally in order to study the reliability of the test.

I am indicating below the nature of the questions by giving a few examples in English equivalents.

(1) Underline one word which is out of place in the following series:—
red, white, round, black, blue.

(2) Fill up the blank in the number series:—
2 4 6 8 10 12 (. . . . .).

(3) A genealogical table was given with a descriptive list. Questions were asked on the relationships subsisting between different persons shown in the genealogical table.

(4) Arrange the following in their proper serial order, and underline the middle term:—
Day, year, hour, month, week.

(5) Underline the figure which does not fit in the series:—
6, 2, 8, 7, 4, 10.

(6) A number of phrases are given in two languages with corresponding equivalents underlined. It is required to pick up other equivalent words by comparison.

The questions were printed in large type on both sides of foolscap-size paper, and were pasted together to form a booklet of 6 pages. The upper half of the first page contained a blank form which the subject was asked to fill up, and included blank spaces for entering the (1) name, (2) religion, (3) caste, (4) date of birth, (year, month, day), (5) place of birth (village, district), (6) the date of the examination (day, month, and year, in the English system), (7) the name of the school, (8) location of the school, (9) class, and (10) subjects of study. In the lower half of the first page definite instructions regarding the procedure to be followed were printed in bold types. A free translation of these instructions is given below:—

Directions.

Inside the book you will find a few simple questions. Answer these questions correctly and as quickly as possible. Begin from the first question and try one after the other in serial order. If you find any particular question to be too difficult pass on to the next question; do not waste your time. You will get one full class-period, try to answer all the questions within this time. Do not give up the attempt until time is up; do not give up your book before time.

* মনন পরীক্ষা (ক)

At the bottom of the page was printed in large heavy types in Bengali: *"Do not ask any question."*

The second page started with the legend in Bengali:—*"Do not write anything on this page."* It contained a blank for entering test scores, teacher's ratings, etc. The actual questions started at the top of page 3, and continued to the bottom of page 6.

Practice Form. In the preliminay experiments of 1927, I had also used a practice form of one single sheet which the subjects were required to complete within 15 minutes. These practice forms were given usually 1 or 2 days before the date of the examination and served to make the form of the examination familiar to the students, and probably helped in eliminating fluctuations due to the shock of mere novelty. I however found it impracticable to use such practice forms in the later experiment.

Procedure of administering the test. Before administering the test I had a personal talk with the head of the institution, and usually arranged an informal conference with the teachers of the schools who invariably acted as Supervisors during the actual administration of the test. I explained generally the purpose of the intelligence test, tried to secure the willing co-operation of the teachers, and impressed upon them the importance of following the standardized directions very carefully.

Before the actual date of the examination I sent one of my assistants (usually Babu Sudhir Kumar Banerji) to the school concerned for compiling the names, age, class, etc., of the subjects from the school register. From these records a roll-form was prepared in which the actual age of the students (calculated to the nearest month of the date of the examination) was entered. On the day of the examination two of my assistants went to the school concerned some considerable time before the hour of the examination, and marked in chalk the roll number of each student on the exact spot on the benches where he was to be seated. Usually 3 and 2 seats were allotted to benches in alternate rows. As the ordinary school bench is about 10 ft. long, this left a gap of from 3 ft. to 5 ft. between the subjects. This arrangement served to prevent "copying" to a great extent. As the students came in, they were asked by the Supervisor to take their seats according to their roll number. As the number of pupils in a single class (or section of a class) never exceeded 40, it usually took only 2 or 3 minutes to get the children properly seated. The Supervisor then distributed the test-books, and requested the students to fill up the blank form on the first page and to read and follow the instructions carefully.

The Supervisors were requested not to answer questions, but only to point out that all necessary instructions would be found in the book itself. They were asked to impress upon the students that it was not necessary to give up the paper before the hour was over. They were also requested to be very careful that no voluntary or involuntary copying took place, and were strictly forbidden to explain any thing, even what they considered to be non-essential or trivial points. I myself or my assistants remained in the school premises throughout the examination hour, and went from one room to another to see that everything was going on smoothly.

Whenever practicable, all the students of a given school were examined simultaneously. In certain cases owing to want of room, the examination had to be conducted in two shifts. In such cases care was, however, taken to ensure that the students of the second batch had no opportunity of talking with the students of the first batch.

The preliminary experiment had shown that most of the students, even in the upper classes, were unable to fill up correctly all the entries in the blank form on the first page. I therefore instructed my assistants to collect all the necessary information at the time of compiling the age register. In a few cases the caste of the students could not be ascertained.

Schools examined. I am grateful to the authorities of the following institutions for allowing me facilities for examining their students.

(1) Santiniketan School (5th August, 1928).
(2) St. Paul's School, Calcutta (26th September, 1928).
(3) City School, Calcutta (29th September, 1928).
(4) Brahmo Girls' School, Calcutta (5th October, 1928).
(5) Mitra Institution, Bhowanipur (6th October, 1928).
(6) Brahmo Boys' School, Calcutta (12th October, 1928).

Scoring the answer papers. A standard list of the correct replies was prepared, and one mark was allotted for each correct answer. Fractions were ruled out, *i.e.*, a reply was either right or wrong. The maximum number of marks for Part I was 23, and for Part II was 37, or 60 for the two parts combined. The totals of marks in each part were added separately. All the scoring and tabulation was done by my assistants under my direct supervision.

The laborious arithmetical computations were carried out by Messrs. Sudhir Kumar Banerji and Jaladhar Sarma, Computers, Statistical Laboratory, Presidency College, Calcutta.

Reliability of the Test.

From the statistical point of view the first problem to be studied is how far the results are stable for the same individual. Suppose we examine one group of students in any particular subject to-day, and suppose we re-examine the same group of students in the same subject sometime later. If we arrange the names of the students in order of merit on the results of the first examination as also on the results of the second, the two lists will not of course be exactly identical. For it is known that great variations sometimes occur in the performance of the same individual and such variations would naturally introduce fluctuations in the test scores. The test material is again limited in scope, and both of them may not be equally suited to the ability of a child. But we certainly expect that there should be a considerable amount of agreement between the two sets of results. The amount of such correspondence may be adopted as a criterion for judging the reliability of the two examinations.

I wish to make it clear that in this part of our investigations we are not in any way concerned with the nature of the entities the test is actually measuring. Whatever might be the thing measured, we want to know how reliable the test is in actually making this measurement.

The Coefficient of Correlation. It will be remembered that the present test consists of two similar parts. The coefficient of correlation between the scores (of the same individual) in the two parts will therefore furnish a measure of the reliability of the test as a whole. The correlation chart for 1212 students is given in Table A.

The actual correspondence of marks in Part I (called the X-variate) and in Part II (called the Y-variate) is shown in Table 1 (p. 34) and Table 2 (p. 39).

In Table 1 the entry is by the score obtained in Part I (*i.e*, by the X-variate), and column 1 gives such scores from 0 to 23. Column 2 gives, n, the total number of students who obtained the corresponding X-score given in column 1. Thus 0 mark was received by 37 students, 1 mark by 31 students, 2 marks by 20 students and so on. Column 3 gives the average score in Part II (*i.e.*, the Y-variate) obtained by all students with the corresponding X-score given in column 1. Thus the 37 students who obtained 0 in Part I

Table A. Correlation Between Marks in Part I and Part II

Marks in Part I

Marks in Part II	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Totals
37																								1	1
36																								5	5
35																		1	—	—	—	1	6	14	22
34																		—	1	—	1	3	7	14	26
33																		—	—	—	4	2	14	15	35
32																		—	—	2	1	7	16	13	39
31																		2	—	3	5	8	10	12	40
30																	1	2	2	2	7	11	18	11	54
29										1								2	2	3	7	8	18	10	46
28										1						—	1	4	3	4	11	14	6	13	57
27										—			1	—	2	1	1	2	5	3	7	9	8	4	43
26										—			1	—	1	2	3	3	1	7	6	12	6	5	47
25										—		2	—	—	2	1	2	6	3	3	1	6	6	—	32
24										—	1	1	—	2	—	—	1	1	3	3	7	7	5	1	32
23										—	1	—	—	—	2	1	2	—	4	2	3	2	6	2	25
22										—	—	2	2	1	2	4	1	3	5	3	3	4	3	—	33
21										—	—	—	1	—	2	1	3	2	9	5	4	1	2	—	30
20										—	—	—	1	3	—	3	1	2	3	8	2	3	3	—	31
19									1	—	1	1	1	3	2	3	3	4	4	3	2	1	—	—	29
18						1	1	—	—	—	—	1	2	4	6	1	2	3	3	3	6	2	1	—	36
17						—	1	—	1	—	3	—	—	2	3	4	1	3	1	3	1	2	—	1	26
16						—	1	—	—	1	—	5	1	6	4	1	3	3	1	3	1	1	1	—	32
15						—	—	2	—	3	2	1	—	1	1	1	1	1	3	4	2	3	—	1	26
14					1	—	—	—	2	2	2	2	1	6	4	1	3	4	—	2	1	2	—	—	33
13			1	—	1	—	—	1	—	4	2	4	1	2	2	4	2	3	2	—	1	1	—	—	31
12			—	—	—	—	3	1	2	1	3	2	5	1	3	1	2	1	2	1	—	2	1		31
11			—	1	1	—	—	2	1	1	4	2	4	2	—	5	3	2	1						29
10			—	1	—	—	—	2	—	8	3	2	6	4	2	3	—	1	—	—					32
9			—	1	—	—	6	3	4	4	4	4	3	1	5	—	—	1	—	2					38
8		1	1	2	2	1	4	2	4	3	4	2	2	2	1	—	1	1							33
7		—	2	2	1	1	3	3	1	1	2	2	1	—	2	—	—	—							21
6	3	2	3	4	3	3	1	3	4	—	2	4	2	2		—	—	—							36
5	1	—	2	4	4	3	3	4	3	4	2	1	2	—	—	—	—	—							33
4	4	2	2	6	2	2	2	3	2	1	—	—	—	—	—	—	—	1							27
3	5	6	4	4	5	2	1	2	—	2	1	1	—	1	—	—									34
2	4	5	1	3	5	2	1	1	3	—	—	—	1	1	—	—									27
1	3	6	2	4	1	3	1	—	—	—	—	—	—	—	1	1									22
0	17	9	2	4	2	3	—	—	—	1															38
Totals	37	31	20	36	28	21	28	29	28	38	37	39	38	44	47	38	37	58	60	69	83	112	132	122	1212

obtained on an average 1·79 marks in Part II. It will be noticed that Y steadily increases with X, showing that those who obtained high marks in Part I also obtained, on an average, high marks in Part II.

In Table 2 the entry is by scores in Part II (the Y-variate). Here also it will be noticed that X increases with Y, showing that those who did well in Part II had also done well in Part I.

We can obtain a numerical measure of the agreement between the scores in Part I and in Part II by calculating the coefficient of correlation between the two sets of marks.

By direct calculation[6] we find that the coefficient of correlation is $+0{\cdot}8781 \pm {\cdot}0024$. This result roughly indicates that the amount of correspondence or agreement between the two tests is about 88%.

We may show the dependence of marks in Part II (Y) on marks in Part I (X) in the following way:—

$$Y = 1{\cdot}317\,(X) - 0{\cdot}93 \qquad \text{...(1)}$$

This equation gives Y, the average value of scores in Part II of those students who all obtained the same score, X, in Part I. Equation (1) shows that the gain of a single mark in Part I of the test is likely to lead to an average gain of 1.317 marks in Part II.

The above equation will, however, be completely satisfactory only if the relation between Y and X is strictly rectilinear. A glance at Figure 1 (p. 35) shows that this is not quite true.

We may investigate the question in the following way. Consider the 1212 different values of Y (*i.e.*, marks in Part II). Let us classify the individual marks in accordance with the scores in Part I. That is, let us collect together the Part II-scores of all pupils who have obtained zero mark in Part I, and call this group, class 0. Similarly we collect in class 1, class 2, class 3, ... class 23, the Part II-scores of all pupils who have obtained 1, 2, 3, ...23 marks respectively in Part I. In case scores in Part I and Part II are absolutely unrelated, any variation of scores in Part I will not affect the scores in Part II. Thus the mean values of Part II-scores in class 0, in class 1, in class 2, ... in class 23 will be practically the same and equal to the General Mean 18.49 approximately. Further, since the above classification has no effect on Part II-scores, the average variation of Part II-scores within each class will also be practically the same and equal to the average variation of all Part II-scores irrespective of classes. On the other hand, in case scores in Part II were completely determined by scores in Part I, those students who obtained the same score, say *"a,"* in Part I, will automatically secure equal scores in Part II, say *"b"* (which need not necessarily be equal to *"a"*, the score in Part I).[7] Hence the sum of squares of deviations within each class will be zero. Also since scores in Part I and Part II acre completely related, any change in the Part I-score will lead to a change in the Part II-score, and the class-means must be, in general, all different.

It can be shown by simple algebraic reasoning[8] that the total sum of squares of deviations from the General Mean (let us call this sum, T,) is equal to the sum of squares of

[6] See Statistical Note 1.

[7] The score in Part II is in general not equal to the score in Part I; but students securing the same score in one part must secure equal scores in the other part.

[8] See Statistical Note 3.

deviations within classes (say W) plus the weighted[•] sum of squares of deviations between class-means (say B). Or symbolically:—

$$T = W + B \qquad (2)$$

When the two variates are absolutely unrelated B=0, and when they are completely related W=0.*

The above relation leads almost inevitably to Karl Pearson's "Correlation Ratio"[10] which is defined by the following equation:—

$$\eta^2 = \frac{B}{T} = \frac{\text{Weighted sum of squares of deviations of Class Means from General Mean}}{\text{Total sum of squares of deviations of individual scores from General Mean}}$$

For two unrelated variates, B=0 and $\eta^2=0$; while for two completely related variates W=0, and B=T, so that $\eta^2=1$. The value of η^2 (and hence of η, the "Correlation Ratio") thus varies from zero to unity, and the numerical magnitude gives a convenient measure of the degree of relation between X and Y.

In this particular case the weighted sum of squares of deviations of mean values in the 24 different classes is 104,482·59, while the total sum of squares of deviations is 131,230·84, the ratio of these two numbers gives the value of the square of the "Correlation Ratio" or $\eta^2=0·796174$, with $\eta=0·8923$.

Now Pearson has shown that for a strictly linear relation between Y and X, the quantity (η^2-r^2) must be zero. In the present instance $r^2=0·771060$, so that $(\eta^2-r^2)=$ 0·025115, which is small, showing a slight departure from linearity.[11]

Let us try to appreciate the significance of this departure from linearity. We can easily calculate what the sum of squares of deviations of mean values would be in case the mean values progressed strictly according to a linear relationship.[12] In the present example it would be 101,187·64. The observed sum of squares of deviations of mean values is 104,482·59. The difference 3,294·95 represents the total deviations from linearity from the 24 different class-means. These 24 class-means, however, allow 23 independent comparisons, of which one comparison is given by the linear relationship. Thus we are left with 22 independent contributions.[13]. Dividing 3,294·95 by 22, we obtain 149·7705, the mean square deviation, or as it is usually called in statistical language, the variance,[14] due to deviations from linearity. Let us now consider the standard deviation for fluctuations within class groups. We have already seen that the total sum of squares of deviations for 1212 individual marks is 131,230·84, of which 104,482·59 is due to deviations between mean values (23 comparisons). The difference 26,748·25 may then be ascribed to 1188 independent comparisons[15] within classes. Dividing 26,748·25 by 1188 we obtain 22·5154. This represents the purely individual fluctuations[16] of scores in Part II among students who had obtained the same number of marks in Part I.

• That is, the sum of squares of deviations of class means multiplied by the corresponding number of individuals in the class.

*Ignoring, of course, the errors of sampling.

10 See Statistical Note 4.

11 See Statistical Note 5.

12 See Statistical Note 6.

13 See Statistical Note 7.

14 Extracting the square root of 149.7705 we get 12.24 as the corresponding standard deviation.

15 1211 independent comparisons less 23 comparisons between class-means gives 1188.

16 The square root of 22.5154 is 4.745, the standard deviation within classes.

We may judge the significance of 149·7705 (variance due to deviations from linearity) by comparing it with the residual value 22·5254. R. A. Fisher[17] has shown how this can be done with mathematical precision. Carrying out the calculations we find that the variance due to deviations from linearity is significantly greater than the purely random variance of individuals within classes (the probability being much greater than 100 to 1).

The Regression of Part II-Scores (Y) on Part I-Scores (X).

The relation between Y and X may therefore be considered definitely non-linear in character, and we may proceed to use a second degree or parabolic curve. The mathematical process of fitting such curves were fully described by Karl Pearson[18] in 1905. Working out the details of the calculation we obtain.

$$Y = 3{\cdot}12 + 0{\cdot}3268\,(X) + 0{\cdot}0386\,(X^2) \qquad (3)$$

The graduated values of Y, scores in Part II, for each class of marks in Part I, are shown in column 3 of Table 1, and the corresponding observed values in column 4. The

Table 1. Regression of Y (Part II Scores) on X (Part I Scores).

X_p	n_p	Values of Y_p		Difference	X_p	n_p	Values of X_p		Difference
		Graduated	Observed				Graduated	Observed	
0	37	3·12	1·79	+1·33	13	44	13·90	14·32	−0·42
1	31	3·49	2·16	+1·33	14	47	15·26	15·87	−0·61
2	20	3·93	4·05	−0·12	15	88	16·71	16·69	+0·02
3	36	4·45	4·17	+0·28	16	87	18·23	18·41	−0·18
4	28	5·05	4·75	+0·30	17	58	19·83	20·66	−0·83
5	21	5·73	4·29	+1·44	18	60	21·51	21·57	−0·06
6	28	6·48	8·18	- 1·70	19	69	23·26	22·32	+0·94
7	29	7·30	7·55	−0·25	20	83	25·09	25·28	−0·19
8	28	8·21	8·11	+0·10	21	112	27·00	26·22	+0·78
9	38	9·20	10·16	−0·96	22	132	28·98	28·88	+0·10
10	37	10·26	11·30	−1·04	23	122	31·04	31·10	−0·06
11	39	11·39	12·95	−1·56	Total	1212			
12	38	12·61	12·47	+0·14					

differences between observed and graduated (or expected) values are given in column 5. Squaring these differences and multiplying by the corresponding value of n_p, the number of individuals in each class, we obtain directly the sum of squares of deviations from the parabolic curve. The actual value is 616·36. Remembering that a parabolic curve introduces 2 constants which correspond to 2 independent comparisons, we notice that the

[17] See Statistical Note 8.
[18] See Statistical Note 9.

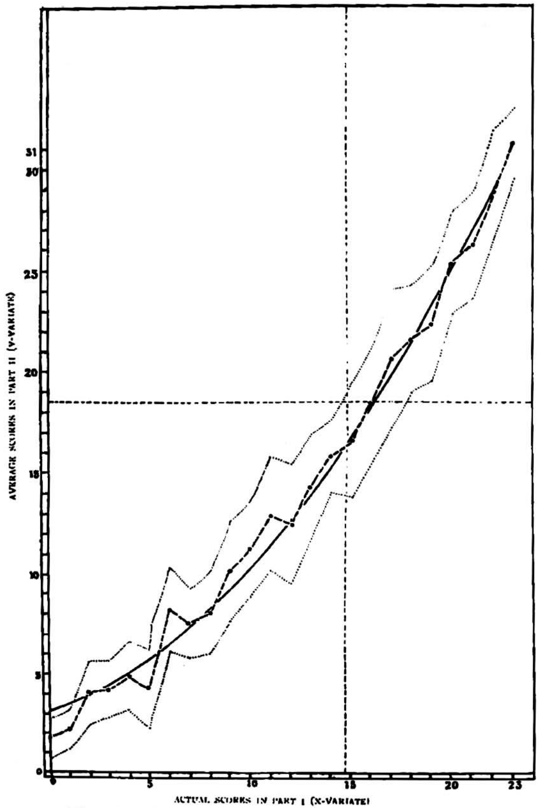

Fig. 1. Regression of Part II-Scores (Y-variate) on Part I-Scores (X-variate).

Average values of scores in Part II (Y-variate) corresponding to actual scores in Part II (X-variate). Observed values shown in broken lines, and graduated values in the continuous line. The dotted line shows the magnitude of the standard deviation (σ) within X-classes (½ σ on either side of the observed average value).

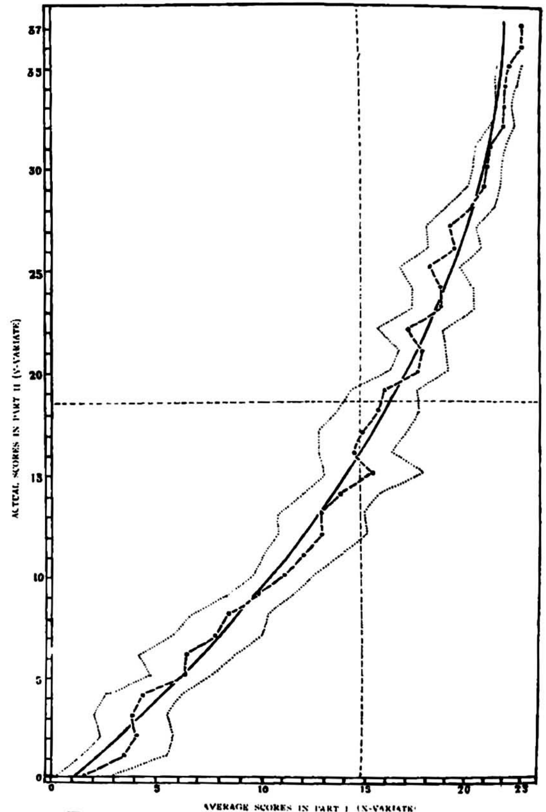

Fig. 2. Regression of Part I-Scores (X-variate) on Part II-Scores (Y-variate).

Average values of scores in Part I (X-variate) corresponding to actual scores in Part II (Y-variate). Observed values shown in broken lines, and graduated values in the continuous line. The dotted line shows the magnitude of the standard deviation (σ) within Y-classes (½ σ on either side of the observed average value).

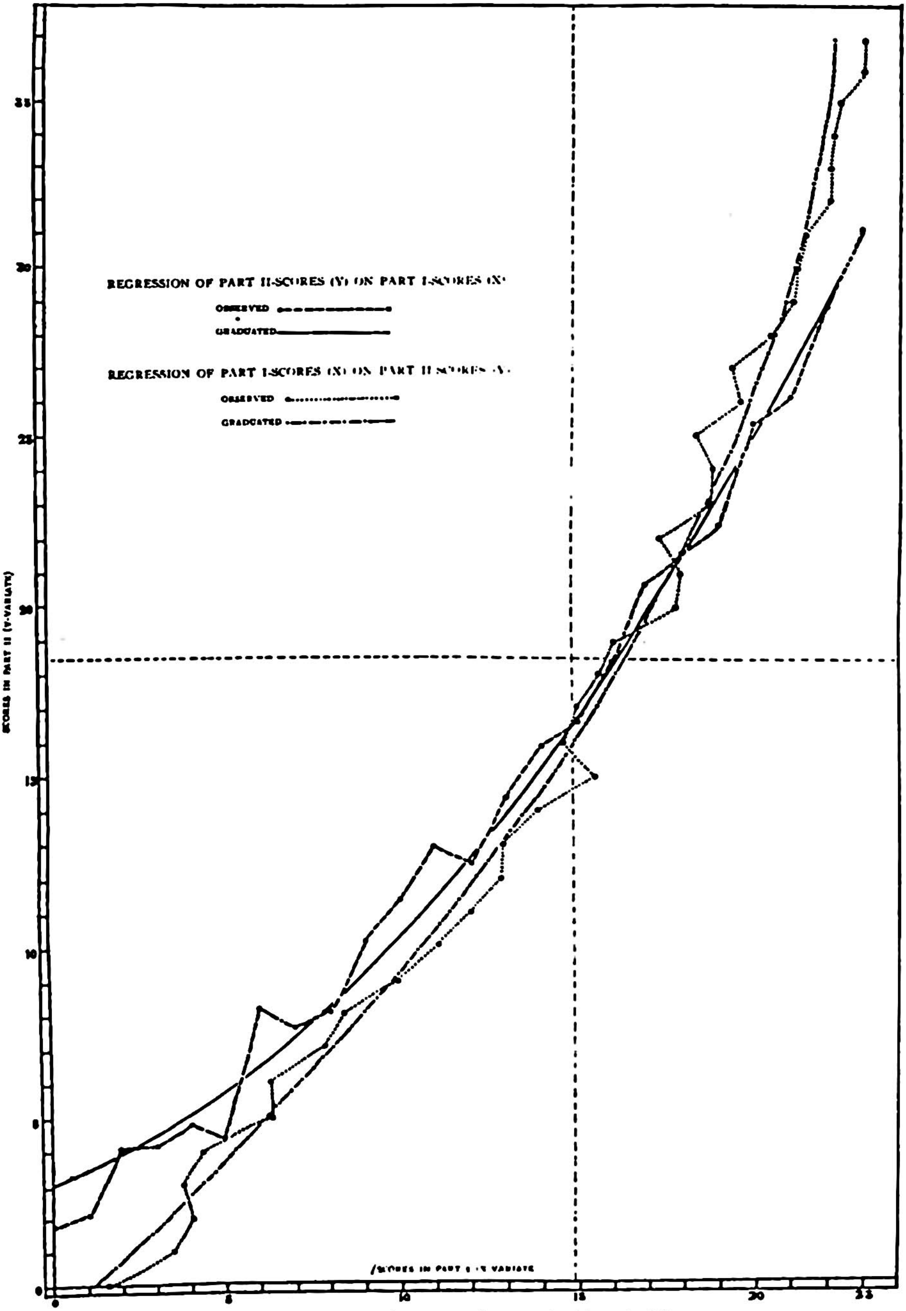

Fig. 3. Regression between Scores in Part I (X-variate) and Scores in Part II (Y-variate).

The two non-linear regression lines shown in Fig. 1 and Fig. 2 are superposed here to exhibit the closeness of association between scores in the two parts of the examination.

available number of independent comparisons is reduced to 23 minus 2 or only 21. Dividing 616·36 by 21, we have 29·35 as the variance for deviations from a parabolic curve. Fisher's test shows that it cannot be considered significantly greater than 22·5154 the purely random variance.[19] That is, the deviations from a parabolic curve are of the same order of magnitude as the purely individual fluctuations within classes. Hence we cannot expect to improve the graduation by using a third or higher degree curve. Fig. I shows the graph of the observed as well as the expected values as graduated by a parabolic curve of the second degree.

The mathematical process on which the above discussion is based is called the analysis of variance,[20] and it is usual to exhibit the results in the following way. (It may be noted that the number of independent comparisons is called the 'number of degrees of freedom' in statistical language.[21])

Nature of Variation	Degrees of Freedom	Sum of Squares of Deviation	Mean Square	Estimated S. D.
Deviations from 2nd order Parabola	21	616·36	29·3505	5·42
„ „ linear Regression	22	3,294·95	149·7705	12·24
Linear Regression	1	101,187·64	...	...
Between X-Groups	23	104,482·59	4542·7313	67·40
Within X-Groups	1188	26,748·25	22·5154	4·48
Total ...	1211	131,230·84	108·3657	10·41

$\eta^2 = 0·796174$ $\eta = 0·8923$

$\eta^2 - r^2 = 0·025109$ $r = +0·8789$

We must be careful about one point. In our sample we have no less than 130 students who are over 17 years of age. They cannot be considered typical representatives of a school population, and it will be desirable to consider what happens when this group of 130 students is omitted from the analysis. I repeated all the calculations after eliminating these 130 students, and I obtained the following results.

Nature of Variation	Degrees of Freedom	Sums of Squares of Deviations.	Mean Square	Estimated S. D.
Deviation from Linear Regression	22	2,404·32	109·2873	10·45
Linear Regression	1	93,385·89	...	...
Between X-Groups	23	95,790·21	4,164·8800	64·53
Within X-Groups	1058	23,757·79	22·4554	4·74
Total ...	1081	119,548·00	110·5902	10·52

$\eta^2 = 0·801270$ $\eta = 0·8951$

$\eta^2 - r^2 = 0·020115$ $r = +0·8838$

[19] See Statistical Note 10.

[20] This method has been recently developed by R. A. Fisher. See Statistical Note 7.

[21] See Statistical Note 6.

It will be seen that the results are in substantial agreement with the analysis of the total sample of 1212. The non-linearity of the relation between the scores in the two parts cannot, therefore, be ascribed to the inclusion of the older group of students.

We conclude therefore that scores in Part II of the test are very closely related with scores in Part I, and the dependence of Part II scores (Y) on Part I scores (X) can be satisfactorily represented by a parabolic curve of the 2nd degree.

$$Y = 3{\cdot}12 + 0{\cdot}3268\,(X) + 0{\cdot}0386\,(X^2) \qquad \ldots\ldots(3)$$

In algebraic equations the connexion between Y and X is rigid; if Y is given as a function of X, the dependence of X on Y is also determined at the same time. This, however, is not true for a statistical equation. Equation (4), for example, gives the *average* value of Y for any assigned value of X. But it is not possible to determine from the equation the *average* value of X for any assigned value of Y.

THE REGRESSION OF PART I-SCORES (X) ON PART II-SCORES (Y).

We must, therefore, investigate separately the dependence of X on Y, *i.e.*, of scores in Part I on scores in Part II. Proceeding in the way already explained, I obtain for the analysis of variance the results shown below. It will be remembered that scores in Part II range from 0 to 37, *i.e.*, we have 38 different Y-groups with 37 independent comparisons between class means.

Nature of Variation	Degrees of Freedom	Sums of Squares of Deviations	Mean Square	Estimated S. D.
Deviations from Linear Regression	36	1,830·34	50·8428	7·13
Linear Regression ...	1	46,046·86	...	...
Between Y-groups ...	37	47,877·20	...	...
Within Y-groups ...	1174	11,841·39	10·0864	3·18
Total ...	1211	59,718·59	49·3186	7·02

$\eta^2 = 0{\cdot}801714$ $\qquad \eta = 0{\cdot}8954$

$\eta^2 - r^2 = 0{\cdot}030654$ $\qquad r = +0{\cdot}8781$

It will be noticed that the Pearsonian Correlation Ratio η^2 for X on Y is 0·8954 and is practically equal to the corresponding coefficient for Y on X ($\eta^2 = 0{\cdot}8923$).

Applying Fisher's *z*-test[22] we find that 50·8427, the variance for deviations from linearity, is significantly greater than 10·0864, the variance within Y-groups (the probability

[22] See Statistical Note 11.

being again considerably greater than 100 to 1). We, therefore, proceed to fit a 2nd degree parabolic curve[23] and find:—

$$X = 1{\cdot}30 + 1{\cdot}0593\ (Y) - 0{\cdot}0132\ (Y^2) \ldots\ldots (4)$$

The graduated and observed values are given in Table 2 and are shown in graphical form in Fig. 2 (p. 35).

...

TABLE 2. REGRESSION OF X (PART I SCORES) ON Y (PART II SCORES).

Y_p	n_p	Values of X_p		Difference	Y_p	n_p	Values of Y_p		Difference
		Graduated	Observed				Graduated	Observed	
0	38	1·30	1·55	−0·25	19	29	16·65	16·03	+0·62
1	22	2·34	3·41	−1·07	20	31	17·19	17·81	−0·62
2	27	3·36	4·00	−0·64	21	30	17·71	17·96	−0·25
3	34	4·36	3·77	+0·59	22	33	18·20	17·33	+0·87
4	27	5·32	4·33	+0·99	23	25	18·66	19·00	−0 34
5	33	6·25	6·30	−0·05	24	32	19·11	19·06	+0·05
6	36	7·18	6·36	+0·82	25	32	19·51	18·38	+1·13
7	21	8·07	7·81	+0·26	26	47	19·89	19·66	+0·23
8	33	8·94	8·43	+0·51	27	43	20·25	19·44	+0·81
9	38	9·76	9·89	−0·13	28	57	20·59	20·49	+0·10
10	32	10·57	11·09	−0·52	29	46	20·89	21·08	−0·19
11	29	11·35	12·00	−0·65	30	54	21·17	21·22	−0·05
12	31	12·10	12·91	−0·81	31	40	21·42	21·38	+0·04
13	31	12·83	12·91	−0·08	32	39	21·65	22·05	−0·40
14	33	13·53	13·85	−0·32	33	35	21 85	22·06	−0·21
15	26	14·21	15·50	−1·29	34	26	22·02	22·19	−0·17
16	32	14·86	14·63	+0·23	35	22	22·17	22·36	−0·19
17	26	15·48	15·00	+0·48	36	5	22·29	23·00	−0·71
18	36	16·08	15·69	+0·39	37	1	22·38	23·00	−0·62

The sum of squares of deviations from graduated values is 346·26 for (38−2=) 36 degrees of freedom, thus leading to an estimated S.D. of 3·101 against 3·176 for purely random variations. We conclude that a second order parabolic regression curve is adequate in this case also.

We find therefore that the relation between X and Y is extremely close, the skew correlation ratio for both the regression of Y on X, and the regression of X on Y being of the order of 0·89. The closeness of the connexion is clearly seen in Fig. 3, p. 36.

[23] See Statistical Note 12.

Correlation for Small Age-groups.

The above result is based on the sample as a whole, *i.e.*, for all age-groups taken together. It is known, however, that the pooling together of heterogeneous material can give rise to a high correlation, even when the component parts are not highly correlated. We must, therefore, enquire whether the high correlation in the present case may be traced to the mixing together of the data for different age-groups.

The most direct way of investigating this point is to determine the correlation between the scores in the two parts seperately for each age-group. This is what I have done. The coefficient of correlation between the scores in Part I and Part II was found out by the product moment formula for each age-group separately. Table 3 shows the coefficient of correlation for one year age-groups.

Table 3. Correlation between Scores in Part I and Part II.

Range in Months	Midpoint (Average age)	Number	Correlation between Part I & Part II — Observed	Graduated	Fisher's transformation (z)
103—114	9 Years	20	+0·8210 ± ·0492	+0·8248	1·1600
115—126	10 ,,	51	·8678 ± ·0230	·8553	1·3333
127—138	11 ,,	113	·8446 ± ·0182	·8761	1·2369
139—150	12 ,,	158	·8862 ± ·0115	·8873	1·4038
151—162	13 ,,	157	·9478 ± ·0055	·8888	1·9185
163—174	14 ,,	192	·8577 ± ·0129	·8806	1·2846
175—186	15 ,,	201	·8620 ± ·0122	·8627	1·3012
187—198	16 ,,	131	·7923 ± ·0219	·8352	1·0776.
199—210	17 ,,	107	·8273 ± ·0207	·7930	1·1794
211 Upwards	18 ,, 6 months	80	·7620 ± ·0316	..	1·0010
100—204	12 ,, 6 months	1082	·8838 ± ·0045	...	1·3927
Weighted Averages	14 ,,	1212	·8751 ± ·0047	...	1·3545

It will be seen that the coefficient of correlation[24] varies from +0·7620 to +0·9478. There cannot be any doubt that even within one year age-groups the correlation between scores in the two parts is very high.

Fisher[25] has shown how we can compare the coefficients of correlation by a precise mathematical method. Applying Fisher's test we find that the value of the coefficient of correlation for the 13-years old students ($r = +0{\cdot}9478$) is significantly greater than all the other values, and the value $r = +0{\cdot}8862$ for 12-years old children is also significantly greater than $r = +0{\cdot}7923$ for 16-years old students. Otherwise the coefficients do not differ significantly.

[24] See Statistical Note 13.

[25] See Statistical Note 14.

We can easily obtain the mean value of r for different age-groups, making due allowances for the differences in the number in the various age-groups.[26] It comes out to be $r = +0{\cdot}8751$. With the already noted two exceptions of $r = +0{\cdot}9478$ (13 years) and $r = +0{\cdot}7926$ (16 years) all the coefficients are in sensible agreement with this value.

The above discussion shows clearly that even for one-year age-groups the value of the correlation coefficient is very high, and of the order of $+0{\cdot}88$, (the lowest observed value being $+0{\cdot}79$). This high value is therefore not due to the mixing of the records, and may be considered to present a real characteristic of the material.

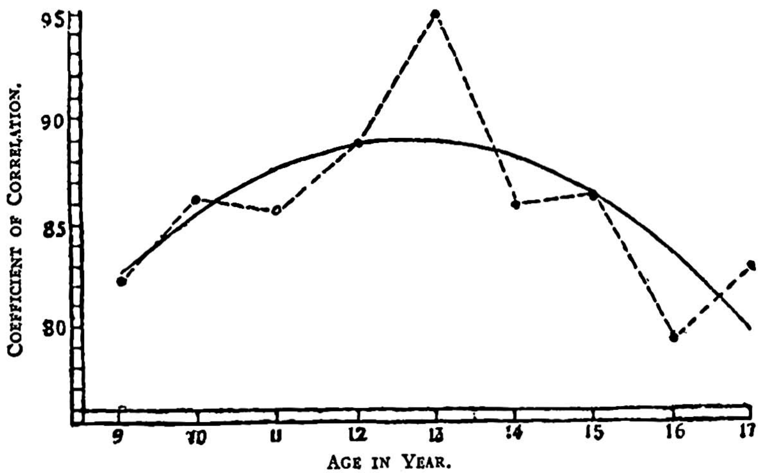

FIG. 4. COEFFICIENTS OF CORRELATION FOR ONE YEAR AGE-GROUPS.
The observed values are connected by the broken line, while graduated values are connected by a continuous line.

At the same time it is rather interesting to find that the higher values belong to the middle of the range, (*e.g.* the two highest values $+0{\cdot}9478$ and $+0{\cdot}8862$ occur for age-groups 13 and 12 respectively) while the lower values occur at either end (*e.g.* $+0{\cdot}8210$, $+0{\cdot}7923$, $+0{\cdot}8273$ for age-groups 9, 16, 17 respectively). If we graduate the successive values of the coefficients of correlations with a parabolic (2nd degree) curve, we obtain the 'graduated' values given in Table 3 and shown in Fig. 4.

There is some evidence to suggest that the numerical value of the coefficient of correlation between the scores in the two parts slightly rises at first, attains a maximum value of the order of $+0{\cdot}89$ between 12 and 14 years, and then slowly decreases. This result is corroborated by the fact that the correlation for students over 17 years of age ($r = +0{\cdot}7620 \pm 0{\cdot}0316$) is distinctly smaller than the correlation for lower age-groups. The magnitude of the probable errors precludes, however, any definite statistical significance being attached to this tendency.

[26] See Statistical Note 15.

In view of the intrinsic interest of the question I thought it worth while to calculate the coefficients of correlation wih a finer unit of grouping of 6 months. The results are shown in Table 4 and Fig. 5.

TABLE 4.

Range in months	Midpoint (Average age)	Number	Coefficient of Correlation	Fisher's Transformation
100–105	8 yrs. 7 months	10	+0·8561±0·0500	1·2785
106–111	9 „ 1 „	6	·7030± ·1373	0·8832
112–117	9 „ 7 „	23	·8783± ·0330	1·3783
118–123	10 „ 1 „	26	·8151± ·0453	1·1421
124–129	10 „ 7 „	41	·8739± ·0252	1·3492
130–135	11 „ 1 „	55	·8339± ·0279	1·2007
136–141	11 „ 7 „	77	·9009± ·0146	1·4768
142–147	12 „ 1 „	73	·8729± ·0189	1·3450
148–153	12 „ 7 „	88	·8974± ·0141	1·4580
154–159	13 „ 1 „	58	·8165± ·0298	1·1462
160–165	13 „ 7 „	104	·8476± ·0188	1·2471
166–171	14 „ 1 „	105	·8888± ·0139	1·4161
172–177	14 „ 7 „	83	·8577± ·0197	1·2846
178–183	15 „ 1 „	112	·8544± ·0173	1·2722
184–189	15 „ 7 „	76	·8423± ·0226	1·2290
190–195	16 „ 1 „	70	·8063± ·0284	1·1163
196–201	16 „ 7 „	54	·8310± ·0287	1·1913
202–207	17 „ 1 „	51	·7786± ·0376	1·0418

The weighted mean value of $r=+0{\cdot}8555$ for 6 months age-groups is slightly lower than $r=+0{\cdot}8751$, the weighted mean value for 1 year age-groups. The correlation for the pooled data (for all ages combined) is $r=+0{\cdot}8822$ and is naturally a little higher owing to the heterogeneity of the material.

The transformed values (Fisher's z) also bring out very clearly the steadiness of the correlation between scores in the two parts of the examination.

The most interesting, and for our purposes, the most important fact is that the correlations retain on the average the very high values of $r=+0{\cdot}8751$ and $r=+0{\cdot}8555$ for even such small age-groups as 1 year and 6 months respectively.

Owing to the smallness of the size of the samples, the differences are now negligible. But we again find that there is a tendency for the correlation to increase to a maximum near the middle of the range, and then to decrease slowly. (Fig. 5).

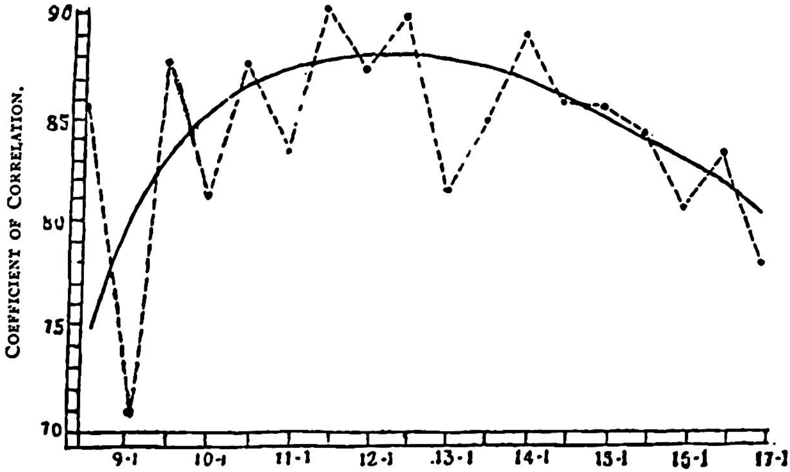

Fig. 5. Coefficients of Correlation for 6 Months Age-groups.
The observed values are connected by the broken line, while graduated values are connected by a continuous line*.

I have intentionally refrained from trying to improve the observed coefficients of correlation by using the correction for a split test owing to uncertainties in the statistical assumptions involved in applying such corrections.[27]

Conclusion.

We conclude that scores in the two parts of the examination are very highly correlated, showing that both parts of the examination are measuring traits which have a great deal (of the order of 80 or 85 per cent.) in common. There is also some evidence to suggest that the correlation is higher in the middle of the range, say between 12 and 14 years but owing to the magnitude of probable errors no definite statistical significance can be attached to this tendency. I believe a reliability of 0·70 or 0·80 (as measured by the coefficient of correlation) is usually considered statisfactory[28] by psychologists. Judged by this standard, the present test may be considered quite reliable.

*Graduation by free-hand.

[27] See Statistical Note 16.

[28] Spearman (13, p. 132), Brown and Thomson (1, p. 132) also Kelley (8, p. 211) who is of opinion that +0.50 is sufficiently high for group predictions, but +0.94 is required for detecting individual differences.

Note 1. The product-moment formula (often called the Pearson formula) was used throughout. Let $\bar{x}$, $\bar{y}$ be the mean values of x, and y respectively, and let N be the total number of individuals in the sample (in this case, 1212). The observed Standard Deviations in the sample are then defined by

$$N . s_x^2 = S(x-\bar{x})^2 \quad \ldots \quad \ldots \quad (1\cdot1)$$
$$N . s_y^2 = S(y-\bar{y})^2 \quad \ldots \quad \ldots \quad (1\cdot2)$$

where S denotes a summation for all N values, and the Coefficient of Correlation is given by

$$r_{xy} = \frac{S(x-\bar{x})(y-\bar{y})}{N . s_x . s_y} \quad \ldots \quad (1\cdot3)$$

In the present case the calculated constants were:—

$$S(x-\bar{x})^2 = 59718\cdot59,$$
$$S(y-\bar{y})^2 = 131230\cdot84$$
$$S(x-\bar{y})(y-\bar{y}) = +77735\cdot24$$

Mean marks in Part I $= \bar{x} = 14\cdot9167 \pm 0\cdot1356$
,, ,, Part II $= \bar{y} = 18\cdot4860 \pm 0\cdot2201$
Observed Standard Deviation for Part I $= s_x = 7\cdot0195 \pm 0\cdot0962$
Observed Standard Deviation for Part II $= s_y = 10\cdot4056 \pm 0\cdot1426$
Coefficient of Correlation $= r_{xy} = +0\cdot8781 \pm 0\cdot0024$

The regression of y on x is given by:—

$$(y_p-\bar{y})/s_y = r_{xy} . (x_p-\bar{x})/s_x \quad \ldots \quad \ldots \quad (1\cdot4)$$

The regression of x_p on y_p, on the other hand, is given by:—

$$(x_p-\bar{x})/s_x = r_{xy} . (y_p-\bar{y})/s_y \quad \ldots \quad \ldots \quad (1\cdot5)$$

It will be noticed that equations (1·4) and (1·5) are not algebraic, and one cannot be derived from the other. This is because the relations given in these equations represent average (and not individual) results. That is, equation (1·4) gives the average value of y_p for any given value of x_p, while (1·5) gives the average value of x_p for any assigned value of y_p.

Note 2. The observed (or sample) value of the standard deviation is given by equation (1.1), in which the square of the standard deviation is obtained by dividing the sum of squares of deviation by N, the total number of individuals in the sample. The best estimate of the square of the standard deviation (usually called the 'variance') for the whole population (of which the sample is only one small portion) is, however, obtained by dividing by $(N-1)$ instead of by N. The substitution of $(N-1)$ for N is intended to correct for the finite size of the sample. It will be noticed that the correction becomes increasingly unimportant as the size of the sample is increased.

Note 3. Let $\bar{y}_p$ be the mean value of the Part II scores in the pth class consisting of n_p individuals (all of whom had obtained the same identical score, p, in part I). Let $\bar{y}_o$ be the general mean score for all individuals.

Let y_{pq} be the individual score of the qth individual within the pth class. Then for each of the individuals in the pth class we have:—

$$(y_{pq}-\bar{y}_o) = (y_{pq}-\bar{y}_p)+(\bar{y}_p-\bar{y}_o) \quad \ldots \quad \ldots \quad (3\cdot1)$$

Summing this equation for all values of q from 1 to n_p, and using S to denote this summation, we have:—

$$\overset{q}{S}(y_{pq}-\bar{y}_o)^2 = \overset{q}{S}[(y_{pq}-\bar{y}_p)+(\bar{y}_p-\bar{y}_o)]^2 \quad \ldots \quad \ldots \quad (3\cdot2)$$
$$= \overset{q}{S}(y_{pq}-\bar{y}_p)^2+\overset{q}{S}(\bar{y}_p-\bar{y}_o)^2+2\overset{q}{S}(\bar{y}_p-\bar{y}_o)(y_{pq}-\bar{y}_p) \quad (3\cdot3)$$
$$= \overset{q}{S}(y_{pq}-\bar{y}_p)^2+n_p . (\bar{y}_p-\bar{y}_o)^2 \quad \ldots \quad \ldots \quad (3\cdot4)$$

since $\bar{y}_p$ and $\bar{y}_o$ are both constants, and $\overset{q}{S}(y_{pq}-\bar{y}_p)=0$. Summing equation

(3·4) for all values of p, we have finally :—

$$\overset{p}{S}\overset{q}{S}(y_{pq}-\bar{y}_0)^2=\overset{p}{S}\overset{q}{S}(y_{pq}-\bar{y}_p)^2+\overset{p}{S}[n_p(\bar{y}_p-\bar{y}_0).]^2 \quad \ldots \quad (3{\cdot}5)$$

For further discussion see R. A. Fisher. (4, p. 218).

Note 4. The extended theory of the statistical relation between two variates was first given by Karl Pearson in a memoir on Skew Correlation (12) in 1905. Using the notation of Note (3), we have

$$\eta^2=\frac{\overset{p}{S}[n_p(\bar{y}_p-\bar{y}_0)]^2}{\overset{p}{S}\overset{q}{S}(y_{pq}-\bar{y}_0)^2} \quad \ldots \quad \ldots \quad (4{\cdot}1)$$

where $\overset{p}{S}$ denotes a summation for all values of p, and $\overset{p}{S}\overset{q}{S}$ a summation for all values of both p and q.

The above equation is equivalent to Pearson's equations (xiii, xv) and (xvi) on page 10 of the same monograph (12). In equation (xix), p. 11, Pearson showed that η, the correlation ratio must always be greater than r, the correlation coefficient. A simplified discussion of the Correlation Ratio will be found in Yule (15, pp. 204-207), Brown and Thomson (1, p. 110), and Kelley (7, p. 238).

T. L. Woo (16) has recently published tables with the help of which the significance of η^2 can be directly ascertained. Unfortunately the table does not extend beyond $N=1000$. For larger samples it is necessary to use the formulæ given on p. 1 of Woo's paper. Although in the present example, there is absolutely no doubt regarding the significance of the association between the scores in the two parts, I am giving below the arithmetical calculations for testing the same with the help of the formulæ given by Woo (16, p. 1). If N is the total number of individuals, and n the number of arrays or classes, we have

$$\text{Mean value of } (\bar{\eta^2})_0 \text{ for no association} = (n-1)/(N-1)$$

$$\text{Variance of } (\bar{\eta^2})_0 \text{ for no association} = \frac{2}{N+1}\cdot\ \bar{\eta}^2\ (1-\bar{\eta}^2)$$

For the regression of x on y, the observed value of $\eta^2=$ 0.796174, $N=1212$, $n=24$. Thus $(\eta^2)_0=23/1211$ = 0·018993. This is the mean value of η^2 which would arise if there were no association between the variates. Also variance of $\eta^2=$ ·000,030,746.314 leading to a standard deviation of $(\eta^2)=$ ·005545. The difference between the observed value (0·796174) and $(\eta^2)_0$ for no association (·018993) is 0·777181, which is something like 140 times the S. D. of $(\eta^2.)_0$

Note 5. In the same memoir Pearson has shown (12, p.11) that $(\eta^2-r^2).\sigma_y$ "has an important physical meaning; it is the mean square deviation of the regression curve from the straight line which fits this curve most closely." When the regression is linear, it is obvious that (η^2-r^2) must vanish. This point is discussed in greater detail on p. 30 of the same memoir. Also see discussion in Fisher, (4, p.225), regarding the test of significance of (η^2-r^2).

Note 6. Fisher has explained (4, p. 222) how we should calculate the sum of squares of deviations of the expected mean values as determined by the linear regression equation. It is most simply obtained by squaring the product sum of the deviations and dividing by the sum of the squares of deviations of the independent variate. In the present case :—

$$\frac{[S(x-\bar{x})(y-\bar{y})]^2}{S(x-\bar{x}^2)}=\frac{(77735{\cdot}24)^2}{59718{\cdot}59}=101187{\cdot}64$$

Note 7. Fisher has shown that the best estimate of the variance in the population is obtained by dividing the sum of squares of the deviations by the number of degrees of freedom. For a sample of given size N, the number of degrees of freedom is obviously

$(N-1)$, since N being constant, we can only vary $(N-1)$ of the individual items. In case a linear equation is used we lose one further degree of freedom through the introduction of one independent parameter. With an equation of the 2nd degree, two degrees of freedom would be lost, and so on. For a fuller discussion of this point see Fisher, Statistical Methods (4), pp. 218, 219. The number of degrees of freedom may be interpreted as the number of independent comparisons possible.

The mathematical discussion is complicated and was given by Fisher in a number of papers (2, 3), but a very clear and lucid account of the important formulæ has been recently given by J. O. Irwin (6).

Note 8. Fisher defines 'z' to be half the difference between the natural logarithms of the two estimated standard deviations. He has shown that the probable error of z depends only on the number of degrees of freedom on which the S. D.'s are based, and has provided tables of values of "z" for two levels of significance, 5 per cent., and 1 per cent. (Statistical Methods, Table VI, pp. 212-216). We find from logarithmic tables:—

$$\log_e (149{\cdot}7705) = 5{\cdot}009092$$
$$\log_e (21{\cdot}5154) = 3{\cdot}114380$$
$$\text{Difference} = 1{\cdot}894712$$

Thus z=half the difference=0·947356, for $n_1=22$ and $n_2=1188$. (It should be noted that n_1 always refers to the larger variance). From Fisher's Table VI, p. 215, we find that the 1 per cent. point for $n_1=22$, and $n_2=1188$, is 0·3045. This implies that a value of $z=0{\cdot}3045$ or greater than 0·3045 would occur by chance in homogeneous material only once in 100 samples. The observed value of z is 0·9474; we conclude therefore that the variance due to deviation from linear regression is significantly greater than the purely random variance.

I have recently published certain Auxiliary Tables with the help of which Fisher's *z-test* can be applied directly without using logarithmic tables. (*Ind. Jour. Agri. Sc.*, Vol. II, Part VI, Dec. 1932).

We calculate the ratio of the two variances under comparison, and in this case find the ratio to be (149·7705/21·5154)=6·96 approximately. Looking up Table IV (11, p. 688-689) in my paper, we notice that the 1 per cent. value of the ratio for $n_1=12$ and $n_2=60$ is only 2·496. The observed value of 6·96 is, therefore, clearly significant.

Note 9. We can write Pearson's equation (lxiv) of his memoir on Skew Correlation (12, p. 29) in the following form:

$$\frac{(Y_p-\bar{Y})}{s_y} = a_0 + a_1 \frac{(X_p-\bar{X})}{s_x} + a_2 \frac{(X_p-\bar{X})^2}{s_x^2} \quad \ldots \quad (9{\cdot}1)$$

or writing $y_p = (Y_p - \bar{Y})/s_y$, and $x_p = (X_p - \bar{X})/s_x$, we get

$$y_p = a_0 + a_1 \,.\, x_p + a_2 \,.\, x_p^2 \quad \ldots \quad \ldots \quad (9{\cdot}2)$$

Here Y_p is the mean value of Y corresponding to the value X_p; $\bar{Y}$ and $\bar{X}$ are the general means, and s_y, s_x the standard deviations of the two variates respectively.

The moment co-efficients are defined by the equation:—

$$N.\, p_{pq} = S\,[\,(X-\bar{X})^p \,.\, (Y-\bar{Y})^q] \quad \ldots \quad (9{\cdot}3)$$

Certain statistical parameters are defined by

$$\beta_1 = p_{30}/\,p_{20}^3, \qquad \beta_2 = p_{40}/\,p_{20}^2 \quad \ldots \quad \ldots \quad (9{\cdot}4)$$

$$\text{and } f_2 = \beta_2 - \beta_1 - 1 \quad \ldots \quad \ldots \quad \ldots \quad (9{\cdot}5)$$

Further if η, r, are the correlation ratio and the correlation coefficient respectively, then

$$a_2 = -a_0 = \pm\sqrt{(\eta^2 - r^2)/f_2} \quad \ldots \quad \ldots \quad (9{\cdot}6)$$

$$a_1 = +r - a_2\sqrt{\beta_1} \quad \ldots \quad \ldots \quad \ldots \quad (9{\cdot}7)$$

In equation (9·7), $\sqrt{\beta_1}$ must be given the same sign as p_{30}, and the sign of a_2 has to be determined from given conditions. It is usually possible to decide by inspection whether the curve should be concave or convex to the x-axis.

By direct calculation I obtain :—

$\bar{X} = 14·9167$	$s_x = 7·0195$
$\bar{Y} = 18·4860$	$s_y = 10·4056$
$r^2 = 0·7710596$.	$r = +0·8781$
$\eta^2 = 0·7961741$,	$\eta^2 - r^2 = 0·0251145$
$\beta_1 = 0·421644$.	$\sqrt{\beta_1} = -0·6493$
$\beta_2 = 2·175542$.	$f_2 = +0·753898$

Hence we get $a_0 = -a_2 = -0·1825$, $a_1 = +0·9966$, and

$$\frac{Y_p - 18·49)}{10·4056} = -0·1825 + 0·9966\,\frac{(X_p - 14·92)}{7·0195} + 0·1825\,\frac{(X_p - 14·92)^2}{(7·0195)^2} \quad \text{.........(9·8)}$$

which may be further reduced to the form :—

$$Y_p = 3·12 + 0·3258\,(X_p) + 0·0386(X_p^2) \quad \text{.........(9·9)}$$

The graduated values were actually calculated from equation (9·8), and values calculated from (9·9) may be discrepant in the last place.

Note 10. We have in this case the following analysis of variance :—

Nature of variation	Degrees of Freedom	Variance×100	Logarithm (Natural)
Deviations from parabolic curve ...	21	2935	7·8845
Deviation within X-classes ...	1188	2252	7·7191

Difference = 0·2654

z is therefore = 0·1327. (It will be noticed that it is sometimes convenient to change the decimal point in both the variances to facilitate writing down the natural logarithm from tables. Further in this kind of work, it is usually sufficient to retain 4 decimal figures in the logarithms)

From Fisher's Table VI. pp. 212 the 5 per cent. point for $n_1 = 24$, $n_2 = \infty$, is $z = 0·2085$ while the observed value of z is 0·1327. The difference between the two variances is therefore entirely negligible.

The ratio of the variances is (2935/2252) = 1·3033. Looking up Table III (Mahalanobis, 1932, p. 687) we find that the 5 per cent. point for $n_1 = 24$ and $n_2 = \infty$, is 1·517. The difference is thus quite inappreciable.

Note 11. We have in this case the following analysis of variance.

Nature of Variation	Degrees of Freedom	Variance	Logarithm (Natural)
Deviations from Linear Regression ...	36	50·84	3·9286
Deviations within Y-classes ...	1174	10·09	2·3112

Difference = 1·6174

From Fisher's Table VI (p.215) we find that 1 per cent point for $n_1=24$, $n_2=60$, is $z=0{\cdot}3746$. Hence the 1 per cent point for $n_1=36$, $n_2=1174$ must be less than 0.3746. The observed value is $z=0{\cdot}8087$. We conclude that z is significantly different from zero, and hence the deviations from linearity are significant.

The ratio of the variances is $(50{\cdot}84/10{\cdot}09)=5{\cdot}0386$. The 1 per cent. point for $n_1=24$, $n_2=60$ is $x=1{\cdot}791$ (Mahalanobis, 1932, Table IV, p. 689). The difference is definitely significant.

Note 12. For the distribution of scores in Part II (Y-variate) I find by direct calculation for the whole sample of 1212:—

$\beta_1= 0{\cdot}028147$, $\sqrt{\beta_1}=-0{\cdot}1678$

$\beta_2= 1{\cdot}761421$, $f_2=+0{\cdot}735275$

The following values are already known:—

$\bar{X}=14{\cdot}9167$ $s_x= 7{\cdot}0195$

$\bar{Y}=18{\cdot}4860$ $s_y= 10{\cdot}4056$

$r^2= 0{\cdot}7710596$ $r=+0{\cdot}8781$

$\eta^2= 0{\cdot}8017136$ $\eta^2-r^2= 0{\cdot}030654$

Thus $a_0=-a_2=0{\cdot}2042$, and $a_1=+0{\cdot}8438$.

The parabolic regression of X on Y may then be written:—

$$X_p=1{\cdot}30+1{\cdot}0593\,(Y_p)-0{\cdot}0132\,(Y_p^2). \qquad \ldots \qquad \ldots \qquad \ldots \qquad (12)$$

Note 13. The probable error of a coefficient of correlation is given by $\cdot 67449(1-r^2)/\sqrt{n}$. In the Biometric Tables (13) the values of $\cdot 67449/\sqrt{n}$ are given in Table V (pp. 12–18), and the values of $(1-r^2)$ in Table VIII (pp. 20–21). The probable errors in the present paper were calculated with the help of the above Tables.

Note 14. Fisher has shown that the following transformation

$$z=\tfrac{1}{2}\,[\log_e\,(1+r)-\log_e\,(1-r)\,] \qquad \ldots \qquad \ldots \qquad \ldots \qquad (14{\cdot}1)$$

leads to an approximately normal distribution of z with a standard deviation of

$$s_z=1/\sqrt{(n-3)} \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots \qquad (14{\cdot}2)$$

The values of z given in column (5) of Table 7 are taken from Fisher's Table V (B), Statistical Methods p.177. The highest observed value is $r=+0.9478$ and $z=1.9185$ for the age-group 13. The lowest value is $r=+0{\cdot}7923$, $z=1{\cdot}0776$ for the age-group 16. Let us compare these two values.

Age-group	r	z	$n-3$	$1/(n-3)$
18 Years ...	+0·9478	1·9185	154	·006494
16 Years ...	+0·7928	1·0776	128	·008720
		Difference = 0·8409,		Sum = ·015214

The difference between the two value of z is 0·8409. The variance of this difference is given by adding together the two values of $1/(n-3)$ and is equal to ·015214, leading to a standard deviation of 0·1233. The actual value of the difference (0·8409) is thus considerably greater than six times its standard deviation (0·1233). The difference appears to be definitely significant, and the two values of r (for 18 and 16 years age-groups) must be considered significantly different.

Note 15. The weighted mean value of "z" can be obtained by multiplying each value of "z" by the corresponding value of $(n-3)$, adding and then dividing by the sum of all values of $(n-3)$. The weighted average value of "z" calculated in this way is 1·3345 (corresponding to $r=+0{\cdot}8751$), with 1103 degrees of freedom. Comparing the value with individual values we find that only $z=1\cdot9185$ ($r=+0\cdot9478\pm\cdot0055$), age-group 13) and $z=1\cdot0776$ ($r=+0\cdot7923\pm\cdot0219$, age-group 16) show significant differences ($0\cdot5640\pm0\cdot0860$ and $0\cdot2829\pm\cdot0934$ respectively) from the weighted average. For fuller details see Fisher, Statistical Methods (p. 170).

Note 16. For a split-test the improved value is sought to be found by the following Spearman-Brown formula (1, p. 132)

$$r=\frac{2(r')}{1+(r')} \qquad \ldots \quad .. \quad \ldots \quad \ldots \qquad (16)$$

where r is the observed, and r the improved coefficient of correlation.

In the present case $r'=+0\cdot8751$ (weighted average for one year age-groups), and hence $r=0\cdot9337$, which is extremely high.

Formula (16) is based on certain statistical assumptions. See Brown and Thomson (1, pp. 130-133, also Chap. VIII) and Kelley (8, Chap. IX). Several years ago I also had occasion to examine certain related questions in connexion with a different topic (10), and I came to the conclusion that the attempt to improve observed or "raw" correlations could not be justified on *a priori* grounds. As regards formula (6), Brown (1, p. 133) mentions that Holzinger has shown empirically that the assumptions made in deducing equation (16) do not always hold.

In fact, owing to both parts of the test having been given under practically identical environmental conditions there is considerable likelihood of accidental similarities in the results in the two parts being introduced through the operation of external causes. The fact that both parts of the test were also identical in form further aggravates the situation. Thus on the whole I am inclined to think that the observed raw correlation is more likely to give an over-estimate of the reliability, and probably furnishes a kind of upper limit for the same.

BIBLIOGRAPHICAL LIST.

1. Brown, W. and Thomson, G. H.: The Essentials of Mental Measurement. (Cambridge Univ. Press, 3rd edition, 1925).
2. Fisher, R. A.: On a Distribution Yielding the Error Functions of Several Well-Known Statistics. *Proc. Int. Math. Congress*, Toronto, 1924.
3. Fisher, R. A.: Applications of Student's Distribution. *Metron*, Vol. V, No. 3 (1925), p. 97.
4. Fisher, R. A.: Statistical Methods for Research Workers (Oliver and Boyd, 2nd edition, 1928).
5. Hamilton, E. R.: The Art of Interrogation (Int. Library of Psychology, Philosophy and Scientific Method. Kegan Paul, Trench, Trübner & Co., Ltd., London, 1929).
6. Irwin, J. O.: Mathematical Theorems involved in the Analysis of Variance. *Jour. Roy. Statistical Society*, Vol. XCIV, Part II (1931), p. 284.
7. Kelley, T. R.: Statistical Methods (Macmillan & Co., 1923).
8. Kelley, T. R.: Interpretation of Educational Measurements (World Book Co., New York, 1927).
9. Kimmins, C. W.: Presidential Address to the Educational Section of the British Association, 1929. *Nature*, 124 (1929), October 5, p. 516.
10. Mahalanobis, P. C.: On Errors of Observations and Upper Air Relationships. Memoirs of the Indian Meteorological Department, Vol. XXIV, Part I (1923).
11. Mahalanobis, P. C.: Auxiliary Tables for Fisher's Z-test for use in the Analysis of Variance. *Ind. Jour. Agri. Sc.*, Vol. II, Part VI, December, 1932.
12. Pearson, Karl: On the General Theory of Skew Correlation and Non-Linear Regression. (Draper's Company Research Memoires, Biometric Series, II, 1905).
13. Pearson, Karl, edited by: Table for Statisticians and Biometricians, Part I (Cambridge Univ. Press, 1914).
14. Spearman, C.: The Abilities of Man (Macmillan & Co., 1927).
15. Yule, G. U.: Introduction to Statistical Theory (Ch. Griffen & Co., 9th edition, 1929).
16. Woo, T. L.: Tables for ascertaining the Significance or Non-Significance of Association measured by the Correlation Ratio. *Biometrika*, Vol. XXI (A), 1929, p. 1.

INDIAN STATISTICAL INSTITUTE

The Indian Statistical Institute is a Society registered under the Indian Societies Registration Act (Act No. XXI of 1860).

Extracts from the Memorandum of Association.

II. The objects of the Society shall be:—

(1) To promote the study of Statistics both pure and applied and allied subjects.

(2) To provide for research and instruction for the advancement of the study and dissemination of knowledge of Statistics and allied subjects.

* * * * * * *

IV. **The membership of the Institute and its constituent bodies shall be open to all persons irrespective of sex, nationality, race, creed or class.**

V. **The Institute shall not distribute any part of its income or profit as dividend, gift, or bonus or in any other way among its members, provided nevertheless that office-bearers, workers, and other employees shall not be debarred from receiving remuneration or participating in grants because of their also being members of the Institute.**

The Institute consists of Associates, Ordinary and Life Members, Fellows, Honorary Members and Patrons. The supreme control including the power of making rules is vested in the members in General Meeting assembled. The President and other office-bearers are elected annually. The management of the affairs of the Institute is in the hands of a Council elected annually by the members of the Institute. There is a standing Technical Committee consisting of members who are specially qualified by their knowledge of statistics to undertake the direction or review of special investigations and technical problems. Members of the Technical Committee are appointed by the Technical Committee itself, and it is provided in the Rules that no opinion in technical questions shall be given on behalf of the Institute without the approval of the Technical Committee.

Membership Fee Rs. 15/- per year includes subscription to *Sankhyā: The Indian Journal of Statistics.*

Detailed information and Membership Forms may be obtained on application to the Hon. Secretary, Prof. P. C. Mahalanobis, Presidency College, Calcutta.