A study of the inconstancy of intelligence in relation to the problems of selection for secondary education at eleven plus

Daniels, John Clifford

.... A S T U D Y ....

of the INCONSTANCY of INTELLIGENCE in RELATION TO THE PROBLEMS OF SELECTION FOR SECONDARY EDUCATION AT ELEVEN PLUS ...

S u b mi t t e d

MARCH 1948

For the Degree of
MASTER of EDUCATION

of

DURHAM UNIVERSITY
(Durham Division)

by

JOHN CLIFFORD DANIELS, B.Sc.
(Dunelmis. 1927).

11, Beaufort Road,
Dialstone Lane,
STOCKPORT. Ches.

... C O N T E N T S ...

March,
1948...

CHAPTER 1.

HISTORY OF THE PROBLEM.

In this work we shall not be concerned with Selection for Secondary Education before the 1902 Education Act. Although a number of Foundation "Free-Place" Scholarships to existing Grammar Schools were given - places which were granted at ages 10 or 11 by means of competitive examinations, these were too few to warrant consideration here and certainly too few to pose the problems of selection which it is our intention to discuss. We will simply say that the children of the working-classes went to the public elementary schools up to the ages of 13 or 14, and that selection for secondary education was simply a question of the size of the salary of the parents. Although, as we shall see, salary-selection is of some importance to our problem, it is only important insofar as it has effects upon the efficiency of other techniques.

The history of the introduction of Universal Education in the nineteenth century is well-known. There is still considerable dispute, however, on the reasons which led to the growth of a large enough public opinion to carry these proposals on to the Statute Book. Too often the history of education is written as the history of great educators whose forceful propagandist efforts eventually succeeded in winning popular support. This

thesis does not subscribe to this point of view. The introduction of universal compulsory education in 1870 was a necessary development in human affairs. The progress of Society, especially its technical development and all the developments associated with it, would have been impossible without a literate working population. It must be remembered that although the major inventions in textiles, transport, and engineering, had been made several decades previously, it was not until about 1860-70 that these began completely to dominate production processes. The "sixties" found the "workshop of the world" running short of recruits to industry of a suitably high educational calibre fully to man the factories. That is why the "great Educators" succeeded in 1870.

We must ask ourselves a similar question about the introduction of municipal secondary schools in 1902. The state machine had considerably expanded during the previous decade. New scientific inventions were now pouring from the scientific institutions which were now in a position to plan fundamental scientific research on certain socially determined problems. The export of large quantities of capital to the Empire during this period was associated with the opening up of these backward areas - not simply in the sense of increasing markets of consumer goods - but of setting up industries of modern type in these countries. The cotton mills of India and the railways of the Empire countries serve as illustrations. All these developments required vast increases in the number of personnel of suitable educational status as given in the "public" and "grammar" schools of the day.

The purpose of the 1902 Act was to provide this secondary-trained personnel. It broadened the field of recruitment to the ablest of the children of the working-classes and the lower professional and middle-class sectors of the population for secondary education. These municipal grammar schools themselves, in their turn, began to set the pace for curriculum refor to broaden the existing syllabuses to include just those subjects which the needs of the time required. The public schools and the old grammar schools followed tardily and protestingly behind. Their protests were strikingly in the Canute manner.

It will be seen that the new Municipal Secondary Schools, although <u>in form</u> modelled on the old Public Schools, were setting themselves tasks which were qualitatively different from the tasks of their models. The models themselves, however, were changing, albeit hesitatingly and slowly and with many misgivings. But schools, because they are part of Society and not citadels above and outside Society, will always change as Society itself changes - or die, out-moded and starved of pupils as indeed a large number of the smaller Grammar Schools did.

It follows, therefore, that - in spite of the obvious presence in our modern secondary schools, grammar, direct-grant grammar, and public schools of today of vestiges of their antecedents - the tradition of the old grammar school was abruptly shattered and replaced at the turn of the century. The 1902 Act was the coup-de-grâce of this tradition and we must, in defining the aims

and purpose of secondary grammar school education, restrict ourselves to a consideration of the methods, curricula, and declared aims of the post-1902 schools.

It is recognised that this proposition is insufficiently argued here, but if this work is not to assume prodigious proportions, little else can be said. The following analogy, however, restates the thesis in a clear form. The British Parliament has many hundreds of years of tradition behind it. But in spite of the persistence during the centuries of ceremony and ritual, to liken the Parliament of King John with the Parliament of 1947 is to make nonsense of any theories of history. They are two extremely different bodies, unrelated in constitution, election, function and purpose. Just so unrelated are the Grammar Schools of 1850 and of 1947.

It is not necessary either to detail here the way in which the implementation of the 1902 Act was accomplished. Sufficient for us to say that it is from this date that modern problems of selection begin to arise. The "Scholarship" makes its appearance and from the outset - because of the restricted number of free-places - competition is keen. Even amongst the fee-paying pupils, by far the biggest proportion of the pupils admitted to the new Municipal Secondary Schools, there is some competition for entrance and intending fee-paying pupils are required to pass a qualifying entrance examination.

The Scholarship Examination now becomes a big feature in the lives of millions of children. Parents become infected with new ambitions - for their sons and daughters to be put on the first

rung of the ladder which leads to the next higher level of the social hierarchy. Local Education Authorities, sensible to public opinion as a body elected by a body elected on almost universal suffrage, now have to consider the Scholarship Examination in all its widespread social ramifications; its influence on the public elementary schools, its contribution to the success or failure of the new municipal grammar schools and it must be said, its unfailing capacity to provide a field for personal intrigue and nepotism. The Scholarship Examination begins to loom large and, in the period of the 1944 Education Act, suddenly bursts into the arena of open public controversy as one of the major educational problems of the day demanding either that a satisfactory solution be found which will compel everyone to feel that "justice is obviously being done" or that it be finally superceded by its opposite, the abandonment of selection and the re-organisation of the form of secondary education accordingly.

During the period between 1902 and 1945, the theoreticians and experimenters in education toiled at the current problems so that today we reconsider the problem of selection with a vast body of experimental data and of theoretical disputation throwing light and casting shadows upon the current discussions. A school of experimental psychology has grown up seeking to objectivise psychological science by the introduction of measurement and computation. These workers have produced startling new theories of the operation of the mind and have reached conclusions which

on an equal basis, vie with (though they sometimes concur with) the conclusions of such feet-on-the-earth physico-chemico scientists as the brain and neurological physiologists. Because of their tremendous influence upon selection, we shall have to return to consider the origins and history of the school of Psychometrists. This is done in detail in Chapter (4). Before this, however, it will benecessary to consider the history of the ideas associated with the purpose of Secondary Education, and to this we now turn.

The purposeof secondary education has been differently defined in different periods. The very use of the word "secondary" reflects some of these changing definitions. The term "secondary" only assumes importance after the 1902 Act to distinguish the sort of education given in the Secondary Schools for the selected few in institutions something similar to the Grammar and Public Schools. It has been shown that these were designed primarily to produce low-grade State and local Government officials and, to a limited extent, a supply of high-grade technicians. But, at the same time, these specially selected pupils had to regard themselves as different from and superior to their erst-while companions of the public elementary schools. The curriculum did this every bit as much as the "atmosphere" of the new schools. So that secondary education became associated with the study of "ideas" of "abstract principle", of "things of the mind". This "intellectual cult" - a turning away from the true source of knowledge, viz. human practice and activity, this elevation of the "printed word" over the "act of doing", arising inevitably from the structure of the Late Victorian and

Edwardian Society, still persists to confuse the real issues of the present. The Norwood Report, unable to free itself from the logical absurdities inherent in the "academic" cult, has provided a butt for a whole generation of scientific psychologists and pedagogues. Yet the current discussions on 11-plus selection are concerned primarily with these ideas for they shape the form of the controversy whilst throwing little light on its content. Selection for secondary education is pre-war days could mean nothing else than the selection of the "gifted few", the selection of those with "academic" ability, the selection of the future middle strata of administrative, professional and technical levels of occupation; the selection of the intellectual élite to give it the sort of education to enable it to play its predestined rôle in society. It is not necessary to detail the effects of this fact here, because below it will be necessary closely to review the selection instruments which the 1902 Act brought into existence. These selection instruments are, of course, the progenitors of the modern selection methods still operated under the 1944 Education Act, Yet the 1944 Education Act alters the nature of the whole problem. This is reflected in the new use of the word "secondary" education. Under the 1944 Act, secondary education ceases to be a type of education and becomes a stage. All full-time education between the ages of 11 and 16 (or 18 in certain cases) is secondary education. Thus any child, on reaching the age of 11, passes to a form of secondary education.

It will be seen that this nomenclature echoes the propaganda of the liberal educationalists in the inter-war years. During

this period certain political parties, notably the Labour Party, together with a number of organisations like the Workers' Educational Association, called for secondary education for all. True, much of this propaganda was shapeless, ill-defined, and sometimes thoroughly confused. But it played a great part in preparing for the 1944 Education Act. "Equality of educational opportunity for all" became a political election slogan on a number of occasions. Yet it was not this propaganda alone which brought the 1944 Education Act into existence. The political party which had championed "free secondary education for all" was still a minority party in the House of Commons, and still only very junior partners in the Cabinet when the Act reached the Statute Book. Why was this so? For exactly the same reasons as the 1902 Act. It was a necessary stage in the development of Society. The war in particular had thrown the problem into sharp relief although, masked by the economic depression, it had existed previously. There was an absolute shortage of trained man-power capable of handling the problem of modern industry, modern administration and modern warfare. The education system was as out-moded as the bi-plane fighters, the Bren gun carriers and anti-tank rifles with which Britain began the war in 1939. It was as out-moded as 95 per cent of the Lancashire cotton looms and the British coal industry of 1939. And, with the economic problems of the war showing the difficulties with a clarity not always discernable in times of peace, the Government began to tackle the question of skilled man-power at its very fountain-head - in the schools. For no Government can afford at any time to neglect its education system.

To do so is simply to commit national hari-kari.

But the educationalists and psychologists, pre-occupied up to 1939 with the problem of selection for secondary education, and conducting their experimental work with a belief in the permanence of the 1938 order of things were caught, like the lay-men - quite unawares by the changed conditions which the outbreak of war indicated. They began to search for answers to new questions of ~~secondary~~ selection, using old methods of approach. It is just because they did this that the 1944 Education Act, so definite and precise in many ways, is so non-committal about the crucial question of the form and content of secondary education. The "experts" had not made up their minds. The Ministry of Education had not made up its mind. Indeed, judging by the frequency with which it issues and withdraws pamphlets on Ministry policy on the subject it has not yet made up its mind. The professional psychologists, especially those associated with the working out of certain schematic interpretations of data collected in working the old system of élite selection within the framework of the 1907 Act, began seriously to quarrel in public about the importance or otherwise of certain isolated sets of data. Hence the Alexander-Burt controversy in relation to the emergence of a K factor at 11 plus or 13 plus, and other similar controversies. What they all failed to do is to take into account the fact that the selection problem is being asked under quite different conditions of society than it was being asked in (say) 1923. It is our intention to ask new questions more related to the necessary stage of society's development.

Other authorities, however, did begin to grasp the essential

difference in approach that would be necessary. It was a Trade Union leader who began to use the phrase - "vertical rather than horizontal selection". It was a leader of a political party who began to use the phrase "bi-dimensional selection". Numerous local authorities began to look around (and appoint) "experts" to solve difficulties which the "experts" for the most part had not seen but the members of the educational committees had. The year 1945 saw springing right to the front of public controversy the question of 11 plus selection. In one town, for example, a public campaign was led by one parent whose child had "failed" the preliminary examination qualifying his daughter to sat for the "scholarship" examination. This case reached the High Court which decided in effect that the Act of 1944 had changed nothing in this respect.

Local authorities hurriedly changed the name of the old "scholarship" examination to County Examination, etc. etc. As a result, we are treated to advertisements in the local newspapers of which thefollowing is representative. -

> The general Selection Examination for all children born between Ausut 1st, 1936 and July 31st, 1937 will be held on Wednesday, March 6th in the Primary Schools of the Authority. This examination will be the means of selecting entrants to the following schools....." (followed by a list of Grammar and Modern Secondary schools.

This is a piece of hypocrecy forced on the local education authority by unresolved questions. Everyone knows that if one still "fails" to get into a "grammar School one is automatically selected to go to the Modern School round the corner, upon which the legend still stands our boldly - "Board School".

How shall we formulate the basic questions which society poses to the educationalist and psychologist? More simply, but directly related, how shall we formulate the questions of the local education authorities in seeking to implement the section of the 1944 Education Act dealing with secondary education?

The formulation presents many difficulties to be sufficiently comprehensive. Here is a considered attempt:

(1) Are there three "types" of children, corresponding to the three types of Secondary School proposed?

(2) Are there three "spheres of life" corresponding to the three types of secondary school?

(3) What sort of children does society need today and tomorrow?

(4) Is there, in real life, a unitary trait, ability, aptitude, mechanism, which we can call intelligence and which is granted to individuals to greater or lesser degree?

(5) If so, can this intelligence be measured?

(6) How far can one predict at 10, the "type" of adult he or she is likely to be?

(7) How can one predict at 10. the amount of "intelligence" he or she will have as an adult?

(8) How far is it possible to make as many types as the education system is designed to make?

(9) How far is it possible to estimate the "demand" of society for each "type".

Thesenine general formulations of the questions at issue can be reduced to a number of more specific questions. Those given below, however, cannot cover the whole field of the nine general questions. -

(1) What validity have the theories of Factorial Analysis of Human Abilities which on the cognitive side, finds four main independent factors of the mind - general intelligence, a verbal factor, a "number" factor, and a "spatial relations" factor? This has a bearing on the nature of that which is measured by intelligence tests, upon the influence of "environment" on test results, upon the relationship between ability as shown in tests and ability as later shown in school and adult life.

(2) How constant is the Intelligence Quotient and measures of the special aptitudes? This has a bearing upon the age of selection or the fruitfulness or otherwise of trying to select children at 11 plus or at any other age.

(3) Do the proportions of the three "types" which certain psychometrists claim to have discovered agree with the proportions of these types known to be need by society at the present time?

These questions are fundamental to any contribution of the psychologist in the problem of 11 plus selection. Before proceeding to consider the results of the experimental work associated with this thesis it will be necessary to present and discuss the evidence already collected and the deductions which the theoreticians have made. Wtih this statement of the problem in these generalised forms we pass on to consider in detail the existing procedures for 11 plus selection. This survey should illustrate our problem and begin to suggest possible answers.

CHAPTER 2.

A SURVEY OF 11 PLUS SELECTION PROCEDURES WHICH HAVE BEEN OPERATED BY L.E.A.'S SINCE 1945.

All the schemes for 11 plus selection bear hall-marks of their history in the selection for the 1902 type of secondary school. This is necessarily so. The majority of L.E.A.'s, indeed, have not altered their selection procedures under the 1944 Act at all except to select a bigger proportion of entrants to replace the fee-payers. This increase in the proportion selected has exaggerated the validity problem. It is comparatively safe to select the top five per cent from a curve of marks which is approximately normally distributed. It is quite another matter to select the top twenty per cent.

Theprocedures group themselves into four main groups, as follow, but in each group there are often major differences as to the type of tests or criteria used:-

GROUP 1.

The Local Education Authority collects a body of information about all those pupils who wish to be considered for transfer to a Grammar School at 11 plus. This information includes one or more of the following:

1. Examination results in English and Arithmetic, set by and marked by officials of the L.E.A.
2. I.Q. from a standardised Intelligence Test.

3. Teachers' estimates of ability in school work.

4. Record of teachers' "marks" over a period.

5. Teachers' "assessment" of personality, usually in the form of the traditional report or even testimonial.

From this information selection of a number of "successful" applicants is made by "inspection" and arbitrary decision. The composition of the Selection Board varies considerably between L.E.A.'s adopting this method.

GROUP 11.

A single order of merit is drawn up by the L.E.A. of all pupils offering themselves for "examination" and consideration for transfer to a Grammar School. This order of merit is used as the only criterion affecting the offer of a place in the grammar school, the highest on the list being made such offers. The order of merit is arrived at in many different ways by different L.E.A.'s. The following are examples:-

(a) A "traditional" examination in English and Arithmetic. By this is meant an examination containing essay type questions. some formal grammar exercises, and the longer type of arithmetical problem. The "traditional" examination is posed against the "objective" test where many short questions with answers objectively defined from the basis of "objective" scoring of answer sheets. In the "traditional" examination the marks given are a dual function of the quality of the answers given and the marker. The "raw" marks (percentages) are added together so that the total score is weighted in selection to the "accidental" standard deviation of the two sets of marks. This usually means weighting arithmetic more heavily than English. Several important L.E.A.'s still operate this system.

(b) A standardised intelligence test result is used to supplement an English and Arithmetic Examination. The precise method of "adding" this result varies considerably from adding the I.Q. to the "percentage" marks in English and Arithmetic to adding the "raw" scores on this Intelligence Test to the marks in English and Arithmetic. The Standard Deviation of most I.Q. distributions is about 15 and this is usually greater than percentage marks in an Examination. It must be remembered however, that the Standard Deviation of I.Q.'s of children who present themselves for the 11 plus Selection Examination are usually much lower than 15. In two cases examined the standard deviations were 8.1 and 5.8 respectively. Often, therefore, by this method the I.Q. is adversely weighted in relation to the Arithmetic and English marks which often have standard deviations above these levels.

(c) A standardised intelligence test is used together with a standardised "objective" test in English and Arithmetic. By this means, if all the results of these latter are expressed in A.Q.'s (Attainment Quotients) with standard deviations about 15 like the I.Q.'s, the adding of I.Q. and the A.Q. gives equal weight to each test. Two authorities investigated doubled the I.Q.'s, thus giving equal weight to the Intelligence Test and the combined Attainments Tests.

(d) An intelligence test is used as the only criterion for preparing the order of merit. A number of authorities adopt this method and apparently neglect the difficulties associated with test reliability, which makes the value of single I.Q. point differences at the ~~central~~ critical selection point (usually 110 - 115) very doubtful.

(e) At least one L.E.A. adds to the results of standardised tests a "mark" based upon the record of the pupil in the primary schools. The contribution of this "mark" is of doubtful significance as it is well-known that the standard deviation of such "marks" is very low because teachers, anxious to do their best for their protégés, bunch their estimates nnear to the maximum.

By far the greatest number of L.E.A.'s adopt one of the many forms of this method of allocation. -

GROUP 111.

This method is one whereby ~~to~~ two determinants are used. -

(a) An order of merit in an examination:

(b) The offer of a place as a result of a high position on this order of merit is only made if the pupil has achieved a minimum standard on another test.

There are two forms of this method -

1. The order of merit is drawn up from the results of an intelligence test and the candidate has to satisfy the examiners that heor she has reached a certain level of competance in English and Arithmetic. This qualifying level is usually low and in one authority would only exclude 30 per cent of the pupils of the area.

2. The order of merit is drawn up on the basis of a test in English and Arithmetic, either of the "traditional" or the "objective" type, and acceptance is conditional upon having recorded a certain minimum standard in an intelligence test. This minimum I.Q. qualification usually excludes more than the minimum attainment qualification required in method 1. A figure of a minimum of I.Q. 110 is fixed by one authority. Taking the standard deviation of the I.Q. to be 15, this would exclude 74.7 per cent of the total child population.

GROUP 1V.

Most teachers and educationalists are acutely aware of the fact that preparation of pupils in certain primary schools for the 11 plus selection Examination has an adverse effect upon the teaching which is given in those schools. The curriculum is consequently narrowed down to lay special and inordinate stress upon teaching the sort of information

which will help their pupils to do well in the Selection Examinations. This important problem is discussed in detail elsewhere, but it is worthy of note that one or two L.E.A.'s have not restricted themselves to formal denunciations of the practice but have devised a method of 11 plus allocation which minimises the advantages gained by such special tuition, at least special tuition given by the schools themselves. One is described below to illustrate the principle.

An intelligence test is administered to all children between certain given ages living in the boundaries of the Authority. From these results the "intelligence standard" of each primary school is calculated. From this and the number of available grammar school places each primary school is given a "quota" of places. The next stage is for each school to test all its eligible pupils in English and Arithmetic by means of standardised tests. The results of these tests are added to each child's I.Q. as found by the original intelligence test, and the selection of the school's quota made from this combined score. Thus it is no advantage to the school's "Scholarship results" as a whole to ensure that the "absolute" standard of the English and Arithmetic papers is high. On the other hand, a series of anomalies can arise.

But the most serious danger is that certain schools will, following their former practice of establishing a reputation for "good scholarship" results, begin to instruct their pupils in the technique of doing intelligence tests. All the evidence goes to show that practice and advice can raise the I.Q. of a child by an average of 10 points. Such practices would undermine the validity of intelligence testing but, reprehensible though they are, it is certain that they are not so uncommon as one would like to believe.

The disadvantages of this scheme, as opposed to a single order of merit, is that it is not "obviously fair" to all. Parents of certain children displaced in the "second round", yet who contributed to the success of the child who displaced him, would feel a grievance at the system. It is axiomatic that any good scheme of selection must satisfy parents as to its "fairness".

This survey is necessarily incomplete. No full report indeed has yet been made in recent years and such a survey would seem to be indispensible to research if we are to come to some firm conclusions about the future of eleven plus selection.

It will be seen, however, from this survey that the Intelligence Tester has won for himself a position of authority which demands that he shall fulfill all that is claimed for him

or abdicate.

The Intelligence Test is made the instrument of what, for all practicable purposes, is an irrevocable decision concerning the whole future of an individual. It is given on one specific occasion in most areas. Although psychologists recommend two or three tests to be given over a period of months during the tenth year, few, if any, L.E.A.'s follow their advice.

The broadest aspects of test reliability, therefore, ar e of the utmost concern, both for each individual and to the educational system as a whole. It is most important to be able to give definite answers to a series of questions about the stability, constancy, reliability, and validity of intelligence tests. The immediate future of 11 plus selection and the tripartite system of secondary education under the 1944 Education Act depends upon the nature of these answers.

CHAPTER 3.

THE HISTORY OF SELECTION PROCEDURES.

The detailed history of 11 plus selection procedures between 1902 and the present time is bound to be sketchy because little or no information has been gathered in a complete and compact form. One traces, therefore, the history of these procedures as they moved, step by step with the development of selection theories by the rising school of psychometry and as they also reflected new and more complicated problems because of their changing field of operation.

The building of municipal secondary schools took place slowly before the first world war. Most towns of 30,000 or more inhabitants did build such schools or took over and adapted existing schools. The number of pupils catered for was small - usually little in excess of 5 per cent of the child population at eleven. The proportion of fee-payers to free-place scholars varied widely from town to town. It follows that in the main, the task a scholarship examination was asked to do was to select the top 2 or 3 per cent of the ten-year olds. This task is, of course, as is well known, comparatively simple. Almost any school examination if made difficult enough could do this fairly efficiently. The tests, usually given to a very small number of children in the actual building of the secondary school, were traditional examinations in English and Arithmetic. The required standard in these subjects was, however, well in advance

of whatis required at the present time and many authorities issued printed syllabuses for their scholarship examinations one or two years before the examination was due to take place. This included "set" books and Shakespearean plays in English. Other L.E.A.'s included detailed History and Geography papers in the e examination. To "pass" such a scholarship examination was no mean achievement. Indeed, the papers did look for marked precocity and blithely assumed that precocity was a reliable guide to estimating the future "success" of the child.

The "scholarship" class under these circumstances was only heard of in isolated cases. The main contribution to helping their pupils to obtain a "scholarship" was to send them precocious pupil careering ahead in the classes of the public elementary school. It was not rare, therefore, even in the nineteen-twenties to find 9 and 10 year olds in St ndard 7, amongst the pupils about to leave. Special coaching after school hours by private individuals, however, was common and indeed the usual practice.

After the 1914-1918 war, secondary school accommodation rapidly expanded. Twice the "normal" number of pupils were often crowded into these schools and overcrowding and "annex" buildings became common features. Also, L.E.A.'s began to increase the proportion of non-fee-paying pupils and a number of progressive L.E.A.'s made "passing the scholarship" a condition of entry into their municipal secondary schools. The percentage selected jumped from 2 per cent to as many as 25 per cent with an average for the

country as a whole of between 10 and 15 per cent. The problems of selection were accordingly complicated. The "selection mark" becomes nearer and nearer the "peak" of the "cocked hat" curve of examination marks - so increasing rapidly the proportion of "borderline" cases.

Before the 1944 Act, the "efficiency" of a selection examination was the percentage of children who "passed" the 11 plus scholarship and "failed" in the secondary school. (Failure here usually meant failure in the School Certificate Examination. A detailed discussion of validation criteria is to be found below). It was, for official purposes, at least, of little concern how many children who could have passed the school certificate examination never entered a secondary school - either because they were never "entered" for a scholarship", or because they "failed" when so entered. Now with the increase of the percentages of children selected for secondary school education, this "failure" percentage began rapidly to increase. Nor did this increase in simple proportion to the percentages selected, but in proportion to some high power of this. By 1924 it had become a "problem" of the first magnitude in some areas. For a number of reasons, it became difficult to detect differences in performance between fee-payers and "scholarship" pupils, a state of affairs which was never true in pre-1914 days.

The selection problems associated with the "shape" of the normal curve at the discrimination point were not the <u>only</u> ones contributing to the crisis. The form of the examination itself

contributed a great deal. One deficiency in this respect was the detailed nature of the syllabuses issued. This entailed specialised extra school work by the junior pupils and in an increasing number of cases of pupils of outstanding ability this was impossible. Again, the teaching of history and geography is well known to vary in content and form between schools and teachers to such an extent that at 10, it is difficult to prepare a suitable examination paper to give every child equality of opportunity.

Nevertheless, by 1922, a number of workers in education began to address themselves seriously to the whole problem of 11 plus selection for secondary education. The time was opportune for other reasons. Suggestive data from other fields of education and selection began to come in on an ever-increasing scale. First, Binet's work on testing mentally deficient children had set the train of experimental and theoretical work which had led up to Spearman's work on factorial analysis and intelligence. The vast experiment in group testing and selection in the American Army also began to throw light on some of these problems of 11 plus selection. So that we have emerging at this time into full prominence a number of workers whose monumental work over the last twenty years has simplified our task in answering our modern questions. A few names to illustrate must include - Sir Cyril Burt, Ballard, Valentine, Thompson, Alexander, Philpotts, Spearman, McClelland and Vernon. These are not by any means the only leaders in psychological research. They are, however, some whose work has

been specially directed towards the problems of "scholarship" examinations and related problems.

VALENTINE'S first investigation concerned the validity of the traditional type scholarship paper. He found, much to everyone's surprise, that the correlation between order in the scholarship examination and in the School Certificate examination was zero. He found, on the other hand, that the correlation between the first term school examination results and the order in the School Certificate Examination was sometimes as high as $r = .6$. He reported that when a group intelligence test was added to these scholarship papers, this correlation (order in scholarship with order in School Certificate) rose to between .3 and .4 .

What is the explanation of this? Without wishing to anticipate some of our later answers, it would seem that the contribution to examination results of the common cultural and intellectual pattern of all stages in a secondary school far outweighs in importance the sort of "intelligence" or ability measured by the Scholarship examination and which, on the surface, one might have thought would have had greater persistance than these results indicate.

Nevertheless, the introduction of intelligence testing into the Selection examination has gone on since the early twenties until very few authorities neglect to include one in their tests.

At the same time, a great deal of research work has been done designed specifically to improve the efficiency of the scholarship examination. This itself was a great step forward. Recognition of the existence of a problem and experiments with even negative results is an important step in the right direction.

This research to improve efficiency of eleven plus selection was conducted on three main lines, two only of which can concern us here to any great extent.

(a) Improving the form and content of purely "intelligence" tests. This poses the question of "validation" of these tests and two main approaches have been made. Firstly, there are those who, following the approach of Binet to its logical conclusion, say that intelligence tests are specific instruments designed for a specific selection task. Binet built up his test in order to diagnose mental deficiency and his sole criterion for including or rejecting any test item was whether or not it helped, in the diagnosis. Secondly, there are those who have followed Spearman's approach to its logical conclusion. Those who wish to improve 11 plus selection intelligence tests by the Binet method do so by choosing items which correlate highly with school certificate results or some comparable criterion. Spearman started not with a practical problem to contend with but, "by taking thought" attempted to define the elements of intelligence. The definition he finally arrived at was the ability to grasp complex relationships. Then he showed by a series of mathematical

and statistical analysis, that this "intelligence" pervaded the ability to do tasks of various kinds to a greater or lesser degree. This he called the "g" factor. Those who apply Spearman's theories to eleven plus selection make the assumption that "g" is an important (usually the most important) factor in determining the type of secondary school which pupils are best suited - e.g. Grammar School pupils need a high "g" ability. To improve intelligence tests for 11 plus selection by this method, means in the classical Spearman way, to construct tests which are reliable and highly saturated with "g".

(b) Improving testing of attainment in school subjects. Research in this field has been towards the construction of more reliable tests of English and Mathematical ability. Again, there are two methods of approach. There are those who have concentrated on constructing attainment tests in English and Arithmetic which give simply highly consistent results given a "common" educational background. On the other hand, there are those who make a special point of the differences of educational background of Junior school pupils. The content and methods of junior school education differs considerably between schools. For this reason certain research workers have concentrated their energies on producing tests which presume to measure English and Arithmetical aptitude. This presupposes validation of test items by a criterion of later success - e.g. School Certificate results.

(c) Research in devising methods of using school records in 11 plus selection. This subject is so vast that little can be said here. Enough to say that the results are very disappointing to those who prize high "consistency" and high "validation" of the criteria.

It must be said that this research has brought some important results. Valentine, in the early twenties - as has been mentioned - found that the correlation between the order in the School Certificate examination and order in the "Scholarship" examination, was zero. By using the best, recently developed, batteries of intelligence and attainment tests this has been increased so that many areas report correlations of .5 or .6 . These seem very high claims in view of the following evidence. Correlations between examination order after one term in a grammar school and order in the 11 plus selection examination for three grammar schools in thr e different areas were found by the present writer to be:

Area (1)	r	=	.54	$\pm$ P.E.	.083
Area (2)	r	=	.04	$\pm$ P.E.	.108
Area (3)	r	=	.32	$\pm$ P.E.	.064

The number of careful researches in this direction seems to be very small yet even if we accept a correlation of .6 as the best attainable "validation" coefficient of the best 11 plus selection examinations, we must ask ourselves if this is satisfactory.

Below it is argued that to be satisfied with this means losses in the production of able recruits to man our modernised industries and our reorganised administrative machinery which are far too heavy for Britain to be able to afford. These losses include those who enter our grammar and technical schools and are later judged "failures", and also those who do not enter our grammar and technical schools but who would have succeeded there. These latter are probably the more important. A deal of evidence exists that the thirteen plus transfer pupils make a great success of grammar school courses even though they failed at 11 plus to qualify for entrance. Again, we must consider the number of adolescents and adults who, after failure at 11 plus, take further courses of instruction in our Technical Institutes and Extra-Mural University classes and succeed in reaching standards far beyond that of the vast majority of the grammar school intakes. Motivation here obviously plays a rôle which completely overshadows the importance of ability as measured by intelligence tests and the like. Yet the claims of the "whole-hogger" intelligence testers would deny the possibility of so working "above capacity". Obviously their "capacity" has increased in these cases. We shall deal in detail with this aspect.

Finally, whist still on this point, the experience of the Fighting Services in searching for and finding able men and women capable of learningcomplex technical matters in a few months, needs to be taken into account. The reserves of ex-grammar school pupils were rapidly absorbed in filling commissioned and technical

posts. They had to look elsewhere. They developed new techniques of selection in this search, and provided many hundreds of thousands (if not millions) of men and women whose education had been confined to that of an elementary school to the age of 14 years but who proved to be as capable in the execution of difficult and complex tasks as their ex-grammar school colleagues.

For this reason it must be frankly stated that the "Scholarship" examinations developed up to the present have failed to give an efficiency of service suited to the demands of our times. Some of the reasons for this failure, it is the main puppose of this research to supply.

At the heart of the problem is the rôle of "intelligence" as measured by 'intelligence' tests. Do "Intelligence" tests - as at present constructed - measure a unique trait which is of dominant importance in determining the manner in which an individual "copes" with a given situation, especially a situation containing novel features? It is not possible to classify problems into intellective and non-intellective. Even more difficult does this become if one used the terms "academic" and "non-academic" as do so many authorities - including the Norwood Report. An academic education is, strictly speaking, that given in a college or academy, i.e. a higher, post-secondary, education. It has come to mean in the language of this decade the education given in grammar-type secondary schools. It is sometimes placed in contra-distinction to a practical education. Many equate academic learning with book-

learning without "practical" or "manual" work to assist in the grasping of principles. If this is academic education it is certainly not that given in the grammar schools. On the whole, these schools are far better equipped for "manual" and "practical" work than any secondary Modern school.

Therefore, the ~~concise~~ crucial question still remains - what does an Intelligence Test hope to measure? Educationally we have to ask the question in this definite form. Does an intelligence test measure the ability to cope with the sort of education given in our secondary schools today? There is indeed a certain Common Denominator in all our secondary schools. By this is meant not the "common core" of subjects but a certain fundamental approach to the whole problem of schooling - (a) class-teaching; (b) subject teaching; (c) common core of subjects; (d) teacher-child relationship - all intergrated together producing the British system of twentieth century education. Do intelligence tests measure the general ability to cope with the essentials of this complexus of social-relationships we call a secondary school?

Before we give a considered answer to this question it will benecessary to review modern theories of intelligence and critically to formulate a reply. This is done in the next chapter. But there is still one further problem whose proper location is here, viz. In all this work, what must be our criterion of success? If it is answered incorrectly we shall be running the danger of constructing an artefact. Put into terms of a practical example the problem becomes clearer. Suppose we are planning a

research to discover if a certain test gives a "good" forecast at 11 plus of ability successfully to complete a grammar school course. There are two obvious courses open - (1) to give the test to pupils who have completed a grammar school education and use the fact as to whether they have been "successful" or not as the criterion, and (2) give the test to eleven year olds and follow their careers through the grammar schools. The objection to the first is that this assumes the correctness of the first selection, and is therefore merely discovering if a given test can do exactly what the test which actually did the selection at eleven did itself do. The objections to the second procedure are more fundamental. Grammar education is not a fixed "impact" upon isolated individuals. The education will be varied to suit the child and thus the measurement of success will be more a function of school and teacher adaptability than of the individual capacity and aptitude of the child. Here again, therefore, there seems to be an artefact.

The criterion in our researches, therefore, if it is not going to fall into the trap of internal inconsistency, must be seen to be <u>historically determined.</u> Once the significance of this is grasped, the problem becomes clearer. That is why we have spent so much time elaborating the history of secondary education, in identifying its determinants, of describing the inter-action of the development of educational theories and social practice. In order to settle the criterion of success of 11 plus selection we shall have to go beyond the bounds of measuring how the

selected individuals fare in this or that type of now existing school. To confine one's criteria to these limits stultifies real research into the problems. Subsiduary criteria there are in plenty - the School Certificate, post-school salary, the total amount of knowledge absorbed during the whole of school career found as the summation of all school and examination marks; one could multiply these into the realm of fantasy. But events, practice, history, education committees, employers, society, will bring us back time and time again to the basic question - are the schools, as a result of their structure and form, producing the sort of social individuals who can operate the society into which they are born? Society changes and with it so does education. Every society gets the education system it needs. If we understand the inevitability of this and understand also how and in what direction society is changing then we can smooth the changes in educational form and content which will be made necessary by the changes in society which we foresee. Nothing definite is stated here, therefore, on the question of criteria. It will be seen that, as we consider our own problem in its historical development, the full importance of the generalised criteria will emerge. It is the only way of escaping out of a vicious circle into which most selection research has lost itself today.

CHAPTER 4.

INTELLIGENCE TESTS AND INTELLIGENCE.

The only truly adequate definition of intelligence at present is the merely formal "intelligence is that which is measured by intelligence tests". But this adequacy begs the question and relegates the psychology of cognition to the level of sophistry. It is rarely realised what a hold-all phrase "intelligence" is for the infinite variety of human ability and aptitude. In fact, "intelligence" is more than anything else a socially approving term. In a rather similar way the word - "race" came to have so many diverse meanings that scientific anthropology has had to cease using the word "race" at all. In like manner, psychologists will be well advised similarly to drop using the word "intelligence" and substitute for it some word which can be given universal precision.

The assumption of such a thing as intelligence is implicit in the very construction of a test for it, and too often the construction and standardisation of the instrument is taken to prove the real existence of that it purports to measure, as a species of mental entity. One might as well claim to know every property of electricity or its inmost "nature" after discovering that a current in a conductor obeys Ohm's Law. Such a word as "intelligence" is incredibly vague and indefinable and lends itself easily to the multivalenciesof the subtler emotive types of argument. The word is taken to be the thing - the symbol is mistaken for the existing - this type of logical error is very easy here.

Many psychologiests have tried to define the nature of intelligence and each definition bears the mark of the makers's philosophy. Thorndike is an Associationist. Intelligence is therefore defined as the multiplicity of associations or mental bonds between experiences. His test CAVD, C for completion of sentences, A. for arithmetic, V for verbal, and D. for distinguishing differences, is constructed on the basis of this Associationist assumption; of creative thought and activity he can give no account.

Spearman's definition arises out of a purely statistical approach. He gives two possible definitions and leaves his readers to make their own choice. The "g" factor isolated out of his tetrad equations was either the mental property of synthetic intelligence which facilitated the bringing together of perceptual relations and the education of correlates, or a unique, specific mental energy related in some complicated manner to the action of the Central Nervous System. Spearman, the logician, defines intelligence in terms of a heuristic metaphysical principle which has no practical importance in the construction of the tests to measure it. The tests are constructed - not to give the greatest measure of the ability to educe correlates, but to give the greatest measure of "g" - the statistical isolate of the tetrad equation. Intelligence becomes, therefore, any attribute which can be measured the most consistently. This approach, which is fundamentally that of all the factorial analysts is fundamentally useless to a scientific and realist approach to the problem of intelligence. The tests used in this research were constructed using this same

principle. The reason for this is that at the present stage of the research it is most important to make a critical appreciation of contemporary psychometry and to show if this approach does indeed stand the realist tests of our criteria.

K.S. Lashley's use of the word "intelligence" in his book BRAIN MECHANISMS AND INTELLIGENCE, differs radically from any accepted usage but for our purpose it is more promising. He measured the degree of intelligence in the behaviour of rats by the efficiency and economy of time with which they learned a certain problem - their ability to run mazes of differing degrees of complexity with the least time, and the least number of mistakes. By carefully controlled experiments during which he operated on the brains of the rats, he was able to show that the pre-operative memory or specialised areas of the brain were much less responsible for intelligent behaviour than the fact that the brain operated as a whole, with loss of intelligence varying directly as the loss of surface area of brain tissue. This "holist" point of view is a physiological corroboration for the views of a school of psychologists best represented by Dr. Mary Fleming. The views of this school will be given further treatment below.

Koehler showed that the intelligent behaviour of his apes was little dependent upon training and memory, than on their ability to make use of experience in grasping the significance of a problem and then to solve it in the most efficient manner.

The psychology of learning it would seem, therefore, to should

throw a greater light upon the operation of intelligence. This field itself, however, is confused. There is a vast mass of observed phenomona and accounts of experiments, but practically nothing of value on principles. No one has succeeded in reducing learning to a simple neurological foundation. Thorndike and Guthrie assert that learning takes place by a simple process of "stamping in" reward for response and repetition. The first conditioning theories were as simple and as open to doubt. Learning does not occur in this mechanical way, by the re-inforcement of a response but as B.F. Skinner points out, there are at least two main types of conditioning - and by far the most important is the purposive conditioning in which an animal acquires the capacity of re-acting to a signal which sets for it the situation to follow, and of acting with implicit foresight. In the experiments of Finch, Culler and Brogden, when the bell rings the dog no longer struggles wildly to escape the shock and instead lifts its paw calmly and efficiently a fraction of a second before the shock is due to come.

Tolman has built up an attractive and coherent theory of behaviour on these lines by showing how signal behaviour and the ability of one segment of experience to release the next appropriate response can be built up in a series and in a

definite hierarchy to account for, for example, the way in which a rat comes to realise what food or water goal it is running for in a maze and the way in which it ~~leaves~~ learns the maze from the goal backwards. Krechevsky insists that rats in learning mazes show a higher degree of variability in behaviour when near learning or running successfully than when their capacity to learn has been damaged by operation when behaviour becomes fixed and stereotyped. He suggests that the intelligent animals are in fact, using a kind of implicit "hypothesis", deliberate trial and error of a kind expected only in the higher forms of life. The apes of Koehler certainly demonstrate this experimental intelligent approach to problems - C.L. Hull set out to study the formation of concepts in human beings by a study of learning. He exposed to his subjects a series of cards bearing Chinese characters, each in a group of six, having a different radical somewhere in the character. In the next group of characters, the same six radicals occurred again and there were altogether six groups shown. The purpose of the experiment was to discover how soon and by what means the common factor of the six radicals - their concepts - would be perceived and learned by the subjects. Hull discovered that mechanical learning failed completely to explain the way in which the significance of the concepts were apprehended, and he noted a growth from a first vague realisation to a species of trial and error behaviour rising to an inductive mode of inference which became more and more successful as the learning trials were repeated and the subjects gave answers nearer and nearer the truth.

So, out of what might be accounted for by the mechanics of conditioning, there appears to evolve spontaneously, "insightful" behaviour and inductive reasoning on the data of experience in the face of a problem-situation. The apprehension, on however primitive and implicit a level, that a problem was facing one, was the first indication of "thought" and intelligent behaviour.

Psychologists have made great use of the "problem-solving" technique in recent studies of the thought processes - Smoke, Selz, Duncker, to mention only a few - but there again there is no agreement on what processes are at work in problem solving, and what relative contribution such lower organisational levels as memory and conditioned behaviour - lumped together as experience - make to intelligent learning and thinking.

We should be not at all surprised to find that the apparent higher faculties had evolved from and bore a similarity to the lower ones. Human memory itself is organised experience, and without memory there could be no "thought". M. Birch, in his article - "The relation of Previous Experience to Insightful Problem-Solving" - cited by Margaret Slauch in SCIENCE AND SOCIETY" (Spring 1947) - shows from his experiments on apes how previous experience with manipulating tools allowed the animals to solve problems involving the use of those tools which they could not solve without the previous apparently unconnected experience. The experiments of Maier on the utilisation of experience by both rats and human beings in solving

problems go to corroborate Birch, and show that memory, mechanical, but more especially organised memory - if one can separate them - is largely involved in intelligent behaviour. Hunter, much earlier on, had claimed to show that ability to learn a maze on the part of a rat and ease of conditioning were correlated.

Intelligent behaviour is closely allied, in human beings, with their ability to refrain from open trial and error behaviour and instead of actually manipulating the situation as an animal does, to "see" where the problem lies and take a short cut to the answer by an implicit trial-and-error of the various possibilities as presented by past experience or suggested by it. By using words and other symbols, we can save ourselves trouble both in remembering and in thinking. Words give structure and significance to the world we experience.

Warren, experimenting with various subjects on the learning of a finger-maze, found that those who verbalised and assigned numbers to the turns of the maze learnt most quickly and accurately; those who visualised came next; whilst the worst were those with motor or muscle imagery. Goldstein found that the difference between the behaviour of the normal person and that of the aphasiac with brain-damage is the "abstract" thought of the first and the enforced "concreteness" of action of the other. The aphasiac forgets persons and things and their names if they have no significance in the immediate situation. Indeed, his trouble is that his world loses "significance" for him, so that words and

symbols literally have no meaning and he "knows" objects only when he can manipulate and make use of them, while only the very simplest relations hold between them. Behaviour is thus much less efficient or intelligent. Findings in experiments on the forebrain of the chimpanzee by Jacobson and in psychosurgery by Freeman and Watts are that capacity to react to signals, to think abstractly, delay response and act with "forethought", is controlled by the action of the frontal lobes and that damage to these strikes directly at the ability to organise memory and purposive behaviour.

It would seem that there is something here like a physiological location and analogue of our "symbolic process". And since this process of thought is so closely connected with words and verbalism it is small wonder that tests for human intelligence should be so heavily weighted with the verbal factor.

One does not make human beings run mazes to test their intelligence, one asks them questions. Nearly all orthodox tests - whether individual or group - rely upon this verbal factor, inherent in human communication. Nevertheless, the avowed purpose of all such tests - the C A V D, Otis, Army Alpha, Cattell, is to test a type of behaviour which is not affected by verbal knowledge or by previous experience, in fact, pure "intelligence". The two demands made of any test ~~is~~ are that it should be:-

1. Reliable:
2. Valid.

Reliability of a test means its self-coherence, the fact that it correlates with itself when used over large populations or across

across periods of time. This suggests that the characteristic for which it tests is a real one and constant, and that the test tests this and nothing else.

Validity means that the test is of use and accurate in isolating real "intelligence" and correlates highly with other estimates of intelligence, e.g. school gradings, observer's estimates. With the question of the reliability of the test is closely bound up the "constancy" of I.Q. Intelligence tests have been used, even in experiments involving them, mainly predictively because they have been discovered to be a fairly accurate guide to what are desirable characteristics in our society, especially, for example, a guide to the desirable characteristics of grammar school pupils. This accuracy of prediction is founded upon the assumption, first that there is a constant and fairly measurible quality called intelligence and secondly that there is a characteristic and regular mental growth. These assumptions are fundamental andhave hardly been challenged since Binet first put them forward.

Looked at from the point of view of the experimenter and the instrument, as the scientist should look, the constancy of the I.Q. is no mysterious constancy of a mental characteristic, but the accuracy of the prediction, whatever it predicts, to the retest. W.S. Neff accumulates a great deal of evidence to show that the asserted constancy of intelligence is untrue. The Iowa studies showed that children moved from poor surroundings to orphanages or foster-homes gained up to 10 points in intelligence. The whole

conflict between "nature" and "nurture" is founded upon the assumption that intelligence is some unitary and static thing. But intelligence is defined only in terms of its test, and when constancy changes, is it the person or the test which is changing? The more valid a test is, according to the argument, the more reliable it should be and the less effect it should have upon experience. Yet the residual effect of testing is well-known, and it is easy to practice upon the usual tests of "pure intelligence". The most that we can allow is that the test isolates something that has a social value and that these results are valuable in being fairly accurate for predictive purposes. It was pointed out by Gray and Moshinsky in their study of social status and intelligence, that the higher intelligence of the progeny of teachers might be due to the fact that intelligence tests make use of the material likely to be encountered in a school curriculum, especially of vocabulary. E.L. Thorndike's finding that tests are valid because they correlate highly with school grades or teachers' estimates can be explained in the same way. The adherents of factor analysis have, abandoned the claim of a unique and uniquely measurable intelligence and content themselves with composing and working through batteries of tests in order to discover how selected mental factors are associated with one another. Thurstone, the proponent of the group and specific factor theory, at first even advocated the existence of something akin to the old mental faculties .

Thomson has not proceeded so far in replacing the "g" of Spearman with the overlap of several differently-weighted group factors. The argument for specific abilities and skills, distinct and apart from general intelligence, is corroborated by the evidence from industrial psychology. Where the evidence indicates that general intelligence is associated with these abilities, this may be so because the more symbolic ability allows its possessor to analyse and understand the way in which the skill may be built up and achieved. If there is a physiological substrate to intelligence, it might well be found in the organisational efficiency which is demonstrated in the working of the associational and symbolic processes. The argument for the physiological and hereditary nature of intelligence depends upon the obvious differentiation of abilities among people, the obvious mental inefficiency and failure to grasp the abstract shown by the moron or the idiot.

In using the term "intelligence" we are - as much as anything - using a socially-approving term, and are in danger of forgetting exactly what a hold-all phrase it is for the infinite variety of human ability and aptitude. It should rather be replaced by the word "cap city" which suggests that intelligence is something potential, not static or incapable of growth, as against "ability" tested by achievement and the "aptitude" so closely connected with interest and opportunity. Gordon, in his study of bargee children, showed exactly what a potential "intelligence" is, andhow dependent for its fostering upon the educational and social environment - falling rather than rising with age when nurture

nurture is repressive. In fact, bearing in mind the evolution of intelligent behaviour in men and animals, its definition as "ability to profit by experience", and its ground in the learning process, it would be safer and more honest to replace our term "intelligence" with the term suggested by R.A. Fisher - "educability".

"Educability" is a term relative to the society in which it is used. Our tests for educability will take account of what abilities it is desired to foster, what characters we wish to build and what purpose there is in and beyond education. The genetic argument does not lose its force, but we should like to know more of intelligence and educability before we allow that it is either a transmitted or an acquired characteristic. The statement that "intelligent" parents tend to produce "intelligent" children is valid, but what is not valid is the Pangloss-like assertion that therefore ability and intelligence will find their own proper level and that everything is as it should be in the social garden.

Intelligence may be a species of "g" factor or general group factor, dependent upon the way an organism is organised, and hence hereditary, but even so the progeny are unlikely to resemble their parents in specific aptitudes and abilities, on the Mendelian argument. The conception that they will therefore retain the same status in society is founded upon an aprioristic view of a class "professionalist" society - not upon any real understanding of the manifold aptitudes of human beings and the manifold needs of society. Secondly, the actual state of affairs is not at all as it is up

it is depicted by the apologists. The percentage of people who rise in society by intelligence proceed to absorb the professions and the means of education and to perpetuate their possession. Not so much is known of human genetics that any dogmatic pronouncements can be made upon the transmission of such a thing as intelligence.

Lastly, if intelligence demonstrates itself in purposive and directed behaviour and the ability to solve problems, why is this type of test never found among the scholastic examinations and attainment-tests which now purport to test ability and educability? The first requirement seems to be to make the test a real test of ability and character and make the problems more alive, immediate and significant to the child. Experience enters the test, so why cannot this be admitted and the ability to use recent or past experience also measured, along the lines of the experiments of Birch, Maier and Duncker on reasoning?

This essay on the nature of intelligence and intelligent behaviour has important practical consequences, illuminating our work. It will be seen that it is quite insufficient to hide behind the definition of intelligence with which we started this chapter - viz. that intelligence is that which is measured by intelligence tests. This hiding away from a qualitiative investigation of the complexus of the totality of personality leads to a position where the research worker is voluntarily drawing boundaries round the sides of his investigations where in point of fact no such boundaries do exist.

Dr. D.M. Fleming's contribution to the 1948 North of England Education Conference has very well summarised the evidence which, coming from a multiplicity of sources and directions, lays new emphasis upon the totality of personality and which relegates the evidence of intelligence tests from its all-dominating position in educational practice which it has gained in the last twenty years. The clinical use of the evidence of intelligence tests - as opposed to its "mass-production" use - will need to be discussed in a separate chapter.

Before this is done, however, it will be necessary to review the experimental work which has already been done on the constancy of the Intelligence Quotient and to that we now turn.

CHAPTER 5.

THE CONSTANCY OF THE I.Q.

A large number of investigations on the constancy of the I.Q. have been described in the Psychological Journals. These are mainly of one type.

A group of subjects was tested on one occasion with a particular battery of intelligence tests and at a later date these same subjects were again tested with the same or similar tests. The results are presented in a variety of forms - the commonest being a "reliability correlation coefficient" and the percentage of subjects who change one, two or more grades between the two testing periods.

We pass on to review in detail the evidence of experiments of this type.

The existence of positive correlation between test and re-test is now taken for granted. Most studies have tried to :-

(a) Get re-test information on younger and younger children:

(b) Get re-test data over longer periods between testing.

(c) Determine the effects on I.Q. constancy of particular types of environmental manipulation.

General Studies of Re-test Correlation in Adults:

Several studies of college students, to discover gains during years at college, have been reported. These agree that those who remain through a college course tend to gain in intelligence test score.

Masters, Schvals and Upshall report corr lations of .78 and .80 for two groups. Wolcott finds correlation of .81 between scores as freshmen and as seniors on Thorndike's test. McConnell reports correlation of .81 similarly on the American Council test. On the same test over a year Levesay reports a correlation of .88 for total score and correlations of ,69 to .80 for individuals selected. for special courses.

Two studies have been made of small groups of adults over ten-year intervals. Davidson reports a correlation of .89 for a group of 50 insurance company employees on a short group test. Garrison retested a group of students whose average at first testing was 25, using Yerkes Point Scale Revised Scoring. Correlations were .58 for a group of 32 men and .76 for a group of 41 women. "These studies", concludes Thorndike (Psychological Bulletin, 1940. P.167), "offer confirmatory evidence of ability of intelligence test performance in adults in fairly stable surroundings, even over a period of years".

It will be seen that here is a tendency to worship the correlation coefficient. What seems more interesting than the bulk correlation coefficients is what happened to those people (and even with correlations of .8 (there are many) whose scores dropped or climbed during the period. Can all the variance be attributed to test unreliability? Obviously not. Then if this is the result in "fairly stable conditions", there is still plenty of movement reported in these investigations. What would be interesting would be a further enquiry into the personal history of each to discover any changes.

Studies of School-Age Children:

Lauderbach & Hause. McCalls multi-mental test given to 150 pupils in grades 4-6 retested after an interval of 11 months. Retest correlation .791. Lincoln and Wadleigh gave the Otis Primary, National A, and Terman A, in successive years at approximately one year intervals. 154 children tested were in Grade 3 at time of first test. Median of all differences in I.Q. was reported as 7.29 points. 37% of changes were 10 points or more. When one considers the narrow margins allowed by the 11 plus Selection examinations and how many examinees would fall within a border of 7.29 points of the critical I.Q. the inefficiency of selection begins to be seen as alarming.

It is interesting here to stop to calculate how many children are involved under average conditions. In an area where 15 per cent of the junior school pupils are selected for grammar school entrance, the I.Q. critical point will be (1.04 x 15) = 115.6. If the average alteration of I.Q. in over 11 months is 7.29, it will be seen that, roughly speaking, half of those who lie between I.Q. 108 and I.Q. 115 will now be over I.Q.115, and presumably acceptable for the grammar school - whilst about half those of I.Q. between 115 and 123 will be now below I.Q.115, and therefore presumably below the standard required for the grammar school. From Guildford's "E" table it will be seen that 22.7 per cent of all 11 plus pupils will have I.Q.'s between these limits and therefore at least 11.3 per cent will have changed, in twelve months, their category of acceptable or not acceptable.

Certain other calculations, over longer periods, suggest themselves but will be left to a chapter at the end of this thesis. This one is introduced here to underline the seriousness of the problem.

Miller. Data reported by Hersh re-analysed, using only those cases who took all tests and correlating scores on successive tests into scores with the same mean and standard deviation. Median difference was found to increase progress only from 5.4 I.Q. points at 1-year interval to 7.3 points at 5-year interval. Corresponding correlation dropped from .85 to .79. It will be noticed that the contrasting of the two methods of reporting data, by median deviation of points and correlation coefficients, shows how correlation coefficients tend to obscure the issues.

Lincoln finds that when children are classed on basis of their initial test, those with high initial test scores tend to decline, and this would seem to throw some light on the bêtes-noir of the grammar schools.

Cattell finds that when children are classified on the basis of average of initial and final tests, children in the higher classification tend to have higher final tests than initial tests. These results illustrate - (1) The tendency of any fallible measure to regress to the mean; (2) the tendency of variability of I.Q.'s to increase with age, i.e. there is a greater likelihood of increase, say, of 10 points between ages 10-15 than 5-10.

Newyck analysed data by Carroll and Hollingworth and by Lawson and found correlation between Stanford-Binet test and

retest with interval of one or two years of .53 to .72. For data based on the Herring-Binet, correlation was .73. For stanford mean change ranged from 7.85 to 9.37 points and for Herring 9.06 points. Newyck concluded that the I.Q. for gifted children was more variable - which is in accord with other groups.

Retarded or Otherwise Handicapped Groups:

Studies by Arthur, Engel, Hoakley, Parker and Woodall on the mentally retarded confirm previous findings of a progressive decline in I.Q.'s of these groups.

Woodall finds an increase in I.Q. for retarded individuals after 16, indicating that mental growth has not stopped at that age.

Schott reports that adult neuro-psychiatric cases show markedlly greater variation between tests than do normal children. The writer has personally confirmed this result in Army work.

Miller suggests that a pseudo-intellectual deficiency may be produced by emotional maladjustment, remediable by careful treatment. Arthur finds that "an examination with Kuhlmann-Binet given by an experienced psychologist to a kindergarten, first or second grade child from non-English speaking home can yield a rating with a high degree of reliability and predictive value as measured by achievement in Kuhlmann-Anders test 5 to 7 years later if the child has had as much as a year in the English-speaking school environment".

Gilden, Macoubrey and O'Neill have endeavoured to analyse the factors associated with large changes in I.Q. between test and retest. Lack of a control group in one study and limited statistical analyses in both preclude any definite conclusions. Pattern of factors seem complex.

Predictive Value of Infant and Pre-School Tests:

Efforts made to develop tests for younger children have resulted in a growing crop of studies of constancy and predictive value of such indices. Evidence on the predictive value has been presented by Bayley, Cunningham, Drescoll, Muehlbein Gesell and others. Hallowell, Herring, Honzik, Hubbard, Kawin, Mowrer, Nelson and Richards, Stutsman, Lymmes, Updegroff and Welmann. These cover a number of different tests and other characteristics andd cannot be accurately generalised.

Here is a rough estimate of the results - taking the test and retest correlations from these studies where tests had been given at fairly definite specified ages, it was possible to tabulate them according to age at first test and interval between test and retest.

Average coefficient in each cell of the table was determined as a rough estimate of degree to which it is possible to predict on that interval at that age. Diversity in tests and variation in range of ability in group being studied must be borne in mind.

No effort was made to apply different weights to different sizes of population tested. It seems clear, however, that adequacy of prediction is combined function of age at which test

is given and length of time over which we try to predict.

TABLE.

Age at Earliest Test.	Interval between tests. 4/12.	4/9 12	10/15 12	16/21 12	22/29 12	30/41 12	42/53 12	Over 53 12
Under 4 Months.	.57	.33	.10	- .03	.09	-	-	-
4-9	.77	.51	.49	.23	.16	.46	-	-
10-15	.78	.66	.50	.45	.33	-	-	.55
16-21	.76	.68	.61	.44	.38	.41	.25	.33
22-29	.82	.74	.68	-	-	-	-	.43
30-41	.87	.68	.66	.49	.57	.57	.56	.66
42-53	.81	.65	.72	.71	.66	.63	.63	.41
54-65	-	-	.76	-	.73	-	-	-

Honzik: working with tests between ages of 21 and 84 months, suggested that size of correlation might be thought of as a linear function of age at time of first test divided by age at time of second test. She found a correlation of .92 for one group and .78 for another between this ratio and the correlation coefficient between test and retest.

Effect Upon Intelligence of Change in the Physical Environment:

Studiesin effect of diet upon learning and intelligence are reviewed by Fritz who concludes that contrary to what would be generally supposed, there is very little experimental evidence at present to indicate that diet markedly affects intelligence or capacity to learn.

Fritz indicates that the results of Maurer and Tsai showing effects of Vitamin B deficiency upon maze learning in rats represents exception to this generalisation.

Segal in London, however, got results with junior school pupils fed with milk, cod-liver oil and orange juice which are at variance with this generalisation.

Balkin and Maurer report a preliminary study of effect of increasing Vit. B. intake of malnourished children. A variety of mental tests were given to 46 children from homes of low economic status in which Vit. B. deficiency was probable. Children were given supplementary feedings of Vit. B. for 15 weeks, and tests

repeated. Gains were found in each test but were small. No control group to measure any practice effect.

Poull matched 41 pairs of seconds from testing at average age of 2/712, second at $4/4_{12}$, average I.Q. was 116 and 111.5 respectively. Correlation between the two tests with average interval of 21 months given as .56 . The Iowa studies of nursery school influences were very thorough.

Binet I.Q. of children from generally superior homes rose markedly during a period in nursery school, but did not rise during summer spent in normal home environment, but Merril-Palmer test results gave less marked results although they were in the same gener l direction.

Gains in Binet I.Q. were maintained by a sample of children located and retested after sever l years of attendance at other than University schools, and when these others were added to the sample the gains were increased. Length of attendance at University is related to intelligence test score in high school and college entrance.

These gains from nursery school education were not related to occupational level of the parent, but the greatest gains were by those who had the lowest original scores and the least gains were by those who had the highest original scores. In an orphanage nursery school, attendance of 200 days or more resulted in some gains of Binet I.Q., whereas a control group exposed to the general orphanage environment for this time showed some loss of I.Q.

Studies by Kawin and Hoeper failed to confirm gains found in Iowa studies, and were attributed by them to practice. But before the lasting rise in Binet I.Q. is assigned definitely to the nursery school experience, there are three other possible factors at work:

(1) Possibility of selection at work.

(2) Possibility that nursery school represents a short circuiting of gain that would have appeared in the years to come as a result of the superior home environment in a cumulative slow development.

(3) Possibility that the children from genetic background represented in these groups tend to do better on the type of intelligence test at the older ages than at the earlier.

No evidence is presented for these points of view and suggestions

Peterson's study, comparing small group of nursery school to other children as they enter and go through kindergarten shows that intellectual increments are not large.

Lamson endeavours to determine to what extent intelligence quotients are increased by children who participate in "vital curriculum", i.e. one which considers their interests and capacities, that requires activity and self-direction, making possible progress at the child's own optimum rate.

Hawthorne found that the average intelligence of pupils retarded in reading, and who improved in reading during remedial reading training at twice the normal rate, showed no corresponding gain in group intelligence test results.

Lowry, on the other hand, found that 50 children given three months' intensive reading drill which produced an average gain of 1.36 grades on 5 reading tests, and 2.72 grades in two speed-of reading tests, showed a gain from a pre-test on form A to a post-test on form B of Otis Intermediate Examination, of 11.76 points I.Q. No control groups for practice effect or possibility of inequality between tests were used, however. Scraggs also finds a gain in intelligence from special reading instruction, Grades of 5th grade negro children who followed an "intensive and extensive work in exercises in vocabulary development fact collection, recognition of central thought, organisation and summarisation, rate and speed of reading and in verbal manipulation" gained more than control group on a variety of verbal and non-verbal intelligence tests. They had a residual gain one and two years later. Durrell finds that children whose reading ability is better than would be expected from Binet I.Q. do better at paper and pencil tests than at Binet.

Conditions of Testing:

Madison and Jordan report reliability of Binet given by student examiners, test - re-test of .65 to .84.

Mayer and Rust study effect of negativism on the scores of young children.

Adkins, Dave and Snedden, report evidence of practice effects on repetition of test or another after brief interval.

Benton, Ferguson, Maller and Zubior, report studies with special incentives. No significant increase over control.

Cattell gives evidence of variation in standards between examiners differently trained.

Lodge reports evidence of a seasonal fluctuation in I.Q. the scores being higher between November 1st and April 30th than during the other half of year. Suggested that this might account for pre-school effect.

•

An important study has been made by Gray and Moshinsky in L. Hogben's POLITICAL ARITHMETIC. This shows some of the background to the 1944 Act with reference to secondary education for all. It shows that in spite of certain misgivings by certain grammar school teachers, the intelligence level as measured by intelligence tests is bound to rise as a result of the non-fee-paying regulation for Municipal Grammar Schools. On the other hand, the fact that so many grammar schools from experience report otherwise would seem to indicate that the introduction of this rule will require a concomitant change in incentives to work inside the grammar schools. These conclusions have, however, only indirect importance for our own research, but the articles are very valuable because of the importance of "social background" which is indicated here. The following summary of their conclusions is given because it illuminates later problems in this work.

Three general conclusions emerge from this study -

(a) We know that a minimum st ndard of educational equipment is and indispensable condition of entry into a large number of occupations. Lack of opportunities for higher education, imposed by the inadequacy of existing educational facilities upon the majority of children whose parents are manual workers, therefore constitutes a serious impediment to free movement in the labour market, and thus an important source of class stratification.

(b) The extent of educational opportunity afforded to the filial generation provides an objective criterion of social inequalities.

In this study they attempt to calculate quantitative indices of the extent of maladjustment of educational opportunity and educational ability at various social levels. They express the hope that some future investigation will accumulate data on a scale large enough to make possible the combination of occupations into wider social categories entirely on the basis of differences in amount of educational opportunity.

(c) The major part of these inequalities in educational opportunity remains after account has been taken of the relative ability of each social class. We are therefore dealing with disparities due to differences in social institutions rather than to genetic inequalities. The ratios are a measure of nurtural as contrasted with natural differences in an important domain of social organisation. Even so, they probably minimise the contribution of institutional agencies to differences in opportunity.

Further light on the direct problem of the Constancy of theI.Q. is given by the various sets of work to revise the Binet scales.

Bobertag[9] used 1908 Binet to study I.Q. constancy in 83 normal children finding a correlation of .95 $\pm$ .025 between test and retest with interval of one year.

Rosenov, using 1911 Binet, obtained r = .82 $\pm$.027 at 10.26 months interval for 69 cases. Cuff, with interval of 24 hours found r = .98 $\pm$.01.

Goodenenough: 300 children retested up to interval of six weeks on Kuhlmann 1922 revision. Range of deviations ~~were~~ was between+39 and - 21 points I.Q. 8.9 per cent gained 20 points or more, 4.9 lost 10 points or more. r for test and retest was .813 $\pm$ .012.

TABLE.

Age.	Boys.	Girls.	Mean I.Q. 1st test.	S.D. 1st.	Mean I.Q. 2nd Test.	S.D. 2nd.	Mean Algebraic Changes.
2	50	50	105.1	13.0	108.1	15.5	3.0
3	50	50	104.4	18.2	107.6	21.7	3.2
4	50	50	109.4	16.6	116.0	15.3	6.6
Total	150	150	106.3	16.2	110.6	18.1	4.3

Gray and Marsden, after making a few adaptations in Stanford-Binet for English children, made important studies on I.Q. constancy:-

TABLE.

Testings.	No.	Correlation.	Range of middle 50 per cent of differences.	Semi (Inter-quartile) ranges of changes.	Median I.Q. changes.	Interval of years.
1 & 2	100	.887	- 2.25 + 7.66	4.95	2.25	1
2 & 3	55	.908	- 3.03 + 3.0	3.01	●	1
1 & 3	63	.836	- 1.0 + 7.25	4.12	3.5	2
All.	218	.883	- 2.7 + 7.0	4.85	1.6	1 - 2
1 & 2	100	.883	- 2.25 + 7.7	5.0	2.25	1
4	371	.854				1 - 3
6	616	.851	- 6.1 + 4.7	5.5	3.3	1 - 5

At an interval of one year, Carroll and Hollingworth retested 52 gifted children 7-9 years at first test with Herring-Binet. Range of I.Q. changes was from +19 to -22. Average change regardless of sign was 9.06. r = .73 ±.044 . Notice low correlation and big range for gifted children.

Other Data:

Test.	Range I.Q.	Mean I.Q.	S.D. of I.Q.'s.
1	108 - 174	135.8	13.8
2	116 - 173	138.8	13.3

It is important for our purpose to give some account of experiments done on these differences in reliability of group tests and individual tests. It is to be noted that the results from studies concerning constancy of I.Q. of both types present a high degree of consistancy. As one method of comparing the results of

individual examination with those of group tests, reliability coefficients found by correlating test and retest I.Q.'s may be arranged in the following frequency distribution, as follows:

r's.	f (individual)			F (Group)		
.95 - .99	5			1		
.90 - 94	15			3		
.85 - 89	20	Median	.832	9	Median	.846
.80 - 84	23			6		
.75 - 79	12	Q.3.	.889	3	Q.3.	.885
.70 - 74	9	Q.1.	.76	4		
.65 - 69	8			0	Q.1.	.779
.60 - 64	3			1		
.55 - 59	1			0		
.50 - 54	1			0		
	97			27		

This represents a rather wide range of reliability coefficients for Stanford-Binet as well as for various group tests. The extremel low coefficients reported for the Stanford Binet were not found for unselected groups. In the last analysis, however, the magnitude of the reliability coefficient for group tests tends to be as high as for individual tests, casting doubt upon the conception of gross unreliability of group testing methods.

This covers the first method of determining the reliability of intelligence as measured by the standard group and individual intelligence tests.

The following chapter deals more specifically with an analysis of the defects of this method and an outline of the alternative approach - an approach which I have called the Clinical Approach.

CHAPTER 6.

THE CLINICAL APPROACH TO PROBLEMS OF INTELLIGENCE.

Up to thispoint we have been considering the problems of intelligence in general. In the essay on the nature of Intelligence we discussed intelligent behaviour ~~and~~ of animals and tried to draw some conclusions about the nature of intelligent behaviour in Man. We also discussed various theories which have been put forward to explain that which is measured by intelligence tests. But the I.Q., it has to be remembered, is a statistical concept. It does not describe any absolute quality of intelligence but simply describes the capabilities of an individual in terms of "average capability". It does not tell us how an individual behaves, but how much better he behaves than certain other people. True, this is valuable information. But there is ~~a~~ much lacking. The work of Koehler on the Intelligence of Apes is an example of the opposite of standard intelligence test methods for this is a purely qualitative, descriptive, deductive method.

To illustrate a point to which we shall return in greater detail later. Thefe have been many studies which purport to describe how intelligence develops and matures in childhood, adults and senescents. The standard method has been to give an intelligence test to many thousands of individuals from the earliest age they can do tests to the oldest age at which men and women are available. The age groups are weighted for (say) social status, if it is obvious

that any intellective selective forces has been in operation, and then I.Q. means for each age group are plotted on a graph. This shows a steady rise up to the ages of 13, 14, 15 or 16 - according to the nature of the test - and then a flattening off until the age of 40 or 50, when the graph slowly declines to old age.

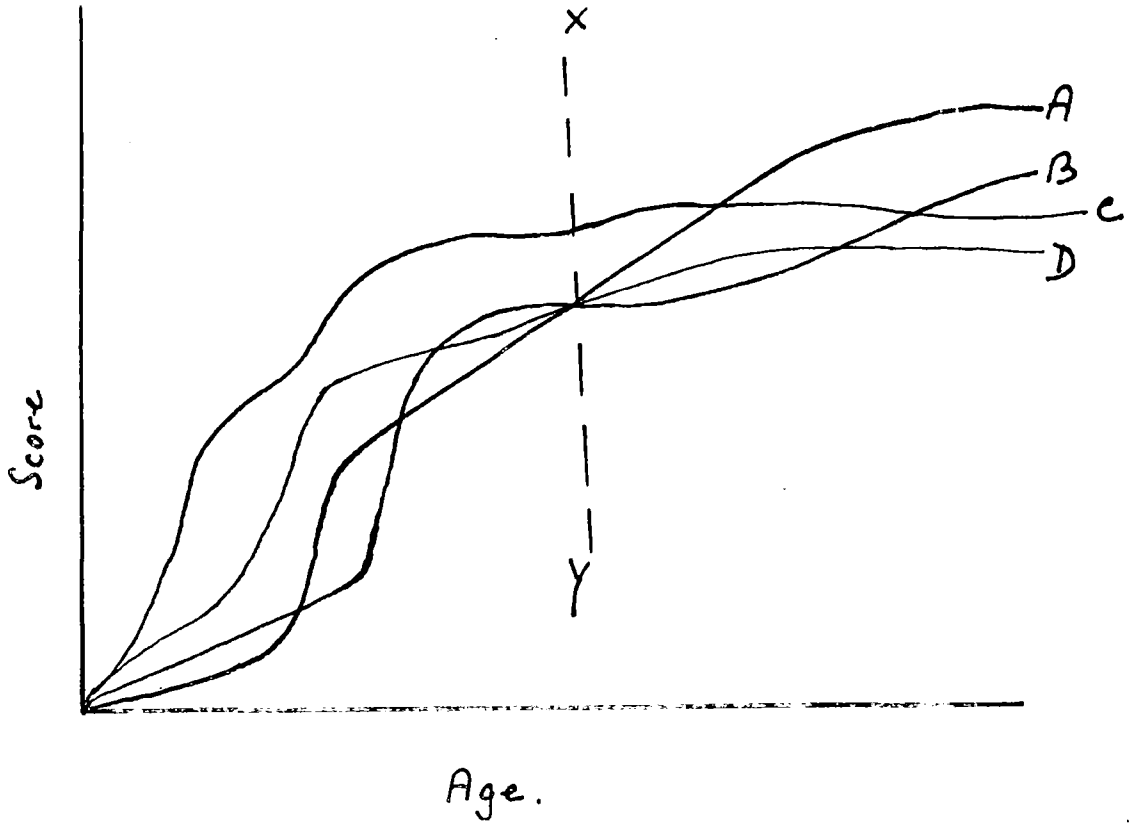
This graph, reproduced so many times, is partly responsible for what must be the biggest howler of psychometry - "Intelligence stops growing after the age of 16". But it says nothing about how individuals grow. Each point on the graph represents - not the intelligence of a certain boy or girl - but that of the "average" boy or girl. He or she is the statistical abstraction. He or she has not yet been discovered.

Now the evidence of test reliability and the influence of environmental changes upon I.Q. leads, on examination, to an acceptance of thepoint of view that intra-personal variability is much greater than the earlier psychometrical workers supposed. Of course, with such an "overall" quantity as intelligence, one wouldhave expected it. In the case of height and weight - which again are "overall" quantities, determined by many environmental and genetic factors, it is well known that individuals do not grow evenly. There is a smooth curve for average height which matches the average intelligence curve. But the individual height and weight charts show marked irregularities. It was obvious to psychological workers a few years ago that we needed individual intelligence charts.

Since this point is so important, let us carry the analogy of height measurements a little further. Suppose the only factor which determines entrance into the Grenadier Guards is the fact of being six feet in height. Suppose, too, everyone wanted to get into the Grenadier Guards on enlistment at the age of 18, but one had to be selected and given special training between 13 and 17 years of age in order to be selected. The Recruiting-Sergeant would be called upon to decide at age 13 if a boy was going to be over six foot in height at the age of 18. But practically no boys aged 13 are 6 feet. The very few that are are an easy problem for the Sergeant. They represent the "geniuses", the old "scholarship boys" of 1910, who used to be picked so easily by the examiners. They will not "shrink", at least. But everyone sees the difficulties of picking six-footers at 13. This is because everyone has seen dozens of "individual charts". We have watched John grow quickly, then stand still for years, then slowly to mature - whilst the pattern of Bill's growth has been very different.

Now individual charts of intelligence growth are much more difficult to make, especially when psychological research always seems to be in such a hurry.

The growth of individual intelligence has been studied most thoroughly during the last ten years by Dearborn. He has given a battery of intelligence tests every year for ten years and plotted the score of each individual child against his age. His findings are simple. The following chart illustrates a few standard patterns:



Here not onepattern of gro th is the same. If the line XY represents age 11, AB & D would have scored the same mark at age 11, but by the time adolescence has arrived they all three stand at quite different levels.

This information itself is very valuable, but even more important evidence has been accumulated, throwing light on these problems. It is evidence concerning the way in which certain children have been given really special attention and whose intelligence quotient has remarkably improved. In one case, a group of children of Intelligence Quotient = 70 - 75 who were classed near-defective, after special tuition in general school subjects over a year, all recorded I.Q.'s over 90 at the end of the year, and one recorded an I.Q. of 140. This suggests a measure of plasticity such that the early workers little dreamed.

It would seem that intra-personal variability has two sides. Absolute level (maximum) of ability certainly has some genetic background, though it is not the whole story. Similarly, the pattern of ability growth is determined to some extent by genetic factors and to some extent by the influence of environment. It has to berecognised that the child is not simply an individual with fixed abilities and with interests and aptitudes which arise simply out of an internal complexus of the individual himself. The individual is at one and the same time a social being. He is only an individual in his relation to society. Similarly, society is not simply the outside, overbearing agent pressing ruthlessly, moulding the individual to a pattern inevitably determined by the stresses of that environment, but the individual makes society as much as society makes the individual. This is not so much a "two-dimensional concept", as described by Dr. C.M. Fleming, but a single contradictory, inter-penetrating interplay of two opposites which resolves itself into the developing wholeness of personality.

For this reason the main researches of factorial analysis with its discoveries of special aptitudes are only valid insofar as they wish to group and classify tests rather than persons. Most recent work, for example, which tends to show the transitory ephemeral nature of the "K" and "M" factors, isolated by tests of mechanical aptitude at 11 plus, could be explained simply from this point of view.

It is, in fact, not sufficient in psychological science to be purely emperical in one's approach to problems of this sort. A study of the theoretical implications before practical work is started would help to make the work of many psychological researches much more fruitful.

The clinical view of estimating intelligence then would look at intellective behaviour from a multiplicity of viewpoints. It is not so much a question of regarding "personality" judgments as important as seeing that it is impossible to differentiate between the two sides ofpersonality, except as a technique of analysis which can only be given significance when an integrated view of the total personality has been made.

No single procedure then is valid for estimating the quality and level of intelligence in an individual. Test results are only one part of the required information. More, information must be of a "historical" kind; it must not only show what the child is but what it has been. It must not only show where the child is, but where it has been. Only then, by a judgment which is best made by a group of investigators who "socialise" their judgments rather than "pool" them, can a useful estimate of the child be made.

Dr. C.M. Fleming sums this up in a report she gave to the North of England Education Conference, 1948. - "Nineteenth century workers concentrated on the observation of individuals and on a study of means by which they could be classified. Twentieth century evidence presents a challenge to thevery act of classification, and it now seems necessary to say that since human beings are social in their nature and react always to a total situation, there is no way of adjusting schooling to the needs of the children, except through the provision of continuous and many-side stimulation (preferbly in a co-educational setting) accompanied by continued study of the achievements, interests and social reactions of pupils as they develop through childhood and adolescence to maturity".

CHAPTER 7.

THE SPECIFIC PROBLEM STATED & THE RESEARCH OUTLINED.

In the last chapter it was demonstrated how the question of intra-personal variability was an important factor in the "efficiency" of eleven plus selection. If it is the design of the administrators to produce as many grammar school successes as limited resources will permit (and to may the administrator's job appears to be just that) he will be interested in two quantities: (a) How many who failed at 11 would, if they had been admitted to a grammar school, have "passed" at 16; (b) How many who passed at 11, "failed" at 16. The work of MacLelland is work which is directed towards the discovery of the proportions of these groups with many different types of "test", which include:-

(a) ordinary examinations:
(b) teachers' estimates:
(c) group intelligence tests.

MacLelland has almost completely exhausted this line. His conclusions are summarised here: "It is estimated that the percentage of the qualifying group who had the ability and attainment necessary for success in a senior secondary course is estimated as 15.4 (16.4 per cent of total population). The percentage who were actually successful is 10.5". "Any prediction of success based on I.Q. alone is most uncertain. Pupils with very low I.Q.'s may pass. Pupils with very high I.Q.'s may fail".

It will be seen that MacLelland's study uses the criterion of success, the judged ability to profit by the given secondary school

course. His problem was to discover what proportion of the 11 plus population had at eleven abilities and aptitudes which would enable them, ceteris paribus, to successfully absorb the instruction given at a grammar type of school (roughly grammar type), and what proportion actually succeeded. His wastage of 4.9 per cent is one of the quantities which interests the administrators. The other quantity, they hope, is made negligible by the 13 plus Transfer Schemes now in universal operation.

But from the point of view of satisfying social needs this sort of answer is not sufficient. Undoubtedly society needs today the largest possible number of really able men and women with an educational background which will fit them to manage, operate and extend the complex manufacturing and administrative concerns of present-day society. One cannot therefore be content with discovering techniques of selection which simply cut losses, under the present methods, to a minimum. There are a whole host of related problems.

(1) If the methods of study and the content of the curricula of present-day secondary schools were altered, it may be that many more 11 plus pupils would have the qualities necessary to be successful in the more advanced socio-technico studies.

(2) If the total complex environment of children in their former years could be altered, this might increase the proportion of pupils who at 11 plus would show capacity to profit by the advanced studies of the grammar schools (or the advanced courses which answers to (1) would produce).

The former is the research ground of the Educationalist whose research lies in the field of curriculum development. He will need to remember the special problems of pupil selection in general and especially the effects of research on problems suggested in paragraph (2).

We have ranged over the problems of eleven plus selection before stating the purpose of our particular research in detail because, in social studies of this sort, it is especially important to unearth and examine all related factors. The day of the one-track research worker in the exact sciences like physics is over today. Such one-track workers in the field of education never had a day. Much research of a painstaking nature has proved sterile because it failed to observe the importance of correlated factors.

Let us enumerate our discoveries to date:

(1) The problems of 11 plus selection for grammar school education have a history which is rooted in the social history of our century. Any answer to the problems therefore needs to take into account the necessary features in the future development of society.

(2) The actual practice by L.E.A.'s of selection for grammar school education follows the general pattern of the prediction of ability at 15 or 16 by means of written tests given at age of 10. No L.E.A. up to the present has succeeded in finding a satisfactory method of applying the firm recommendation of the Norwood Report to use, as the main selection agency,

teachers' records of the children's abilities, aptitudes and personality in the junior school, supplemented by intelligence tests.

(3) That further examination of the meaning of the word intelligence, which is loosely applied to the ability measured by intelligence tests, shows it to be, on the whole, a socially-approving term. The word "intelligence" covers so many types of ability that it is for scientific purposes best dropped altogether. We continue, wrongly, to use the term to cover that which is measured by intelligence tests. We also come to the conclusion that to regard intelligent behaviour as behaviour governed and directed by a sector of the mind, leads to incorrect conclusions. Behaviour has always to be determined as the total reaction of the total personality. Thus to attempt to build selection instruments on th assumption that those children with "intelligence" are those most likely to succeed in grammar schools, flies in the face of all psychological experience. It is one thing to say that those who do welll in "intelligence" tests at 11 plus also tend to do well at gr mmar schools. It is quite another thing to say that intelligent children succeed at grammar schools and unintelligent children do not. On the contrary one might, to take one socially-approved definition s y that those who do succeed at grammar schools are intelligent.

(4) We then surveyed the fields of research in detail dealing with the constancy of the I.Q. We showed that the "mass methods" demonstrated that it was harder to predict intelligence from one set of results the longer the pe iod ahead one was trying to predict.

This was relatively independent of the type of test. This proves that the longer other factors are allowed to react on the individual the more those other factors had their effects. Some of the"other factors" seemed to be little dependent upon such direct things as food. The most important factor appears to be the nature of the social environment - the quality of friends, acquaintenances, and especially guardians. Thus 11 plus selection by the present methods presupposes that nothing can be done about the changing environment of the child. It views "efficiency" from the viewpoint of existing society only. "Losses are bound to occur" - "it is the job of the selector to see that his methods make losses a minimum".

It is probable that present techniques have actually reached their maximum percentage efficiency. Different prodedures seem to produce much the same results. It would appear, therefore, that it would be more productive to tackle the problems of the scarcity of ability from one or the other angles suggested above. We have set ourselves the task of contributing towards the second approach - Can the number of able children who are capable of beinggiven an advanced education to man the many new administrative and technical posts required by modern society, be increased by altering the total environment of the growing child?

Each determinant in use in selection procedures at present in use in selection needs to be investigated in turn. We start at the simplest and probably the most fundamental - intelligence.

Specifically, our problem can be stated as follows:

Children differ in their rates of maturation. This the evidence of test reliability and the studies of Dearborn show. Can any factors which are responsible for these differing rates be isolated? These factors are either specific inherent factors or factors of environment. If they are not one they are the other. It will be easier to design an experiment to discover environmental influences and to attribute the variance not accounted for by these to inherent factors.

The experimental work which we now describe is a first contribution to this problem. A later research is planned on a larger scale to confirm the findings of this first experiment and to widen the field of investigation. A parallel research has just been started - at the suggestion of the present writer - to investigate by similar methods the cases where there are significant discrepancies between the intelligence test level and the level of performance after one year in a grammar school.

CHAPTER 8.

THE METHOD IN DETAIL.

The method of the experiment may be briefly described as follows:

A. A battery ofintelligence tests was prepared. This battery was constructed to give also measurements of "V", the verbal factor, and "K", the spatial relations factor.

B. This battery of tests was given to boys of eleven who had been in a grammar school one term. It was given again to the same boys exactly twelve months later.

C. A number of boys were selected from the remainder as having shown significant increases or decreases in their scores when allowance was made for the general increase of ability during the twelve months.

D. The environmental, school and home, background of the boys was investigated to discover any factors which could account for the change in intelligence level.

We will now proceed to give the details of preparation and collection of the data.

A. The Battery of Intelligence Tests:

All these tests were constructed by the present writer to serve the special purpose of this investigation. In preparing the tests three aims were held in view. -

(a) The tests should be valid and reliable:

(b) They should be highly saturated with a "g" factor and be capable of grouping into two main classes; tests to measure g and V, and tests to measure g and K:

(c) They should give a wide spread of marks over the highly selected population which constitutes a "year" of a grammar school.

In each case the standard pattern of test construction was followed. A large number of items for a particular test was prepared and these were given to colleagues on the staff of a school to comment upon. All doubtful items were eliminated and eventually a test which was twice the length aimed at for the final test. This test was then given to a "year" of a grammar school, and a total test mark for each individual was found. The orderof merit list was divided into halves, 50 per cent who were "good" on this test and 50 per cent who were "bad". A tetrachoric of the following percentage form was then prepared for each item:

ITEM:

	Pass	Fail	
Good	20	30	50
Bad	35	15	50
	55	45	100

Tetrachoric correlations to be regarded as validity coefficients were then calculated for each item. Since the number of coefficients required numbered many hundreds and the labour of calculating so many tetrachoric correlations could have been prohibitive, a special table was prepared giving the tetrachoric correlation for all

possible percentage tetrachoric tables within the constraint of the criterion being split on the 50 per cent "good"/"bad" line. This table, which has been accepted for publication by Professor Thomson, editor of the new "Journal of Psychological Statistics", eventually saved a great deal of time. It is to be hoped that it proves useful to many other research workers in the field of test, questionnaire and inventory construction. A copy is appended to the present thesis.

At the same time an order of difficulty was drawn up for the items of the test. From the data, validity coefficient and order of difficulty items were selected to give high validity and a wide distribution of marks. Selection to give high validity, of course, is not straightforward nor easy. For the greatest usefulness in contributing towards high validity of the test as a whole, test items should have a high correlation with the criterion and low correlation with the other items. To follow this to its logical conclusion a factorial analysis of a matrix of inter-correlation of every item would need to be done. But this work would be prohibitive. The simplest method recommended by Guildford in "Psychometric Methods", was adopted. All items with validity below .25 were neglected. This was accurate enough for our purpose. In order to select items to give a wide distribution with the selected population, items which had less than 5 per cent or more than 90 per cent passed were also rejected. A final test was then constructed.

The details of the construction of the most important non-verbal test - the Figure Reasoning Test - are given below to

illustrate the work done on all the tests.

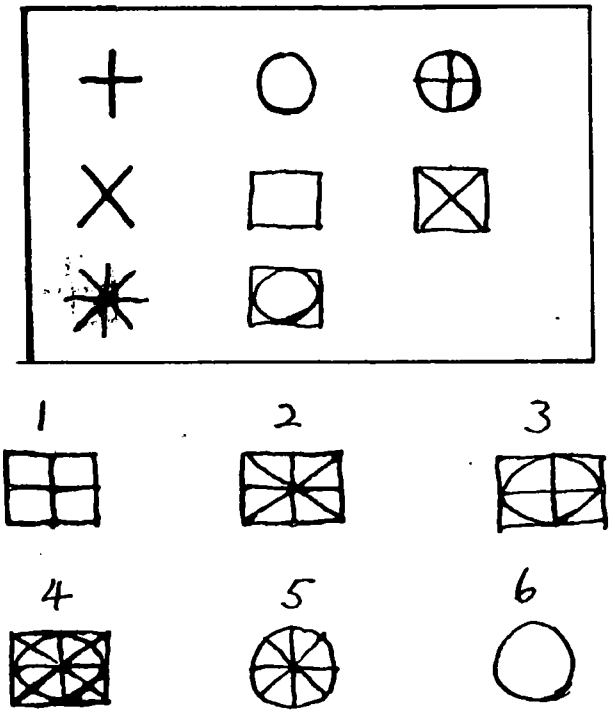
A hundred items were prepared and reduced to 90 by individual rejection of doubtful items. Validity coefficients were calculated together with an order of difficulty. The table given below summarises the qualities of the selected 45 items:

Item No.	Validity Coefficient.	Percentage Passed.
1.	.32	89
2.	.28	87
3.	.34	84
4.	.61	84
5.	.25	83
6.	.37	81
7.	.60	80
8.	.41	79
9.	.37	76
10.	.42	76
11.	.31	72
12.	.68	70
13.	.57	66
14.	.61	65
15.	.34	64
16.	.27	62
17.	.64	62
18.	.41	58
19.	.44	54
20.	.34	54
21.	.27	50
22.	.28	50
23.	.31	49
24.	.30	49
25.	.30	45
26.	.61	41
27.	.34	42
28.	.31	40
29.	.33	36
30.	.38	36
31.	.40	36
32.	.32	29
33.	.29	28
34.	.37	26
35.	.41	20

Item No.	Validity Coefficient.	Percentage Passed.
40.	.28	18
41.	.29	11
42.	.25	10
43	.26	8
44.	.25	5
45.	.27	5

The tests, constructed in this way for this purpose, are now described.

Test 1: The Figure Reasoning Test: This is the test whose construction is described above. This consists of a square containing eight figures in three rows of three with a figure to complete the third row and third column missing. The answer is required to be written on a separate answer sheet. Here is one example. A copy of the 45 items and six examples are appended. -



The subject is required to choose one of the six figures given below the design which he feels completes it. He is given instructions as to how to solve this type of question in sixe examples. These examples cover all the main reasoning processes required to solve the items and thus the test is in part a test of the ability to learn a new task, to grasp its significance, and apply his new-found knowledge and skill. It is essentially a learni ng test as well as a reasoning test for the items in it have no absolute order of difficulty in the same way as has Kohs Blocks designs. Harder items are easier because of the fact that the subjects have learned the methods of solution by dint of solving the previous item. It will be seen that the items of this test are not arranged in strict order of difficulty because this proved to be impossible. Each time the order is changed the difficulty is changed. At some points, therefore, a "logical" order was used as the criterion.

This test would be expected to show a very high "g" factor and some "K" factor.

The form of the test is, in some ways, like Ravens' Progressive Matrices 1938 Test, but the present author, in war-time collaboratio n with Raven, developed independent theories of matrices constructions which abandoned, on theoretical grounds, the "wall-paper" type of item.

Test 2: Vocabulary Test: This test was the familiar synonyms test but was constructed by going back, in the first instance, to the Binet method of selecting a sample of words. The first word of eacn

fifth page of an English dictionary was chosen. This comprised 123 words. A number of these were eliminated because of the difficulty of securing suitable synonyms, and the remaining 80 were examined by the same general technique as described above. Here is a copy ofthe final test:-

END.	FAME.	CONCENTRATE.
sharp. place.	success. renown.	divide. boil.
exit. finish.	bravery. kingly.	impose. centralise.
GEM.	OPPOSE.	DEFER.
jewel. portrait.	prevent. put.	conquer. suspect.
paper. plant.	assist. resist.	judge. delay.
ENEMY.	DECLINE.	DEVIATE.
fighter. foe.	improve. mismanage.	agree. swerve.
foreigner. tank.	deteriorate. divert.	suspend. pretend.
FAIR.	PROBABLE.	CAVITY.
kind. rich.	simply. eventful.	serious. hollow.
quiet. just.	likely. successful.	safety. judge.
CREST.	FRAGMENT.	ELECT.
hat. top.	rock. piece.	tall. govern.
plant. brush.	crack. cut.	find. choose.
CONFINE.	SPECIMEN.	IRRITATE.
imprison. ravage.	peculiar. sole.	destroy. annoy.
decreel dig.	sample. sequence.	hurt. scold.
MAIM.	MUTINY.	CONSOLE.
injure. kill.	rebellion. retort.	pulsate. soothe.
execute. operate.	battle. rudeness.	annex. treat.
CONSUME.	DEFEND.	FRIGHT.
expend. drag.	guard. destroy.	noise. cry.
fill. detach.	short. rely.	fear. hurt.
LUBRICATE.	DETAIN.	DESPICABLE.
plane. burn.	journey. hold.	profitable. indolent.
oil. speed.	pull. get.	puerile. contemptible.
DELICATE.	GAOL.	IMITATE.
sensitive. pretty.	shop. prison.	copy. amuse.
smooth. sweet.	bird. score.	chap. dishonest.
FAULT.	ALIEN.	IMPLORE.
flaw. lose.	foreign. dirty.	tall. retreat.
broken. spite.	black. dwell.	give. beseech.

HUE.		PROVOKE.		INFIDEL.	
colour.	thief.	defend.	fight.	unbeliever.	murderer.
tree.	boy.	burn.	challenge.	negro.	knight.

COMMENCE.		LANGUISH.		EXERTION.	
please.	dream.	torture.	pine.	sweat.	heavy.
begin.	fall.	smart.	pain.	fear.	effort.

MODEL.		PREVAIL.		ILLUSION.	
photograph.	wood.	overcome.	waste.	picture.	evasion.
pattern.	utensil.	work.	usual.	belief.	deception.

LASH.		NARRATE.		CONTAGIOUS.	
condemn.	punish.	print.	lie.	happy.	fatal.
whip.	break.	escape.	relate.	catching.	serious.

PORTION.		DURABLE.		PERIPHERY.	
number.	part.	soluble.	hot.	area.	weight.
fruit.	Gum.	lasting.	expensive.	contents.	boundary.

This test, it would be expected, measures "g" and "V". It should be noted that a vocabulary test measures not so much intelligence <u>at the moment of taking the test</u>, as the total effect of intelligence over a number of years. Babock gives a great deal of evidence to show that people whose brains have suffered deterioration or mutilation (e.g. in G.P.I.), have a higher vocabulary score than a "reasoning" score. The difference between these is taken to be a measure of mental deterioration. Similarly for our purpose, vocabulary level measures much more directly than any other test the way in which a child has been able to use a "favoured" background and intelligence to increase the range of his vocabulary.

<u>TEST 3. Series Test</u>: This is a version of the Shipley reasoning test and uses one or two of its items. It is essentially a test of mental flexibility. Close examination of the items will show that

if one attempts to solve an item by the method employed in the previous questions, little success would be gained. High perseveration would be ~~at~~ a disadvantage. In this respect the test is the opposite of the matrices test. It uses both word series and numerical series and it would be expected to measure g plus V plus n but for our purpose we shall neglect the n factor, since no other test is designed to include many significantly numerical examples: The final 40-item test used was as follows:-

Example:	A	B	C	D	.
	2	4	6	8	. .

Each dot represents a figure or a letter:

1. full empty hard soft heavy . . . . .
2. 16 14 12 10 .
3. bread eat water . . . . .
4. z y x w v u .
5. arm hand finger leg foot ...
6. spend lend face mace ten p..
7. chain yard foot . . . .
8. escape scape cape ...
9. Sunday Tuesday Thursday Saturday . . . . . .
10. late ate tall all chair hair snap . . .
11. today yesterday Monday . . . . . .
12. A Z B Y C X D .
13. 2/6 (30) 3/2 (38) 5/4 (64) 1/4 (. .)
14. 12321 23432 34543 456 . .
15. mist is wasp as pint in tone . .
16. 614 416 268 862 943 . . .
17. steep e rare r fools o summer .
18. plate cup saucer crockery
    bread meat egg . . . .
19. grandfather father . . .
20. taste tongue see . . .
21. E/W NE/SW SE/NW S/ .
22. board bad start sat spent . . .
23. 10 16 18 36
    2 4 6 4
    5 4 3 ,
24. 1 mutton 2 pork 3 beef
    (.) pig (,) cow (.) sheep.

25. pull drag (S) tall short (O) hot cold (.)
26. tot tot bard drab 537 . . .
27. 6 7 9 12 . .
28. 30 = 2/6 (Y) 18 = 1/6 (Y) 20 = 1/10 (N) 36 = 3/4 (.)
29. 1 4 9 16 . .
30. foot inch shilling penny 12 .
31. kcftd dkcft tdkcf . . . . .
32. ewe (Y) drake (Z) man (Z) woman (Y) bull (.)
33. another 1234567 roan 7312 ton . . .
34. lag leg pen pin big bog rob . . .
35. tar pitch throw corridor passage voyage recline ... falsehood
36. tar tas fib fid fot fow hip ...
37. 1 3 7 15 31 ..
38. two w four r one o three .
39. 3214 82 73 154 55 22.
40. amet team poson spoon sole lose banule . . . . . .

TEST 4: "Always" Test: This is a test of familiar pattern. The subject is required to underline the word which represents what a given object "always has".

Example: A CAT always has: FUR: Kittens: ball: mouth. This test is verbal in form and measures the ability to recognise essentia. In effect it is another form of classification or sorting test. The fundamental nature of this aspect of intelligence secures the test a valuable place in the battery. Again, work on persons suffering from brain injury has shown that the inability to distinguish between superficial and fundamental qualities is a measure of the extent of damage to the brain surface. It would appear that classification tests, measure the mass action of Lashley. The Vigotsky sorting test and the Trist-Hargreaves sorting test are the more sensitive instruments for measuring what is measured here in Tests 4 and 5. This test measures g and V.

The final form of the test as given is as follows:

RIVER.
boats. fishes.
logs. water.

CAT.
ball. fur.
kittens. ribbons.

HOUSE.
roof. fireplace.
paint. pordh.

SLEDGE.
bells. horses.
runner. ropes.

DOG.
kennel. master.
licence. nose.

TREE.
leaves. roots.
fruit. nuts.

KITCHEN.
cupboard. mirror.
table. floor.

MOTOR CAR.
cushions. speedometer.
wheels. windows.

LIBRARY.
books. chairs.
magazines. tables.

SHOE.
buttons. sole.
laces. lining.

FIRE.
coal. wood.
heat. paper.

SNAKE.
legs. poison.
rattles. skin.

HILL.
birds. brooks.
height. grass.

CHIMNEY.
bricks. opening.
ivy. smoke.

DOOR.
hinges. knocker.
lock. top.

GARDEN.
plants. flowers.
weeds. vegetables.

SHIP.
sails. engine.
stern. anchor.

FISH.
scales. speckles.
spines. stomach.

LEOPARD.
mate. cubs.
prey. spots.

KNIFE.
blade. handle.
sharpness. owner.

BOOK.
print. title.
author. pages.

PARROT.
sppech. cage.
feathers. food.

SHOP.
assistants. food.
cigarettes. commodities

TRUNK.
lock. handles.
weight. strap.

GROTTO.
cavity. pools.
fissures. darkness.

RESERVOIR.
fishes. contents.
sand. fence.

ROW-BOAT.
oars. keel.
bottom. rudder.

CLOTHING.
warmth. wool.
buttons. texture.

TOWN.
people. town hall.
church. school.

ANIMALS.
life. offspring.
tail. eyesight.

CINEMA.
seats. box-office.
attendants. screen.

RADIO.
loudspeaker. dial.
case. wires.

TABLE.
surface. legs.
polish. wood.

TEACHER.
school. pupil.
books. wages.

ROAD.
kerb. direction.
stones. lamp-posts.

ENVELOPE.
flap. stamp.
address. gum.

CUP.
handle. glaze.
contents. shape.

PICTURE.
subject. paint.
figures. beauty.

DOCTOR.
medicine. bag.
knowledge. smile.

BOX.		RAILWAYS.		HAPPINESS.	
cover.	sides.	track.	station.	satisfaction.	wealth.
hinges.	nails.	signals.	locomotive	laughter.	comfort.

SCHOOL.		WOOD.		PURSE.	
teachers	blackboards.	knots.	grain.	money.	leather.
desk.	maps.	pitch.	bark.	capacity.	clip.

TEST NO. 5: Classficiation Test: Here is a straightforward classification test in which the subject is required to select one thing out of four which is different from the remaining three. Closer inspection of the items reveals that in the majority of them, the four things may be grouped into at least two pairs - sometimes more. Thus to observe the rule of excluding only one, all superficial qualities have to be rejected and the attention of the subject is directed specifically to the essence of the likeness.

This test also measures g plus v. The few numerical examples are of insufficient importance to contribute much variance to a correlational matrix:

EXAMPLES: (a) knife table (b) carpet sideboard.
fork spoon table chair.

- - - - - -

cabbage	bacon	square	circle	petrol	coal
coal	beef	triangle	oblong.	wood	coke
pen	chalk	yacht	battleship	plough	mower
paper	crayon	rifle	sword	harrow	tractor
house	bungalow	garden	field	ostrich	hawk
cottage	caravan	cemetery	church	eagle	stork

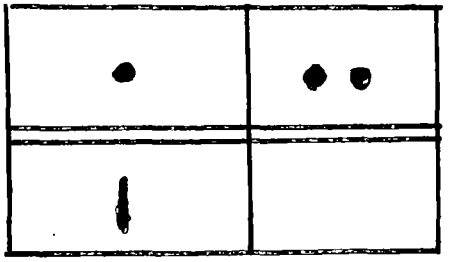
horse goat	robin giraffe	smell sight	taste colour	photograph record	book speech
skirt frock	coat shoe	teacher carpenter	lawyer doctor	cup pipe	tunnel chimney
tree snake	ball bird	pound shilling	onnce ton	factory mine	farm cinema
cup plate	saucer spoon	2 6 3 9		3 7 5 9	
hill river	valley plain	violin piano	harp trumpet	train bus	motor-car bicycle
beef bread	mutton pork	statue rock	poem song	newspaper radio	telephone magazine.
4 6 9 8		cabbages carrots	turnips potatoes	iron wood	vinegar alcohol
wool fur	flax cotton	8 27 18 9		pencil worm	tube cow
grass fly	fish man	hour week	minute second	justice wealth	kindness mercy
strawberry damson	plum apple	glass steel	air water	red blue	purple yellow
				18 7 16 13	

TEST NO. 6: Figure Classification Test: This is another non-verbal test of the classification type. Its items are constructed with the same basic principles in mind as test No. 5 but with purely non-verbal, non-mumerical examples. The test would be expected to measure g plus K. A copy of the whole test is appended but an example is quoted below:

(Underline the one different from the rest).

TEST NO. 7. Figure Analogies Test: This is a transposition of the familiar type of verbal analogies test into purely non-verbal form. Again this measures a skill of observing that which is essentially the same and that which is different in two figures and so building a fourth that it has the same essential likeness and differences as a third figure. Notice that it is also of the "Recall" type - i.e. the subjects are not asked to select from possible answers but to picture the required figure for themselves and to draw it. In this it ressembles the series test.

This test would be expected to measure V and K. An example is quoted below and a copy of the full test appended:



These are all the subtests used in the battery of Intelligence tests.

The tests were built up and standardised using a grammar school population first to fourth year in another school other than where the main investigation was made. Summarising the general properties of the tests before any statistical analysis of them is made, we find the following picture:

Test No.	Name.	Time.	Factors Measured.
1.	Figure Reasoning.	20 mins.	g and K.
2.	Vocabulary.	10 mins.	g and V.
3.	Series.	15 mins.	g and V.
4.	"Always".	7 mins	g and V
5.	Classfication.	7 mins.	g and V.
6.	Figure Classification	7 mins.	g and K.
7.	Figure Analogies	10 mins.	g and K.

From such a battery, three separate sets of measurements can be made for each individual. -

(a) A measure of the "g" ability of the individual by some form of control of all seven tests:

(b) A measure of "K" by some form of combination of tests 1, 6 and 7.

(c) A measure of V by some form of combination of tests 2, 3, 4 and 5.

There are several ways of combining these tests to give the best measurements of g, V and K. They could be weighted according to the time spent on them, according to the number of items, according to their reliability (e.g. split-half reliability), according to their loadings in the factors which it is required to measure, and, finally, they could be weighted equally.

In order to make a correct decision on the best method to adopt, a full statistical analysis of the tests needs to be made. To such an analysis we now turn.

It was first necessary to examine the internal consistency of the new battery of tests. This is given by the split-half reliability of the tests. Each split-half correlation was corrected by the Spearman-Brown formula and gives the following results. The marks used in this analysis were those obtained by the 120 pupils used in the main investigation on the first application of the test.

Test 1.	Figure Reasoning Test.	45 items.	Split-half reliability)	.95
Test 2.	Vocabulary Test.	48 "	"	.94
Test 3.	Series Test .	40 "	"	.90
Test 4.	Always Test.	45 "	"	.95
Test 5.	Classification Test.	40 "	"	.95
Test 6.	Figure Analogies Test.	24 "	"	.91
Test 7.	Figure Classification.	20 "		.89

Next, another "test" result was added in the form of a scholastic examination mark. At the end of their first term in the school these pupils were given an intensive written examination in the subjects of their curriculum. Percentage marks were given by the subject master and a total scholastic mark obtained simply by adding these raw percentage marks. In spite of the crudity of this marking system, the raw total ~~shows~~ serves as a useful estimate of

attainment in school subjects. Thus, in all there are eight sets of marks. Inter-correlations of these eight tests were then calculated. In each case it was of course possible to calculate a product-moment correlation, and this was therefore done.

The following table gives the resulting matrix of correlations. Test 8 stands for the examination mark. The table is rearranged to put the tests in hierarchical order. -

Test.	2.	4.	7.	6.	1.	3.	5.	8.
2.	x	.632	.670	.605	.659	.632	.527	.291
4.	.632	x	.645	.580	.678	.577	.442	.298
7.	.670	.645	x	.644	.676	.560	.368	.194
6.	.605	.580	.644	x	.646	.551	.401	.262
1.	.659	.678	.676	.646	X	.421	.296	.114
3.	.632	.577	.560	.551	.421	x	.455	.230
5.	.527	.442	.368	.401	.296	.455	x	.444
8.	.291	.298	.194	.262	.114	.230	.444	x

In order that the names and numbers of the tests should not be confused, the tests are re-grouped as follow:

Test A. Vocabulary Test.
" B. Always Test.
" C. Figure Classification Test.
" D. Figure Analogies Test.
" E. Figure Reasoning Test.
" F. Series Test.
" G. Classification Test.
" H. Examination Mark.

This matrix was analysed by the Lawley's method of Maximum Likelihood for the main reason that it provides a valuable exercise in this little-used method and in spite of its length, where accurate results are required on which a test of goodness of fit can be made this method seems to be well worth the trouble. In order to illustrate the method the results are given in detail.

INTERCORRELATIONAL MATRIX.

-	A.	B.	C.	D.	E.	F.	G.	H.	
A.	.	.632	.670	.605	.659	.632	.527	.291	
B.	.632	.	.645	.580	.678	.577	.442	.298	
C.	.670	.645	.	.644	.676	.560	.368	.194	
D.	.605	.580	.644	.	.646	.551	.401	.262	
E.	.659	.678	.676	.646	.	.421	.296	.114	
F.	.632	.577	.560	.551	.421	.	.455	.230	
G.	.527	.442	.368	.401	.296	.455	.	.444	
H.	.291	.298	.194	.262	.114	.230	.444	.	
TOTAL.	4.016	3.852	3.757	3.689	3.481	3.417	2.931	1.833	
Estimated Communality.	.670	.678	.676	.646	.678	.632	.527	.444	
GRAND TOTAL.	4.686	4.530	4.433	4.335	4.159	4.049	3.458	2.277	31.927
Factor 1.	.83	.80	.78	.77	.74	.72	.61	.40	Divisor 5.650

To obtain first approximations to the first and second factors, Thurstone's Centroid method was applied. This table shows how the first factor loadings were isolated. The totals of each column of correlations were found. This table has been arranged in simple hierarchical order, i.e. the totals diminish left to right. The "communalities" of each test were estimated by equating them to the highest correlation in the column. The communality of a test is the sum of the squares of all the group factor loadings of that test which can be isolated from the correlational matrix. This has to be guessed to make a first approximation and Thurstone's rule, followed here, is usually adequate.

These guessed communalities are added to the column totals giving the row "Grand Total". This row is summed (31.927) and the square root of this sum (5.650) becomes the "divisor". Each quantity in the row "Grand Total is divided by 5.650 to give the loadings in factor 1 (4.686 ÷ 5.650 = .83).

The next step is to find the "residual correlations" after factor 1 has been allowed for. The following table shows the working methods. -

FACTOR I RESIDUALS.

	A	B	C	D	E	F	G	H	Sign
Saturation Co-efficient.	.83	.80	.78	.77	.74	.72	.61	.40	
Observed Correlation. A.83	.670 .699	.632 .664	.670 .647	.605 .639	.659 .614	.632 .598	.527 .506	.291 .332	
Residual.	-.029	-.032	+.023	-.034	+.045	+.034	+.021	-.041	+
Observed Correlation.B. .80		.678 .640	.645 .634	.580 .616	.678 .592	.577 .576	.442 .488	.298 .320	
Residual.	-.032	+.038	+.021	-.036	+.086	+.001	-.046	-.028	+
Obs.Crrlation. C .78			.676 .608	.644 .601	.676 .577	.560 .562	.368 .476	.194 .312	
Residual.	- +.023	- +.021	- +.068	- +.033	- +.099	+ -.002	+ -.108	+ -.118	-
Obs.Crrlation. D .77				.646 .593	.646 .570	.551 .554	.401 .470	.262 .308	
Residual.	-.034	-.036	+.033	+.053	+.076	-.003	-.069	-.046	+
Obs.Crrlation. E .74						.678 .548	.412 .533	.296 .451	
Residual.	- +.045	- +.086	- +.099	- +.076	- +.076	- +.130	+ -.121	+ -.155	-
Obs.Crrlation. F .72						.632 .518	.455 .439	.230 .288	
Residual.	+.034	+.001	+.002	-.003	-.121	+.114	+.016	-.058	+
Obs.Crrlation. G .61							.527 .372	.444 .240	
Residual.	+.021	-.046	-.108	-.069	-.155	+.016	+.155	+.204	+
Obs.Crrlation. H. .40								.444 .160	
Residual.	-041	-.028	-.118	-.046	-.178	-.058	+.204	+.282	+
A lgebraic Sum	-.149	-.210	-.318	-2244	-.476	+.227	+.544	+.611	Divisor
Absolute Sum.	.259	.288	.472	.350	.890	.349	.570	.957	2.00
Factor II	-.074	-.105	-.159	-.122	-.238	+.113	+.272	+.305	

Factor 11 is calculated in the following way.

The part of each correlation accounted for by Factor 1 is calculated by finding the product of the loadings of factorI in each pair of tests:

For example -

Test A Factor I = .83
Test B Factor I = .80
Inter-correlation of A & B = .632
Of this (.83 x .80) = .664 of the .632 is accounted for by factor I.
The residual correlation for A against B is (.632 - .664) = -.032
For each cell of the original matrix of correlations a residual correlation is calculated.

Each row is then multiplied by a unitary weight positive or negative to bring out the pattern as clearly as possible. An algebraic sum (i.e. taking into account the signs) of the residuals is found. An absolute sum (i.e. neglecting the signs) of the residuals is found. This row of Absolute Sums is added and the square root of the answer gives the divisor. The algebraic sum for each column is divided by this divisor (2.00) to give factor II e.g. Test A (-.149 ÷ 2) = -.074.

These factor loadings for each test are finally tabulated in the following table. -

First Approximation to Factors I and II.

Test.	Factor I.	Factor II
A	.83	-.074
B	.80	-.105
C	.78	-.159
D	.77	-.122
E	.74	-.238
F	.72	+.113
G	.61	+.272
H	.40	+.305

Fifty-nine per cent of the variance of matrix of correlations is accounted for by these t o factors. One expects a bigger percentage than this for a battery of Intelligence Tests but the two tests, G and H, undoubtedly bring down the expected percentage considerably.

At this stage, an assumption is made that two factors fit the table.

Lawley's method of applying the principle of Maximum Likelihood to Factor Analysis is then applied for further clarification of the original centroid estimates of factor loadings.

Method of Maximum Likelihood:

In the first table each cell in the diagonal is completed with 1,000. The following table then shows the various stages of the re-estimation of two factors from the first centroid approximation.

	A	B	C	D	E	F	G	H
Factor I	.83	.80	.78	.77	.74	.72	.61	.40
" II	-.07	-.11	-.16	-.12	-.24	+.11	+.27	+.30
Specific Variance	.306	.348	.366	.393	.395	.470	.555	.750
(a)	2.712	2.299	2.131	1.959	1.873	1.532	1.699	.533
(b)	9.7147	9.3221	9.3244	8.9400	8.9422	8.2396	6.6022	3.9878
(c)	8.8847	8.5221	8.5444	8.1700	8.2022	7.5196	5.9922	3.5878
(d)	.8348	.8007	.8028	.7676	.7706	.7065	.5630	.3370
(e)	-.229	-.316	-.437	-.305	-.608	+.234	+.186	+.400
(f)	-17863	-.8627	-1.0142	-.8742	-1.1797	-.4432	+.0466	+.2748
(g)	-.7163	-.7527	-.8542	-.7524	-.9397	-.5532	-.2234	-.0252
(h)	-.7713	-.7398	-.7417	-.7092	-.7120	-.6527	-.5201	-.3113
(i)	+.0550	-.0129	-.1125	-.0432	-.2277	+.0995	+.2967	+.2861
(l)	+.0800	-.0187	-.1634	-.0627	-.3307	+.1445	+.4309	+.4155

h^2 = 113.28
i/h = .093956
p = -.9239
h^2 = .47418
i/h = 1.45224

The stages in the calculations resulting in two new estimations of the two factors (lines (d) and (l)) are as follows. -

Specific Variance: Sum the squares of both factors for each test and subtract from unity. The specific variance measures for each test experimental error plus specific factors for the test. Thus test H will have a large specific variance (.750) because many more factors other than the cognitive factors "g", V, N. etc., are measured in a set of examination results.

(a) Divide Factor 1 by the specific variance
(e.g. .83 ÷ .306 = 2.712)

(b) This row is given by the internal products of row (a) with the rows of the table of correlations (amended as above i.e. 1.000 in the diagonals). Thus row (b) for column A =
(1.000 x 2.712) + (.632 x 2.299) + (2.131 x .670) + (.533 x .291)

(c) Subtract from the figures in row (b) the corresponding loadings in factor I
e.g. (9.7147 - .8330) = 8.8847.

h^2 is next found and is given by the internal product of rows (a) and (c) e.g.
(2.712 x 8.8847) + (2.299 x 8.5221) + (.533 x 3.5878) = 113.28

$\frac{1}{h}$ is found by finding the reciprocal of the square root of h^2

(d) Multiply each item in row (c) by $\frac{1}{h}$. This gives the second approximation to factor I.

(e) Divide the trial loadings in factor II by the specific variance
(-.07 ÷ .306) = .229

(f) This row is given by the inner product of row (e) with the rows of the correlation tables.
(-.229 x 1.000) + (-.316 x .632) etc.

(g) This row is obtained by subtracting the trial loading in factor II from row (f).

p is calculated by finding the inner product of rows (d) and (e).

(h) This row is found by multiplying row (d) by p

(k) Subtract row (h) from row (g)

h^2 is given by the inner product of rows (k) and (e)

$\frac{1}{h}$ is found as the reciprocal of the square root of h^2

(I) This row, the second approximation to factor II is given by multiplying row (k) by $\frac{1}{h}$.

The following table shows the second approximation to factors 1 and II. -

SECOND APPROXIMATION.

Test.	Factor I.	Factor II.
A	.8348	+.0800
B	.8007	-.0187
C	.8028	-.1634
D	.7676	-.0627
E	.7706	-.3307
F	.7065	+.1445
G	.5630	+.4309
H	.3370	+.4155

It will be seen in comparing the tables that the factor loadings of Factor I have remained very much the same in the tests of high Factor I loading and have changed considerably (In tests G_2 H_1 factor II has changed considerably) but the same pattern is displayed.

Lawley's procedure is now repeated until there is no practical difference between the estimation of factor loadings assumed and factor loadings finally emerging from the calculations.

Actually, after another re-estimation the following factor loadings were obtained:-

THIRD APPROXIMATION.

Test.	Factor I.	Factor II.	Communality.
A	.8354	+.0828	.7047
B	.7993	-.0263	.6396
C	.8114	-.1627	.6778
D	.7664	-.0719	.5912
E	.7930	-.3602	.7244
F	.6944	+.2009	.5119
G	.5552	+.4636	.5232
H	.3260	+.4552	.3135

It will be seen that the differences between the Second and Third approximations are very small. The procedure was therefore stopped at this stage.

The "goodness of fit" of these two factors to the correlational matrix was tested at this stage by the chi-squared test. The residual correlations were calculated by the same method as was described in the description of the Centroid Analysis. The effect of the two factors was eliminated and the following table shows the residual correlation Matrix.

RESIDUAL CORRELATIONS.

	A	B	C	D	E	F	G	H
A.	(.295)	+.034	-.003	+.030	-.023	-.038	-.025	+.019
B.	-.034	(.360)	-.007	-.035	+.052	+.017	-.010	+.049
C.	+.003	-.007	(.322)	+.013	-.010	+.021	-.017	-.007
D.	+.030	+.035	-.013	(.409)	-.019	-.030	-.005	-.040
E.	+.023	+.052	-.010	+.019	(.276)	-.086	-.001	-.003
F.	-.038	-.017	-0021	-.030	+.086	(.488)	+.010	+.074
G.	+.025	-.010	-.017	+.005	-.001	-.010	(.477)	+.052
H.	-.019	+.049	-.007	+.040	-.003	-.074	+.052	(.687)

In the diagonals of this table are placed the specific variances of the tests.

Chi-squared for this table is found by finding the sum of the 28 terms found by squaring the residual correlation and dividing by the product of the numbers in the corresponding pair of diagonal cells.

Thus the term for cell B/E = $(+.052)^2 \div (.276 \times .360)$. This sum is multiplied by 120, the total number in the sample. This gives the value of chi-squared.

The appropriate number of degrees of freedom is calculated from the following formula:-

$$\frac{1}{2}\left((n - m)^2 - n - m \right)$$

where n = the number of tests
m = the number of factors assumed.

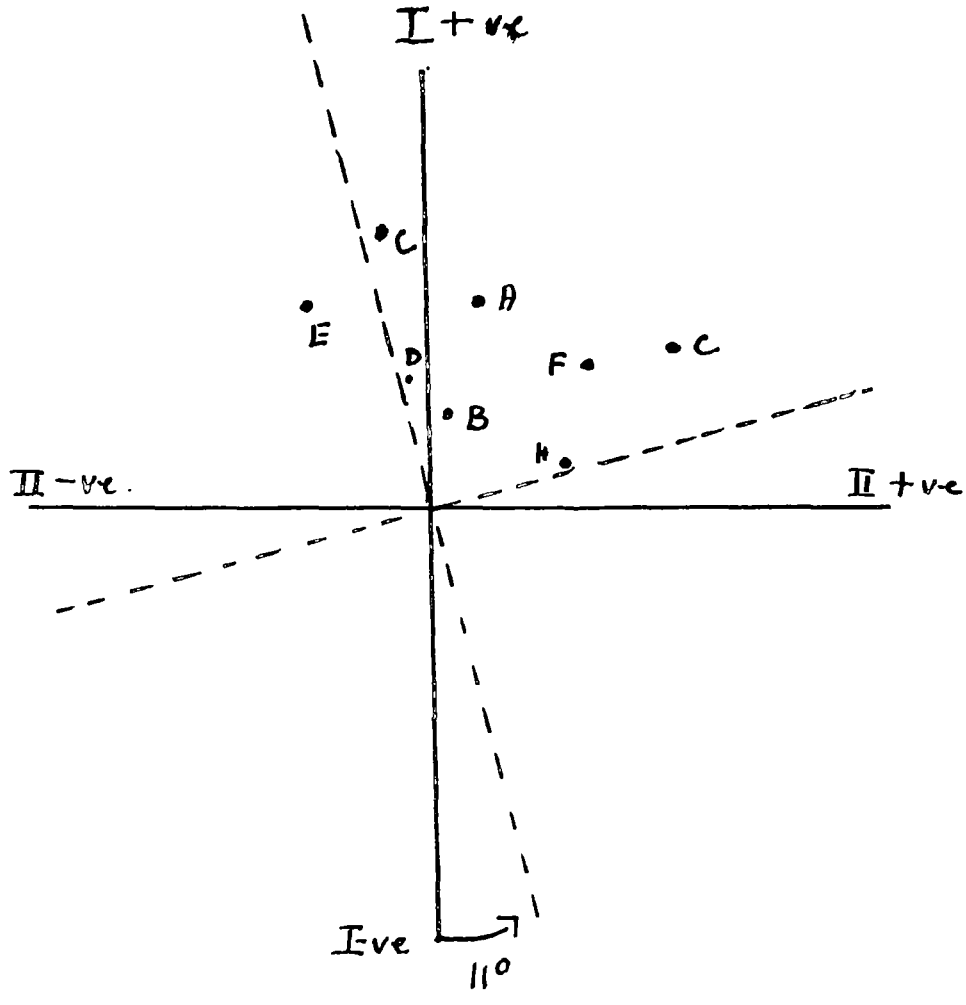
In this matrix n = 8 m = 2 df = 13
Chi-squared = 20.510

Entering Fisher's Table of Chi-squared distribution at df = 13 we find that p lies between .10 and .05.

Thus the differences between the theoretical residual correlational matrix assuming 2 factors only and the experimentally found matrix can be taken to be insignificant.

We may therefore draw the conclusion that there are no grounds for assuming more than two general or group factors to explain the original correlational matrix.

ROTATION OF AXES.



Factors calculated by this method usually lack psychological significance, however. For example, the factor loading of -.3602 in Factor II for test E cannot easily be identified as a real variable in human nature.

Now each pair of final factor loadings for each test can be located as points on a graph.

The relative positions of these points is unchanged if we rotate the axes but the values of the factor loadings may be re-estimated in terms which are psychologically significant.

The aim of the rotation which has been done in these results has been to eliminate negative factors. It will be seen on inspection of the factor loadings that this is best accomplished by rotating anti-clockwise about 11 degrees.

The general formulae for computing the new co-ordinates are:-

$$K\,\frac{1}{1} = K_1 \cos Q + K_2 \sin Q$$

$$K\,\frac{1}{2} = K_2 \cos Q - K_1 \sin Q$$

Where K_1 and K_2 = the original factor loadings for I and II

K_1^1 and K_2^1 = the new factor loadings after rotation for I and II

Q = angle of rotation.

For most purposes, however, purely graphical methods suffice so this method is used here for illustrative purposes. The points representing each test are plotted accurately on graph papers. The axes are then rotated and the new factor loadings measured from the new axes.

FACTOR LOADINGS AFTER ROTATION
OF THE AXES 11° ANTICLOCKWISE.

Test.	I	II
A	.800	.250
B	.790	.134
C	.832	.000
D	.770	.080
E	.846	-.200
F	.640	.340
G	.456	.560
H	.232	.504

This rotation of the axes through 11° anticlockwise would appear to give the best psychological meaning to the factors in spite of the fact that Test E (the Figure Reasoning Test) still has a negative second factor.

The first factor "g" is found to be appreciably present in all the intelligence tests and it would seem, therefore, wise to weight each of the tests equally in calculating a pupil's "g" score. It is hardly worth while weighting in accordance with Thompson's method.

The second factor is easily identified as the "V" factor. The isolation of this second factor indeed does nothing more than to confirm that in constructing the battery they group together, as was planned, as follows:

V Tests:	A	Vocabulary Test "V" loading	.250
	B	"Always" Test	.134
	F	Series Test	.340
	G	Classification Test	.560

Non-"V" Tests (probably "K" tests):

C Figure Classification Test.
D Figure Analogies Test.
E Figure Reasoning Test.

The battery of tests therefore seems to fulfill all that we shall demand of it. For each pupil taking the full battery of seven tests we are therefore able to give a measure of three factors:

(1) From a total score - "g" score.
(2) From total of A.B.F.G. - "V" score.
(3) From total of C.D.E. - "K" score.

It how remains to be described how these three measures were found.

During the calculation of the product-moment correlations the means and standard deviation of the tests were calculated.

These are as follows:	Mean.	Standard Deviation.
Test A.	31.4	7.9
B.	29.1	6.8
C.	10.1	2.8
D.	14.8	3.?
E.	25.0	7.1
F.	23.5	4.8
G.	30.6	5.1

Now in another work the present author has prepared tables for translating raw scores of tests of known mean and standard deviations into scores in a scale of mean = 50 and standard deviation 15. In every case, raw scores were promptly turned into these "standard scores" by means of this table.

We are now in a position to record the first stage of the results. The battery of tests was given in December, 1946 to a group of 120 first-year pupils in a Boys' Secondary Grammar School. Exactly a year later all these boys were retested with the same tests and under the same conditions.

The first result to be recorded is that the mean scores in each of these tests rose as would be expected. But the absolute value of the rise is much smaller than was expected. The standard deviaition remains almost identical. Here are the means and standard deviations of the seven tests in their second application. -

	Mean.	Standard Deviaition.
Test A.	32.9	7.9
B.	31.1	6.9
C.	11.4	2.8
D.	15.8	3.1
E.	27.1	7.0
F.	24.6	4.8
G.	32.4	5.0

Although they are not used in the detailed examination of the changes in intelligence, the correlations for each test - first result against second result - are interesting in themselves. These inter-correlations are done with raw scores:

Test.	Correlation Score/1946. Score 1947.
A.	.71
B.	.74
C.	.66
D.	.61
E.	.68
F.	.71
G.	.70

This set ofresults is quite in line with those recorded in Chapter 5. The general pattern of test "reliability" applies well fo this battery of tests if one takes into account the highly selected population we are using in our experiments. For the total population we would expect, if the pattern of the results were the same, that these correlations would be at least .1 higher if corrections were made for the selected nature of the group tested.

Unfortunately, it has proved, up to the present, impossible to obtain the standardisation of the battery on a full 11 plus year group to allow the corrections to be made, even if one used the approximations of the Doolittle method.

The next step was the most important in the whole of the statistical work. Three sets of measures were obtained each year as is described above. These "Total" scores were correlated, giving the following results. -

(1) "g" factor measures r (1st year/2nd year) = .70 ± SE.046
(2) "V" " " r (1st year/2nd year) = .72 ± SE.043
(3) "K" " " r (1st year/2nd year) = .658 ± SE.049

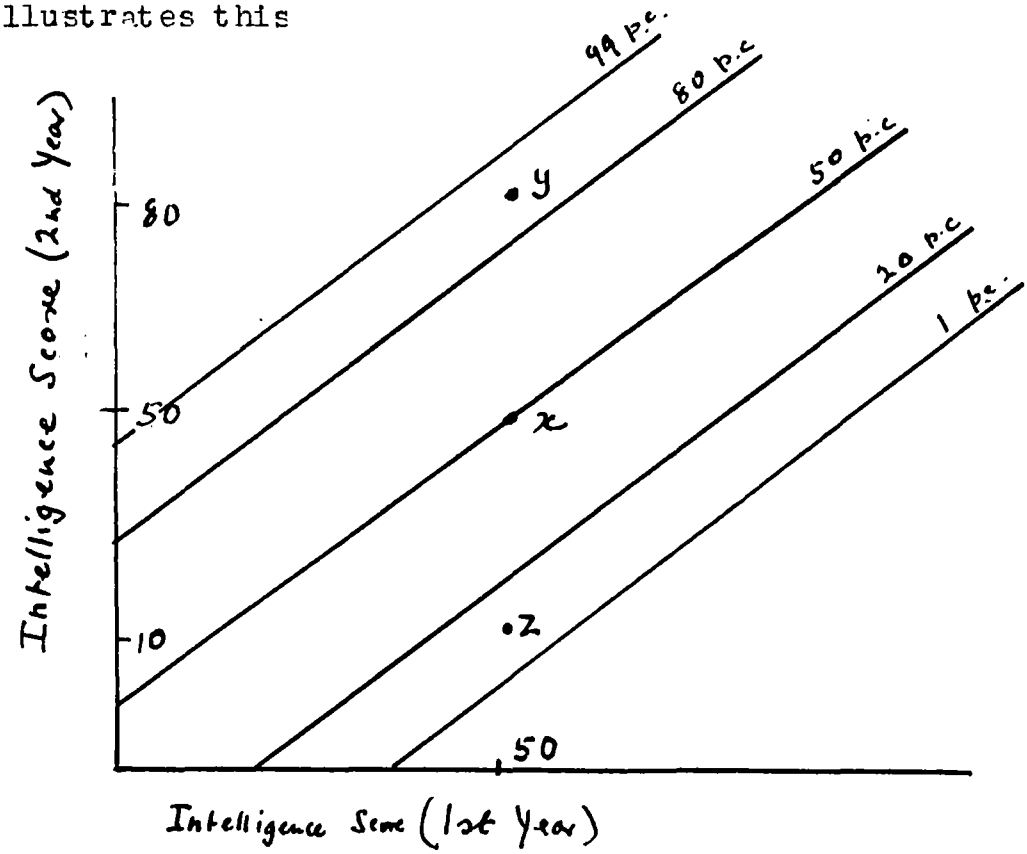
These correlations do not differ significantly and for this reason one graph will serve the purpose of all three tests of significant "discrepancy".

This graph was drawn giving first the regression line predicting the most probable score in the second year, from a knowledge of the first. This line is drawn by plotting the straight line graph

$$Y = (ryx)\left(\frac{\sigma_y}{\sigma_x}\right)(X - Mx) + My.$$

This was done for each "measure". The standard deviation of these "forecasts" is given by the expression

Itis therefore possible to draw a series of lines parallel to the regression line indicating the probability of a particular second year score arising from a given first-year score. The diagram below illustrates this



Suppose candidate X had a score of 50 in the first year test and a score of 50 in the second-year test. This is represented by the point x. He actually scores the most probable result. But suppose X scored not 50 but 80 in the second year test. This is represented on the graph by point y. He has not scored the most probable result after one year. He falls in area B which means that there was over a 5 to 1 chance against him scoring as high as 80 in the second year test.

Similarly, if he scored 10 on the second year test (represented by the point Z), he has achieved this low result with odds of 5 to 1 against him. Thus 40 per cent of all pupils will fall in areas A. B. E. and F.

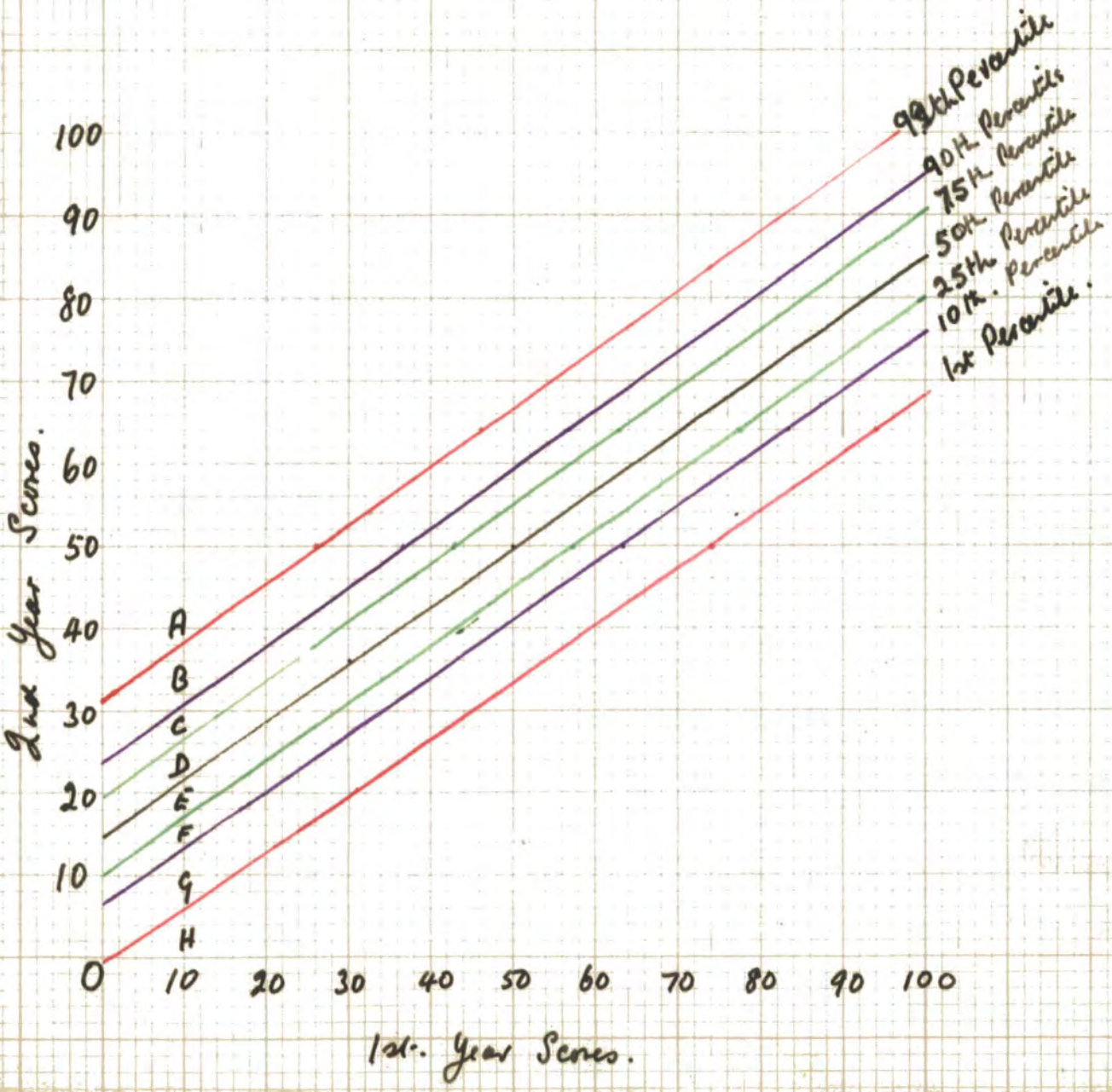
The actual graphs obtained for the three"factor" measures are given in the graphs which now follow. Notice that here the percentile "lines" drawn are the 1st, 10th, 25th, 50th, 75th 90th and 99th.

The first stage of the investigation concerns only those pupils who fall in areas A. B. G. and H. These number 23. To the examination of these 23 we now turn.

Regression of 2nd Year Scores on First Year Scores. $r = .70$

Standard Deviation of Both Scores = 15
Mean " " " = 50

Standard Deviation of Estimate. = 10·7



CHAPTER 9.

ANALYSIS OF THE SUBJECTS WHO HAVE SHOWN SIGNIFICANT CHANGE IN GRADING.

The twenty-three pupils so selected consisted of nineteen who were placed in areas A. B. G. and H. from a consideration of their two total "g" scores. Three more showed significant changes if only their "g and V" scores were taken into account, and one more showed discrepancy when only his "g and K" was taken into account. All the nineteen with discrepant "g" scores had discrepant scores in both (V and K) and (V and g). This is rather unexpected but with these numbers tests of significance of this fact would reveal nothing. The following table lists their scores:

PUPIL.	("g") Score. 1st Yr.	2nd Yr.	("g" & V) Scores 1st Yr.	2nd Yr.	("g" & K Scores. 1st Yr.	2nd Yr.
A	30	50				
B	34	50				
C	40	60				
D	44	56				
E	46	58				
F	50	60				
G	51	62				
H	57	66				
J	64	74				
K	73	76				
L	81	82				
M	25	24				
N	36	30				
O	48	30				
P	51	34				
Q	58	44				
R	65	44				
S	74	56				
T	83	62				
U			80	86		
V			69	50		
W			31	48		
X					76	86

Inspection of this table shows that the subjects A. to T. (i.e. those with significant changes in their "g" scores) have been arranged so that A to L showed significant increases in "g" during one year, whereas M to T showed significant decreases. There are thus 11 increases and 8 decreases. In the "V & g" changes are two increases, and one decrease, whilst the sole "g & K" change is a significant increase. This makes in all 14 increases and 9 decreases.

The next task was to select a control group. This was done in the following way. All the pupils lying in the "g" scores in areas D and E, comprising 62 pupils in all, were arranged in alphabetical order of names and the first twenty on this list were taken as the control group. These had scores which did not appreciably change during a year. The mean of their "g" scores in the first year was 53.9.

In all, therefore, 43 children had to be investigated in the qualitative way outlined above. We will now proceed to describe the details of this investigation.

Two main sources of information were open:

(1) Evidence of the pupil's teachers:

(2) Evidence from an interview with the investigator.

Unfortunately, at this stage it proved impossible to carry these investigations any further. It was hoped at the start to be able to visit the home and talk to the parents, but there are many "social" difficulties standing in the way of such fundamental investigations in England at the moment.

(1) Evidence from the teachers: All teachers (15 in number) were given the following questionnaire to fill in about the pupils which they had taken during this first year in the grammar school. Each form was taught by 7 masters:

(1) Without looking at your mark book, would you say that this boy has risen or fallen in his form in your subject?

(2) Now consult your mark book. Has he in fact risen or fallen in his form position in your subject?

(3) Have you noticed any significant change in his attitude to school work during the year?

(a) Has he taken more time than the average to "settle down"?

(b) Would you say he has become significantly "naughtier" or significantly better behaved?

(c) Has his written work, especially homework, significantly changed during the year - for the better - or worse?

(d) Have you any knowledge of any event in his own life or his family life which could have radically affected him in any way? If so, what?

(e) Have your noticed any physical changes during the year. If so, what?

(f) Does he seem "maturer" than the average for his age, or less mature, or average?

(g) Any further relevant information.

This evidence from questionnaires, for statistical purposes, had to be marshalled and graded into a standard form. In this case the evidence of seven masters had to be averaged out. This was done in the following way:

Question 1: If four or more masters guessed that he had risen in form position, the boy was given a plus mark for this.

Question 2: As for Question 1.

Question 1 and 2: Where answers to question 1 and 2 agree the boy was given a mark 0. Where the master "guessed" that he had risen when in fact he had fallen he was given a mark - and vice versa, a mark plus.

Question 3: In sub-questions a, b, c, and f, a "majority" opinion of masters is taken.

(a) plus mark for the answer "no"; - mark for answer "yes"

(b) plus mark for "better behaved; 0 for no change, and - for worse behaved.

(c) plus for improved written work, 0 for no change, and - for worsened written work.

(d) plus for any event recorded by any one master which in the opinion of the present writer was an event of such importance that it should have materially assisted the pupil, and - mark for any event which would tend to set back the pupil. 0 for no event of importance.

(e) plus for improved health as indicated by general health or rapid sturdy growth. - for a falling off in his health standard, for continued absence for health reasons, etc.

(f) plus for maturer; 0 for "average", and - for less mature.

(g) not marked, but commented on in the text later.

Care was taken to mix the papers of the control group and the group being investigated before handing to the masters. They were given no clue as to who were in the control group and who were not.

The following table lists the results of this investigation.

	QUESTIONS.								
Pupil.	1.	2.	1 & 2.	3a.	3b.	3c.	3d.	3e.	3f.
A.	+	+	0	+	0	+	-	0	0
B.	+	+	0	+	0	+	0	+	+
C.	-	+	+	+	-	0	0	0	0
D.	+	+	0	+	0	+	0	+	+
E.	-	+	+	+	-	-	0	0	+
F.	-	-	0	+	0	-	-	0	0
G.	+	+	0	+	0	+	0	+	+
H.	+	-	-	+	+	+	0	0	0
J.	+	+	0	+	0	+	0	0	+
K.	+	+	0	+	0	+	0	+	+
L.	-	-	0	+	+	0	0	0	0
M.	-	+	+	-	-	-	-	-	-
N.	-	-	0	-	0	+	0	-	+
O.	+	+	0	+	0	0	-	0	0
P.	-	-	0	-	-	0	-	0	-
Q.	-	-	0	-	-	-	0	+	-
R.	+	-	-	+	-	0	0	0	-
S.	-	+	+	+	-	-	0	0	-
T.	-	-	0	-	0	0	0	-	0
U.	+	-	-	-	+	+	0	+	0
V.	-	-	0	+	0	+	-	-	0
W.	-	-	0	+	0	0	0	0	0
X.	+	+	0	+	0	0	0	0	0
AA.	+	+	0	+	0	+	0	+	+
BB.	+	+	0	+	0	0	0	0	0
CC.	-	-	0	+	0	-	0	0	0
DD.	+	-	-	-	+	0	0	0	0
EE.	-	+	+	+	0	+	0	-	-
FF.	-	-	0	+	0	0	0	0	0
GG.	-	-	0	-	-	0	0	+	+
HH.	-	-	0	+	0	-	0	-	0
JJ.	+	+	0	+	+	-	0	0	0
KK.	+	-	-	-	0	0	-	0	+
LL.	+	+	0	+	0	0	0	0	-
MM.	-	-	0	+	-	0	-	0	0
NN.	-	-	0	+	0	+	0	0	0
OO.	+	+	0	+		+	0	-	0
PP.	-	-	0	+	0	-	0	-	0
QQ.	+	+	0	+	0	0	0	+	+
RR.	+	-	-	-	-	0	0	0	-
SS.	+	-	-	+	-	+	0	0	0
TT.	-	-	0	+	0	0	0	0	-
UU.	+	+	0	+	0	-	0	0	0

Contingency tables for each of these sets of data were prepared, and now follows an investigation of their significance. For convenience, those whose intelligence significantly changed are callled the α group. Those whose intelligence rose during theyear are called the β group, and those whose intelligence fell the γ group. Thus $\alpha \rightarrow \beta + \gamma$

Question 1:

Table showing the relationship between rising in intelligence score and the "guess" of masters whether or not they have risen in form position:

Guessed Rise in form position.		α		
		β	γ	
	+	9	2	11
	−	5	7	12
		14	9	23

Chi-squared = 3.884 = 1 P < .05

We may safely conclude therefore that the same factors which determined the change in level of ability measured by the intelligence tests in some way or other made a subjective impression upon the masters who took these pupils.

A further table compares the results of the group who significantly changed their intelligence level with those of the control group. Yet another table compares the results of the control group with those whose intelligence grading significantly rose:

Guessed rise or fall in form position:

	+	−	
Control α	11	9	20
	11	12	23
	22	21	43

By inspection there is no significant relationship here.

Guessed rise or fall in form position:

	+	−	
Control β	11	9	20
	09	5	14
	20	14	34

Chi-squared = .29 n = 1 Plus between .7 and .5

We cannot therefore conclude that the masters were able to distinguish between those who changed their intelligence and those who did not by the "total impression" which pupils made on them. Obviously this first "guess" of the masters is a very complex judgment and whilst the changes reflected by the intelligence tests do not go un-noticed, there are side by side with intelligence other factors which dispose the teacher to

estimate progress or retrogression of their pupils. This is a most important conclusion. It is believed that it is of considerable importance to those investigators working on the construction and usage of School Record Cards. The present author intends to follow this conclusion, with a properly designed full-scale experiment to confirm it in a more detailed manner than this particular set of results allows.

Question 2:

We may investigate first the results of this question by the same methods as were used for Question 1.

Actual rise in form position,		α	β	γ
	+	9	3	12
	−	5	6	11
		14	9	23

Chi-squared = 2.1 n = 1 P lies between .1 and .2

We cannot conclude that marks gained during the first year at school (form tests, homework marks, and examination marks) can significantly be related to an increase or decrease in intelligence test score. There are indications in the table of a trend in this direction but that trend might very well be attributable to chance distribution. When the relationship between answers to Questions 1 and 2 is investigated we obtain the following table:

Actual Rise in form position.		Guessed Rise in Form Position. +	Guessed Rise in Form Position. −	
	+	15.	5.	20.
	−	7	16	23
		22	21	43

Chi-squared = 8.5 n = 1 P < .01

It will be seen that there is a highly significant relationship between the "guesses" of the teachers and the actual recorded changes in form position. Partly, of course, this is due to memory of the actual figures but the chief part in this relationship is played by the ability of teachers after a year with a set of pupils to get an unconscious "grasp" of the abilities of their charges. This subtle ability of teachers is slightly better at estimating changes in intelligence than in attainment, which in itself is quite remarkable and worthy of following up.

Column 3 of the table which deals with the measure of agreement between the answers to Questions 1 and 2.

The first table measures to what extent masters' "guesses" agreed with reality for the α group and the control group.

Agreement.

	+	0	-	
Control	1	15	4	20
α	4	16	3	23
	5	31	7	43

Chi-squared (by Snedecor's Method) = 1.91 n = 2
P lies between .3 and .5

There is thus no evidence to show that masters find it more difficult to "remember" or judge the merits of the control group than those whose intelligence has significantly altered.

The answers to the various parts of Question 3 are more directly related to our researches.

Question 3 (a) deals with the complex phenomenon which teachers call "settling down". It is indeed a strange new world which little boys and girls of eleven are thrust into when they transfer to a Secondary School. They come from the Primary School where for a year at least, they have been "top dogs" into a world where they are very much the "under dogs". Big boys and girls laud their power and experience. New tasks, new teaching methods and hosts of new teachers to contend with. This is especially true of those pupils transferred to a Grammar type of Secondary School where the added complication - that the boy who was the bright boy of his Junior School now turns out to be the dull one of the highly selected population which constitutes a "grammar" year. It is scarcely to be wondered that individual differences in the ability of pupils to find their feet and settle down are shown quite markedly. Of the 43 children examined in this investigation, 10 showed some difficulty in "settling down".

When oneexamines the α group against the control group the following table results.

	+	−	
Control	16	4	20
α	17	6	23
	33	10	43

There is obviously no significant difference between the ability of the α group to settle down and the ability of the control group so to do.

When the β group is analysed side by side with the control group the following picture results.

	+	−	
Control.	16	4	20
β	13	1	14
	29	5	34

Chi-Squared = 1.08 n = 1 P is approx. 0.3

There seems to be no significant difference between the ability ofthese groups to "settle down".

When the β group is placed beside the γ group the following table results:

	+	−	
β	13	1	14
γ	4	5	9
	17	6	23

Chi-squared = 6.7 n = 1 P is less than .01

It would seem, therefore, even if we make allowances for the small number in one cell of the table that there is a highly significant difference between the ability of the two groups.

Those pupils whose intelligence score rose did in fact find it easier to settle down in the first year, according to the judgment of their masters. This is a highly significant fact. It was pointed out above that this "settling down" is a complex phenomenon which would need further detailed investigation in order to unravel its many strings. But it is certainly something which is connected with the total psychological make-up of theindividual and is probably not so much due to the more direct environmental factors in personality. The main contributory factors are "innate" ones which reflect the ability of the individual to cope with the world of people. Introversion, extroversion, the "W" factor, schizoid - manic depressive traits and so on, would probably, in the opinion of the writer, be highly represented in this ability. But this is only speculation aroused by the interesting result obtained. It does, however, indicate the lines which a further research could profitably take.

Question 3 (b) deals with another complex phenomenon - behaviour in school. By behaviour here is meant the limited feature of general behaviour, the willingness or otherwise of the pupil to co-operate ith the school in general. This is the popular, school-masters' use of the word. It is a purely subjective judgment and varies accordingly from teacher to teacher. Improved behaviour is therefore difficult to estimate accurately, but it is remarkable how close the agreement of masters was on

the question of improvement or retrogression on this point. Masters agreed better in deciding the change of behaviour than they usually do in estimating the level of co-operation.

The α group is first compared with the control group.

	+	0	-	
Control	3	13	4	20
α	3	13	7	23
	6	26	11	43

There is obviously no significant differences here. The following table compares the β group with the control group but it unites those whose behaviour has not significantly changed with those who showed improvement:

	+ & 0	-	
Control	16	4	20
β	12	2	14
	28	6	34

Again, it is obviously by inspection that there is no significant difference.

The β and γ groups are now compared:

	+ & 0	-	
β	12	2	14
γ	5	4	9
	17	6	23

Chi-squared - 2.6 n = 1 P = approx. .1

Thus there are not sufficient grounds for concluding that the boys whose intelligence increased also increased in co-operation with the school, although larger numbers might indicate such a trend.

Question 3 (c) deals with the written work of pupils. Now written work in a Grammar school reflects more than capacity to do work. It is not so much an attainment test as another fruit of question 3 (b) - the co-operation with the school. Most of the written work is in fact homework, so that this represents also those factors which are concerned with the co-operation of the parent and home with the school, or, more accurately, the ability of the home to co-operate with the school. It is difficult for a child who has several older and younger brothers and sisters in small houses and with the shortage of fuel to have the sort of conditions favourable for the production of good written homework. Again, the "free-time" of the pupils vary a great deal in these days when so many children (30 per cent in this particular school) do paid work of some kind or another. Together with this there are also factors of so complex a causation that we have to describe them as intrinsically individual, whilst recognising that somewhere they have objective causation which has not been exposed.

Firstly, group β is compared with the control group. + and 0 are combined:

	+0	-	
Control	15	5	20
β	12	2	14

Thereis, by inspection, no significant difference. When β is compared with γ the following table results:

	+	0	-	
β	12		2	14
γ	6		3	9
	18		5	23

Chi-squared = 1.1 n = 1 P = approx. .3

Thereis thus no evidence to show that boys whose intelligence test rose considerably, improved their written work in greater proportions than those whose intelligence fell.

Question 3 (d) concerns known events which have occurred in the life of the pupil during the year which may have had some bearing upon his efficiency as a pupil.

It is a sad comment on this world that no masters could discover any pupil who had experienced events which could have materially changed the pupil's efficiency in the positive direction. No parent even won a large figure on the football pools enabling a larger house to be purchased. There were a number of events, however, in the negative direction - eight in all. These could be classified as follows:-

(1) Death of one parent. (2)
(2) Serious accident. (1)
(3) Serious illness of parent (3)
(4) Broken homes. (2)

Here is group γ compared to the control group.

	0	-	
γ	5	4	9
Control	18	2	20
	23	6	29

Chi-squared = 7.5 n = 1 P is less than .01

Even allowing for the small numbers in some cells, this result is so highly significant that it is a justifiable conclusion that there is evidence to believe that "catastrophe" in the lives of pupils occurred more frequently amongst those whose intelligence level deteriorated than amongst the control group.

The next table compares groups β and γ

	0	-	
β	12	2	14
γ	5	4	9
	17	6	23

Chi-squared = 2.6 n = 1 P = approx. .1

This comparison shows an insignificant relationship but taken in relationship with the previous result, it may safely be concluded that "catastrophe" in the life of the pupil is reflected in some reduction of his capacity to perform in intelligence tests. As far as can be ascertained, none of these eight events were so immediate to the second time of testing that the result could be attributed to the direct, immediate "shock" effect.

This is a most interesting conclusion. It seems to the present writer that it points to the factors which are of fundamental importance to changes in intellective capacity. It is not so much the type of home in general which is important. It is not socio-economic status as such which determines how intelligence grows. It is rather the complex working out of the intimate interpersonal relation of the home and its deep effects on the ego which determines the rate of growth of intelligence. As in all these investigations of relationship between rate of intelligence growth and personality changes, the most important fact seems to be that it is the developmental changes of the deeper processes of an individual's personality which are primary. Events which are classified as trauma affecting the whole emotional and psychic inner life of the subject are more likely to influence the rate of growth of intelligence than more superficial changes, however radical these may appear at first sight.

Question 3 (e): This question concerns the physical fitness of the selected pupils. There is no consistent health record card kept covering the period at school so that it was necessary, in analysing the factor, still to use the subjective judgments of the masters.

The β group is first compared with the control group, and $+ \propto 0$ are grouped together:

	$+_0$	-	
Control	16	4	20
β	14	0	14
	30	4	34

Chi-squared = 3.1 n = 1 P is greater than .05

With such small numbers in the cells, this chi-squared must be regarded as insignificant. The control group is now compared with the γ group:

	+0	-	
Control	16	4	20
γ	5	4	9
	21	8	29

Chi-squared = 1.8 n = 1

This is insignificant.

Finally, the β & γ groups are compared:

	+0	-	
β	14	0	14
γ	5	4	9
	19	4	23

Chi-squared = 12.4 n = 1 P < .01

Even making allowance for the small number in some cells, this is significant evidence from which to conclude that there is evidence to show that pupils whose health has deteriorated recorded a falling off in intelligence level. The extent of the relationship is probably not high but it is sufficient to demonstrate that physical health, mental health and ability are inter-related factors.

The final question which can be subjected to statistical treatment concerns "maturity". The age at which boys come into adolescence varies considerably. The masters who made judgments

on maturity were asked to take into account the various well-known features connected with sexual maturation - e.g. voice breaking, hair on chin, interests and general "poise".

The α group is first compared to the Control group. + and 0 are combined.

	+ 0	-	
Control	16	4	20
α	18	5	23
	34	9	43

On inspection there is no significant difference between these two groups.

The β and γ groups compared give the following table:

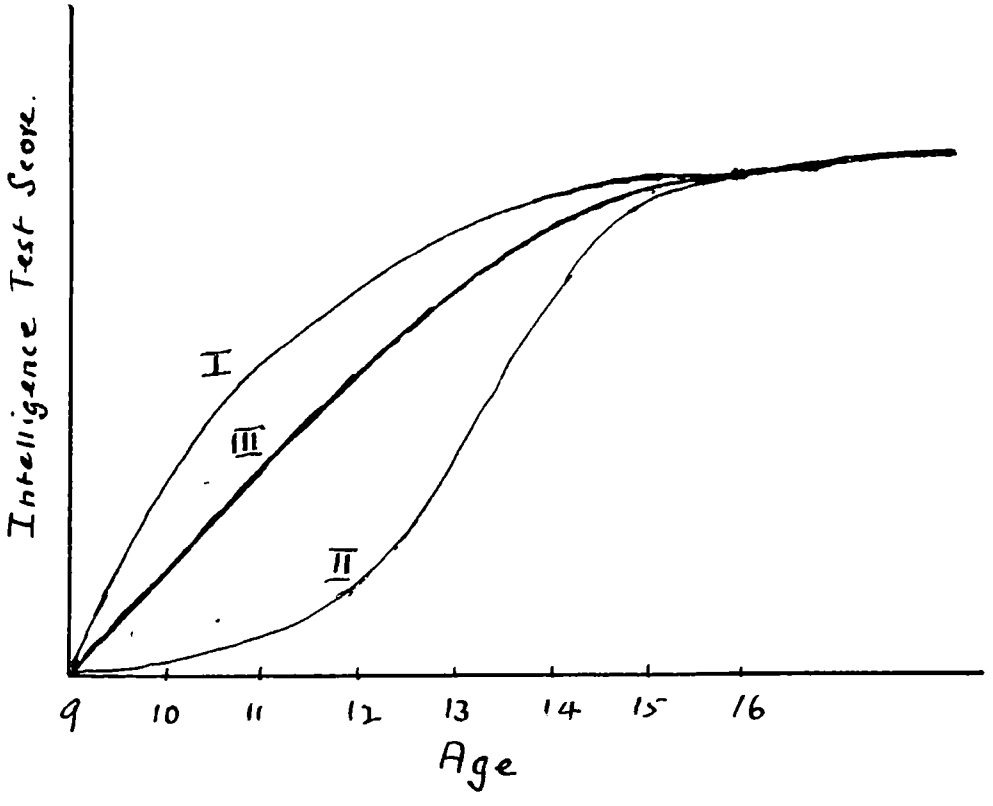
	+ & 0	-	
β	14	0	14
γ	4	5	9
	18	5	23

Chi-squared = 9.9 n = 1 P is less than .01

This means that even taking the small numbers in certain cells into account we may reasonably draw the conclusion that those pupils who matured earlier than the rest showed improvements in their intelligence level.

Now this result is at first glance rather surprising - for the following reason. In The graph showing the average intelligence test scores against age, to which we referred in an earlier chapter,

the rate of growth of intelligence would appear to fall off with approachingmaturity. Yet when one studies individuals it is precisely those boys who mature early who make a spurt in ability to do intelligence tests. This is illustrated by reference to the following graph which is only meant to be illustrative - i.e. not accurately plotted, but which is the only shape which will conform to the above data



Curve I represents the development in intelligence of a boy who matures (sexually) late; i.e. between 11 and 12. His rate of growth is shown as slower than the average (Curve III). Curve II represents the development in intelligence of a boy who matures early i.e. between 11 and 12; the rate of growth is faster than the average.

There appears a contradiction in this. How is it to be explained? This is yet another example of the dangers of using the statistical "average" to draw conclusions about specific individuals. Growth to maturity is, of course, a complex physio-psychological process. It means to the individual, more than any other single fact, the opening of an entirely new world of experience. It gives new fields for the exercise of cognitive abilities but, more than that, it provides new drives and urges for the whole personality in the business of living. It raises complex problems of individual responsibility in quite new forms. These facts, it would seem, are reflected in the narrow field of ability to do intelligence tests. If the emergence of new "drives" and "motives" can influence basic abilities in this way, how much more would we expect it to influence those abilities (like the ability to do a real job of work for example) which derive more directly from the demands of the world outside the individual's own psyche. This accords with the experience of teachers who find that maturity (both sexual and social) does in fact affect considerably the ability of pupils to profit from instruction. We must, in drawing conclusions, especially remember this point.

Yet there is no real contradiction here. It is only an apparent one, drawing special attention to the dangers of deductive arguments in psychological science.

The last section of Question 3 gave masters the opportunity of recording any special features not covered in the foregoing questions. There were only four entries in all here, and these were of value only in assisting the investigator himself when he interviewed the pupils individually.

The last stage of the investigation, that of a personal interview with all the 43 boys who were reported upon by the masters, proved to be of little value in throwing any special light on the problem.

An attempt was made to confirm the ratings of masters on "maturity" andon special events in the life of the pupils which may have some special significance. In the case of estimation of "maturity", this proved to be very difficult in a short interview and was abandoned. Only one further case of "special events" was brought to light, but this was not investigated because it was of so recent origin that it could have only reflected itself as an "immediate" causation. The pupil, anyhow, was in the control group.

It was also intended to try to estimate the nature of the home back-ground - whether the parents co-operate in helping the child to do his school work, his home-work, etc. This, it is well-known, is an important factor in determining success in a grammar school.

It was found very difficult to make such estimations of home co-operation simply by interviewing the pupils. The results seemed to be so unreliable that no analysis was made.

This individual interview approach, therefore, on the whole proved to be non-productive. The reasons for this would seem to be that the technique of child-interview is still not properly worked out or mastered by the present writer.

CHAPTER 10.

....CONCLUSIONS....

In this thesis we have ranged over the whole problem of selection for secondary education at eleven plus. We have analysed the nature of these problems and in some attempted accurately to ask specific questions, direct answers to which will illuminate the whole. Finally, we have carried out a planned piece of research in search of an answer to a concise question and have been able to give some direct answers.

During the argument it has been important to state at times, the gaps which it is felt have to be filled in to give a more complete answer to the general questions, and have at times formulated certain general proposals for further research. It is the intention of the author to press forward, himself and in co-operation with other workers, to carry out this research planned for the future.

Since at each stage various definite conclusions have been drawn and stated it will not be necessary to restate them. It only remains for certain general conclusions to be drawn.

The question of selection at 11 plus was approached from two angles (1) What is a suitable criterion from which the efficiency of selection is best related? (2) To what extent is the variability of intellectual growth attributable to factors which are within our control?

The researches of the author lead him to answer in the following way.

(1) The criterion of success is a historically determined one. It varies with the nature of the social needs of the day which always pound urgently at the educator's door. We are moving towards a society whose production and inter-personal (social) techniques will demand a far higher proportion of first-class able people than ever before in human history. We are, in fact, moving towards a society which may be characterised as a society of the democratised genius. This requires that the education system seeks means of developing the potentialities of every single individual. More than that, it must utilise all available knowledge of the plasticity of human personality and ability, creating conditions for the fullest development of all.

(2) The answers to the second question follow naturally from the answers to the first when combined with our specific results together with the rapidly accumulating body of research data pointing in a similar direction.

We conclude firstly that the limits of prediction accuracy have been reached by the best selection techniques at present in use. Further improvement is impossible because the remaining variance is due to factors determining the uneven development of human ability. The factors which lead to this uneven development are complex integrants of the inner development of the individual and the impact of an ever-changing environment on him.

We have discovered certain of these environmental factors but have shown that they are scarcely traceable to isolated events or conditions.

Intelligence tests are constructed principally to smooth out the effects of such change-producing factors, yet it is shown that the results of these are not so stable as an inspection of the "average" (statistical average) child might lead one to conclude.

How much more variable and plastic must be total social performance which is the result of an integration of "intelligence" "drive", "motives", "knowledge", and the character of the individual's emotional connative life!

It would seem, therefore, to the author that our researches have contributed to the point of view which declares that selection into three "types", described in the Norwood Report, should be abandoned as impossible and wasteful; and it supports the point of view of those who say that secondary education should be carried on in Comprehensive Schools whose conditions in regard to space, equipment, quality of staff and especially staffing-ratio, should be far higher than obtains at present in <u>all</u> the municipal secondary schools of today.

- - - - oOo - - - -

March, 1948.

REFERENCES.

1. ADKINS, D.C. Effects of Practice on Intelligence Test scores. J. Educ. Ps. 1937. 28.

2. ARTHUR G. Prediction Value of Kuhlmann Binet for Inmates of State School for Feebleminded. J.Appl. Ps.1933 17.

3. ARTHUR G. Prediction Value of Kuhlmann-Binet Score for a Partially Americanised School Population. J.Appl. Ps.1937.21.

4. ASHER, E.H. Inadequacy of Current Intelligence Test for Kentucky Mountain Children. J. Genet Ps.1935.

5. BAKKEN E.R. & MAURER S. Variations in Psychological measurement associated with increased Vit. B. Complex Feeding in Young Children. J. Exp. Ps.1934. 17.

6. BAYLEY N. Mental growth During First 3 years. Genet Ps. Monogr. 1933 14.

7. BENTON A.L. Influence of Incentive upon Intelligence Test Scores of School Children. J. Genet Ps. 1936. 49.

8. BROWN R.R. Time Interval Between Test and Retest in Relation to Constancy of I.Q.

9. CANDADY H.G., Effect of "Rappont" on I.Q. A New Approach to Racial Psychology. J. Negro Educ. 1936 5.

10. CATTELL, P. Do Stanford-Binet I.Q's of Superior Boys and girls tend to Decrease or Increase with age. J.Educ. Res. 1933. 26.

11. CATTELL, P. Standford-Binet I.Q. Variations Sch. and Soc. 1937. 45.

12. COFFEY H.S. and WILLMANN B.L. Role of Cultural Status on Intelligence Changes of Pre-School Children. J. Exp.Educ. 1936. 5.

13. CRISSEY O.L. Mental Development of Children of Same I.Q. in Differing Institutional Environments. "Child Development". 1937. 8.

14. CRISSEY O.L. Mental Development as Related to Institutional Residence and Educational Achievement. Univ. Iowa St. Child Welf. 1937. 13.

15. CUNNINGHAM. Infant I.Q. Ratings Valuation After Interval of 7 years. J. Exp. Educ. 1934. 3.

16. DANE K.J. Effects of Practice Upon Intelligence Tests. Brit. J. Educ. Ps. 1938. 8.

17. DAVIDSON C.A. Permanency of Mental Ability Measures. J. Appl. Ps. 1938. 22.

18. DAWSON S. Environmental Influence on Mentality. Brit. J. Ps. 1936. 27.

19. DRESCOLL G.P. Development Status of Pre-School Child as Progress of Future Development.

20. DUNELL D.D. Influence of Reading Ability in Intelligence Measures. J. Educ. Ps. 1933. 24.

21. ENGEL A.M. Constancy of I.Q. Mental Defectives. Nation. S. Sch. 1937. 19.

22. ENGLISH H.B. & KELLIAN C.D. Constancy of I.Q. at Different Age Levels. J. Consult. Psy. 1939. 3.

23. FERGUSON H.H. Incentives and an Intelligence Test. Aust. J. Psy. Phil. 1937. 15.

24. FORAN T.G. Constancy of the I. Q. a Review in the Univ. Amer. Educ. Bull. 1929. 4. No. 10.

25. FORAN T.G. Supplementary Review of Constancy 1929. No. 9.

26. FRITZ M.F. Effect of Diet on Intelligence and learning. Ps. Bull. 1935. 32.

27. FORFEY P.H. & MCHLENBEIN J. Validity of Infant Intelligence Tests. J. Genet Ps. 1932. 40.

28. GARRISON S.C. Retests on Adults at 10 Year Interest. Sch. and Sol. 1930. 32.

29. GESELL H. AMATRADA, CUSTNER & THOMPSON. Biographers of Child Development. N.Y. Hocker 1939.

30. GELDEN H. & MACOUBREY C. Factors affecting Constancy of I.Q. of Problem Children. Smith Coll. Stud. Soc. Work. 1933. 3.

31. GARDENAUGH F.L. Kuhlmann-Binet for children of school age; a critical evaluation. Minneapolis. 1938.

32. HALES W.M. Results of Group Intelligence Retest with Reformatory Group. J. Juvenile Res. 1937. 21.

33. HALLOWELL D.K. Stability of Mental Test Ratings for pre-School Children. J. Genet. Ps. 1932. 40.

34. HAWTHORNE J.W. Effect of Improvement of Reading Ability on Intelligence Scores. J. Educ. Ps. 1935. 26.

35. HERRING A. Experimental Study of Reliability of Bahler Baby Test. J. Exp. Educ. 1937. 6.

36. HINTON R.T. Effect of a Constant and Superior Environment upon I.Q.'s of One Hundred and 17 Children. Ps. Bull.1938. 35.

37. HIRSCH N.D.M. Experimental study upon 300 School Children over 6 year Period. Genet Psy. Monogr. 1930. 6.

38. HOAKLEY Z.P. Variability of I.Q.'s Proc. Amer. Ass. Stud. Feeblemin. 1932. 37.

39. HOLLINGWORTH L.S. & KAUNETZ R.M. Intell. status of Gifted Children at Maturity. J. Genet Ps. 1934. 45.

40. HONZIK M.P. Constancy of Mental Test Performance During Pre-School Period. J. Genet Ps. 1938. 52.

41. HUBBARD R.M. A Study of the Reliability and Validity of Bahler Infant Scale. J. Genet Ps. 1935. 47.

42. JORDAN J.S. Reliability of Stanford-Binet I.Q's. given by Student Examiners. J. Educ. Res. 1932. 26.

43. KAWEN E. & HOEPER C. "Comparative Study of Nursery School v. a non-nursery School Group. Univ. Chigago 1931.

44. LAMSON E.E. To What Extent one I.Q.'s Increased by Children who Participate in a Rich, Vital, School Curriculum. J. Educ. Ps. 1938. 29.

45. LEAKY A.M. Nature-Nurture and Intelligence. Ganet Ps. Monog. 1935. 17.

46. LINCOLN E.A. Stanford-Binet I.Q. Changes of Superior Children. Ich. & Soc. 1935. 41.

47. LINCOLN E.A. Study of Changes in I.Q.'s of Superior Children. J. Educ. Res. 1935. 29.

48. LINCOLN E.A. Stanford-Binet I.Q. Change on Growth Study. J. Appl. Ps. 1936. 20.

49. LINCOLN E.M. & WADLEIGH V.C. Change and Constancy in Group Test I.Q.'s Sch. & Soc. 1932. 35.

54. LITHAWER D.B. & KLINEBERG O. Study of Variation in I.Q. of Group of Independent Children in Institution and Foster-Homes. J. Genet. Ps. 1933. 42.

55. LIVESAY T.M. Does Test Intelligence Increase at the College Level. J. Educ. Psy. 1939. 30.

56. LODGE T. Variation in Stanford-Binet I.Q's of Pre-School Children according to Months of the Examination. J. Psych. 1938. 6.

57. LODGE I. Retests after ten Years. J. Educ. Ps. 1930.

58. LARGE I. & HOLLINGWORTH L.S. Adult Status of Highly Intelligent Children. J. Genet Ps. 1936 49.

59. LOWRY. Increasing the I.Q. Sch. & Soc. 1932. 35.

60. MALLER J.B. & ZUBIN J. Effect of Materialism upon Intelligence Test Scores. J. Genet. Ps. 1932. 41.

61. MASTERS H.V. & UPSHALL C.C. Study of Years Made by Normal School Studies in Intelligence Test Scores. J. Educ. Res.1934.

62. MATEER F. Pituitary Acceleration of I.Q.'s. Ps. Bull.1934 31.

63. McCONNELL T.R. Change in Scores on Psychological Examination of American Council on Education from Freshmen to Senior Year. J. Educ. Ps. 1934. 25.

64. McNEMER Q. On Inadequacy of Reliability Co-efficient when used with I.Q. Data. Ps. Bull. 1936. 33.

65. MILLER E.A. Emotional Factors in Intellectual Retardation. J. Ment. Sc. 1933. 74.

66. MILLER W.S. Variation of I.Q's Obtained from Group Tests. J. Educ. Ps. 1933. 24.

67. NELSON V.L. & RICHARDS T.W. Studies in Mental Development. Performance on Gesall Items at 6 Months and Productive Value for Performance on Mental Tests at 2 and 3 years. J. Genet. Ps. 1938. 52.

68. NELSON & RICHARDS. III . Performance of 12 Months-old Children on Gesell Schedule and Productive Value for Mental Status at 2 and 3 Years. J. Gen. Psl 1939. 54.

69. NEMZCK C.L. Constancy of I.Q's of Gifted Children. J. Educ. Psy. 1932. 23.

X 70. NEMZCK C.N. The Constancy of the I.Q. Ps. Bull.1933 30.

71. PETERSON T.J. Preliminary Study of Effects of Previous Nursery school Attendance upon 5 year old Children entering Kindergarten. Univ. Iowa St. Child.Welf.1938. 14.

72. POULL I.E. Effect of Improvement in Nutrition on Mental Capacity of Young Children. Child Dev. 1938. 9.

73. SCHELL M. Infection by Intestinal Protozon in Relation to Intelligence of Siblings. Child. Dev. 1933. 4.

74. SCHOTT E.L. I.Q. Change in Foster-Home Children. J.Appl. Ps. 1934. 18.

75. SCOGOL M.V. An evaluation of Certain Intelligence Tests. J. Appl. Ps. 1934. 18.

76. SHERMAN M. & KEY C.B. Intelligence of Isolated Mountain Children. Child Dev. 1932.

77. GRAY. The Nation's Intelligence.

78. FLEMING, C.M. The Growth to Maturity. (North of England Education Conference).

79. HAMLEY, H.R. The Testing of Intelligence.

80. SPEARMAN, C. The Abilities of Man.

81. SPEARMAN, C. The Nature of Intelligence.

82. BURT, C. Factors of the Mind.

83. BURT, C. Mental and Scholastic Tests.

84. KELLEY. Cross Roads in the Mind of Man.

85. BALLARD. Mental Tests.

86. THOMPSON G.H. The Factorial Analysis of Human Ability.

87. THOMPSON, G.H. The Essentials of Mental Measurement.

88. GUILDFORD. Psychometrid Methods.

89. FISHER. Statistical Methods for Research Workers.

90. FISHER. Design of Experiments.

91. MURPHY, MURPHY & NEWCOMB. Experimental Social Psychology.

92. PAVLOV. Works, Vol. I and II.

93. VERNON, P.E. The Measurement of Abilities.

94. KOCHLER. The Mentality of Apes.

95. THURSTONE L.L. The Vectors of Mind.

96. ALLPORT G. Personality.

97. BLACKBURN J. Psychology and the Social Pattern.

98. WATSON J.B. Behaviourism.

99. MEAD, M. Growing Up in New Guinea.

100. MEAD, M. Coming of Age in Samoa.

101 GINSBURG M. Studies in Sociology.

102. LASHLEY, K.S. Brain Mechanisms and Intelligence.