A Study of certain aspects of selection for secondary education in an East Anglian county in 1966

Riggulsford, P.

ABSTRACT OF THESIS.

P. Riggulsford: A Study of Certain Aspects of Selection for Secondary Education in an East Anglian County in 1966: M.Ed.

The chief aspects studied in the enquiry were accuracy (percentage correct) and speed (percentage attempted) in completing standardised tests taken at the age of eleven and their relationship with academic performance three years later in secondary schools. A brief review of the 11+ examination itself was included.

The secondary assessments were scaled using standardised verbal reasoning tests. For larger schools, product moment correlation coefficients were calculated 'within' schools and later 'pooled'. For samples below 25, rank-order correlations were obtained for comparison. In all 553 boys and girls in 11 modern and 6 grammar schools were studied. The table illustrates the trend of the findings:-

Correlation between secondary school assessments and measures obtained from the 11+ examination.

English.	'EQ'	'Accuracy'	'Speed'
Boys (137)	0.547	0.507	0.478 (67 b.only).
Girls(245)	0.411	0.358	0.283 (140g.only).
Mathematics.	**'AQ'.**	**'Accuracy'**	**'Speed'**
Boys (195)	0.449	0.406	0.245 (67 b.only).
Girls(245)	0.457	0.448	0.157 (108g.only).

The figures appear to indicate that accuracy was almost as

efficient as a predictor of final order as the subject quotients and was also more closely associated than speed with the secondary rankings.

A review of 70 'pairs', matched for sex, age and verbal reasoning ability at the age of eleven, gave similar indications.

A study of 192 children in 2 grammar and 2 modern schools indicated consistent standards of accuracy at the age of eleven, in different tests.

For comparison of the relationship suggested for 'accuracy', a brief survey was made of other correlates previously found in the assessment of attainment in secondary schools.

Although the samples were small, the present evidence may well have value in 11+ 'borderzone' considerations and also for practising teachers.

ACKNOWLEDGEMENTS

I wish to take this opportunity of expressing my gratitude to those without whose help the project would not have been possible. They include:-

Professor F.V.Smith, for his constant encouragement, advice and assistance.

The Chief Education Officer for the County, for his personal interest, encouragement and support.

Many members of the Staff of the Education Department, and especially Mr.Colin Norman, for so readily making available tests, scores, record sheets and other information; for their courteous and prompt help despite their other commitments.

Mr.David Peter, the Chief Education Officer for Darlington, for his kind interest and support in the original enquiry in 1962/63.

Professor J.Nisbet, for making so much information available prior to the publication of his latest book and for suggesting several additional sources of reference.

The Head and Assistant Teachers in schools in the area, who have so willingly co-operated with the provision of information, and especially those who voluntarily marked most of the Secondary Verbal Tests used in the enquiry.

The Authors of books, theses and articles mentioned in the following pages, and acknowledged in more detail in the lists of reference.

My sister, Mrs.G.M.Whiteside, for arranging, at some inconvenience to herself, for the production of the typescript.

My wife and children, for their patience and sympathetic understanding throughout the enquiry.

October 1969. P.H.Riggulsford.

A Study of Certain Aspects of Selection for Secondary Education in an East Anglian County in 1966

Table of Contents

/continued

Table of Contents continued

1. General Introduction

At a time when the country is turning to Comprehensive Secondary Education and the "eleven plus" examination is disappearing from one area after another, it may seem unusual to look yet again at such an examination and aspects of it bound up with prediction of subsequent performance in academic subjects in secondary schools.

The factors (not necessarily in order of importance) which motivate this enquiry include:-

(a) a desire to complete work, interrupted by family illness, for a higher degree,

(b) curiosity, particularly about the possible relationship between "accuracy under pressure" and success in school work,

(c) the fact that in many areas, for a variety of reasons, including the sheer impossibility of reorganising secondary education because of lack of suitable buildings, the "eleven plus" examination will continue for at least another five years, if not longer (*please see footnote) and

(d) the fact that even where secondary comprehensive schools exist, in most cases children are still "allocated" to "streams" or "sets" and consideration of assessments based on standardised tests including "intelligence"/verbal reasoning tests, is important. This statement should be slightly amplified, without comment as this aspect of organisation is a complex one involving arrangements for large numbers of teachers as well as children, both groups containing members with markedly varying interests, abilities and motivating forces. It is perhaps sufficient to note that the omission of the use of standardised

*The footnote is given at the end of this section so that it can be read as a whole.

tests when "homogeneous" groups of children are being gathered together for teaching purposes will permit many unstandardised errors to exert an influence.

Modern views of "intelligence" emphasise complexity, specificity to the situation and motivating forces at work. They generally tend to discredit Verbal Reasoning tests as anything other than a fairly rough assessment of ability to engage in a particular kind of "convergent" thinking, to use Professor Liam Hudson's phrase. The point which I considered of interest was whether accuracy at the expense of speed was a desirable attribute in children's application to the tasks before them. From this standpoint, the "eleven plus" examination seemed a fruitful source of information, obtained under reasonably standardised conditions. In the County concerned the examination is optional. This implies that the majority of the children who attempt the papers will be highly motivated to do well. Of course, some will be driven by parents, teachers or convention. The assessment could be reviewed in two ways: (i) against performance in the examination itself and (ii) in the light of subsequent attainments in the secondary school.

It seems to me that this question of "accuracy under pressure" may well be relevant to teaching method at the present time. Some teachers like to emphasise near perfect performance of a given task before tackling the next. This attitude seems likely to influence children, especially those who wish to do well in school examinations and so should be reflected in the approach to the "eleven plus" hurdle. Thus a conclusion that a "follow up" review of children in secondary schools showed that success was more likely to be associated with speed in tackling "eleven plus" papers rather than accuracy, would have considerable implications. Conversly, if accuracy was shown to be more "profitable", an attribute which may well be considered under attack in modern adult life as well as in school,

would receive renewed enhancement. As far as I am aware this has not been the subject of any previous enquiry.

Before considering this subject in more detail it is desirable to fill in the background by reviewing the "eleven plus" examination as a whole.

---

*Footnote:

Wilson Marks writing in "The Teacher" on June 20th. 1969 reported a speech made a few days earlier by Mr.Edward Short, Secretary of State for Education, in the House of Commons. Referring to the contemplated new Education Act, Mr.Short said, "If the anti-comprehensive authorities are allowed to stall (until the Tories are in power) they could be safe from change for at least another decade. Furthermore, the bill will have to recognise that comprehensive schemes could take anything from 2 to 25 years to implement.

"Eight Local Education Authorities have positively refused to submit schemes. In 102 Authorities comprehensive schemes cover the whole or most of their area, and in a further 22, schemes have been approved for part of the area. Another 12 have schemes 'in the pipeline'. Of the remaining 29, quite a number are in the process of discussing plans."

On the same day the Times Educational Supplement included a report from a large Education Authority (Essex) of a plan to introduce comprehensive secondary education in one of the areas of the County "in 4 or 5 years time". (Some other parts of Essex have comprehensive schools - the point is that in the country as a whole there will be many areas where "selection at eleven plus" occurs for several years to come).

2. The Main "Eleven Plus" Selection Procedure And "Follow Up" Review

(a) Introduction

"A visitor from Mars, reading the general educational press and the daily papers would probably come to the conclusion that the predictive ability of the "11+" is either zero or negative. A perusal of research reports and the more technical journals would lead him to believe that the examination is highly valid. In fact, the validity coefficients obtained.... demonstrate a level of predictive ability unmatched by any other selection programme in Education, or in personnel selection in industry and the armed forces.. .. If, however, one looks at the mistakes.... it is clear that the prediction is far from perfect." (Wiseman, S. 1961)

That Professor Wiseman's remarks concerning the emotional heat engendered by the examination as evidenced by "popular" press reports are still relevant, is shown by an article in "The Daily Telegraph" on August 29th. 1969, "Why 60,000 children start next term in the wrong school", which includes this interesting paragraph, "The chairman of the Board of Governors of a primary school in Richmond, Surrey, writes: 'We are very perturbed about selection, which seems to be more unfair than the old 11-plus examination.' ".

While the "eleven plus" examination may gradually be passing into history it clearly has exerted and continues to exert a tremendous influence upon the lives of thousands of children. "The idea of Secondary Schooling is, of course, very man made and arbitrary ..... Today and in recent decades there have been many societies in which the demand is for more social and educational equality. This demand has naturally tended to favour theories of varied and developable attitudes rather than theories of unitary "intelligence" and relatively fixed abilities. In relation to secondary school learning this has

meant an emphasis on prolonging and varying learning opportunities. It is a considerable operation to work this ideal into the official educational system and perhaps a still greater one to work it into the fabric of society." (McFarland, H.S.N., 1969) "Largely for historical reasons, one type of secondary education, which is normally provided by the grammar school and which gives access to the universities and other avenues of professional training, enjoys much greater prestige than any other. It was inevitable in the absence of provision at this level for every aspirant,....., that competition for places should arise and with it the need for some form of selection." (Smith, F.V., 1957). "It is known half the parents in the country desire this type of education, although Grammar School places are available only for some 20% of children..." (Vernon, P.E., 1954)

The importance of "grammar schools" as opposed to other routes to higher education and the professions can really be traced back to the influence of Sir Robert Morant, the Permanent Secretary to the Board of Education in the early years of this century. In the eighteen-nineties, as the pressing need to provide higher education for a larger proportion of the population became apparent, there were really two possibilities. The first was to strengthen the "public" and endowed "grammar" schools (already improved following the Endowed Schools Act of 1869). The second would have been to legalise the development of the "Higher Grade" schools with which many enlightened School Boards, notably in London, were achieving great success. The 1894 Royal Commission (the Bryce Commission) considered ways of establishing secondary education and recommended the establishment of a government department and of local authorities. In 1899 the Board of Education came into being and Local Education Authorities were set up within the newly created

County and County Borough framework in 1902.

The Cockerton judgement of 1899 virtually ended the Higher Grade Schools and Sir Robert Morant's design of the new Code for Secondary Schools ensured that they were modelled on the old endowed "Grammar" Schools, with their tradition of academic classical education.

"The Education Act of 1902 recognised that, although a course of studies similar to that followed in the endowed grammar schools was appropriate for only a minority of children, such children were nevertheless to be found amongst all classes of the population. Under its terms the new local education authorities...(were given)... the 'power to assist in providing scholarships for and paying or assisting in paying the fees of students'. In 1907 the Board of Education made their grants to recognised secondary schools conditional upon the allocation of a proportion of places - 25% for most schools - to entrants from the elementary schools." (Yates, A. and Pidgeon, D.A. 1957)

The age of admission to secondary schools varied from 10 to 13 and there was pressure both from the grammar schools and the Board of Education to stabilise the arrangements. Dr. Entwistle, quoting figures from the Board of Education Reports for the years 1923 - 25, shows the marked reduction in the general age of transfer after 1913, concluding, "Eleven to twelve is more and more becoming the predominant age of entry". He goes on, "It is not unreasonable to suppose that these figures had a profound influence on the Hadow Committee (1926) when they came to recommend a fixed age of transfer". (Nisbet, J.D., and Entwistle, N.J. 1966)

"There is a tide which begins to rise in the veins of youth at the age of 11 or 12. It is called by the name of adolescence. If that tide can be taken at the flood, and a new voyage begun in the strength and along the flow of the current, we think that it will move on to

fortune. We therefore propose that all children should be transferred at the age of 11 or 12." (Hadow Report, 1926) The term "11+" in connection with the examination for entry to secondary achools occurs several times in "The Reliability of Examinations".(Valentine, C.W. and Emmett, W.C., 1932)

Thus, briefly, the historical determinants of the "eleven plus" examination are reviewed, although, as I trust the quotation from Professor McFarland's book implied, "grammar school education" has changed over the years and continues to do so.

References in Section 2(a)

Hadow Report	1926:	The Education of the Adolescent	H.M.S.O., London. (p.xix)
McFarland H.S.N.	1969:	Human Learning: A Developmental Analysis	Routledge and Kegan Paul (pp.77 and 83)
Nisbet J.D. and Entwistle N.J.	1966:	The Age of Transfer to Secondary Education	University of London Press. (p.60)
Smith F.V.	1957:	Secondary School Selection: edited by P.E.Vernon	Methuen, London. (Preface)
Valentine C.W. and Emmett W.C.	1932:	The Reliability of Examinations.	University of London Press.(pp.32-34 and 98)
Vernon P.E.	1954:	Selection for Secondary Education by J.J.B. Dempster.	Methuen, London (Foreward)
Wiseman S.	1961:	Examinations and English Education	Manchester University Press. (pp.150-151)
Yates A. and Pidgeon D.A.	1957:	Admission to Grammar Schools.	N.F.E.R., Newnes, London. (p.3).

2(b) Review of previous literature

This section deals with previous "follow up" enquiries and the content of the "eleven-plus" examination generally in use. Several factors, sociological and non-intellective aspects of personality included, influence attainment levels found in secondary schools besides those customarily examined in selection procedures. These other factors are mentioned later in Section 5.

However, these very factors also influence the classroom attainments of junior school children and must be borne in mind in designing "eleven-plus" examinations. As Professor F.V.Smith, in the Preface mentioned earlier, notes -

"As a matter of historical fact, it was in the attempt to penetrate the masking effect of some of these features of educational advantage deriving from social inequality, that the so-called intelligence tests were first devised. In the early years of this century Binet and Simon were concerned to discover to what extent the educational backwardness of groups of Parisian children was due to lack of schooling or to poor innate ability. This followed the pioneering work of Galton in Britain, who was interested in discovering innate abilities of a high order as well as in the problem of assessing cases of subnormality. In the 1920's Godfrey Thomson, working in the county of Northumberland, was able to show that there were appreciable discrepancies between the attainments of children and what might have been expected from their performance in a test situation which sought to exclude attainment and assess innate ability. Since they are designed to minimise the advantages of differential schooling and preparation, so called intelligence tests of the type applicable to groups have been included in many procedures of selection in the attempt to be as fair as possible to all candidates. ...."Clearly too, since

assiduity is an important factor in any task, and intelligence tests are sometimes held to favour the quick and "slick" candidate, who may not be adequate in preparing an argument or in written expression, there must be some review of the evidence on other selective devices, such as attainment in school examinations, the use of teachers' estimates and the essay." (Vernon, P.E. - Editor 1957). There are also administrative considerations. "The majority of standardised tests, despite their variation, yield information in an hour or less which can otherwise only be obtained in some much more time consuming way." (A.W.Heim, 1954). Professor Valentine also noted this advantage, describing how in a large city unskilled clerks marking mental tests produced in "far less time" an order of merit "remarkably similar" to that of the entrance examination for secondary schools. (Valentine,C.W. 1932).

Researches in the 1920's and 1930's included Professor Valentine's study of some 1600 pupils in 22 different schools in the areas of 10 different authorities. He related the order of merit in the entrance examinations in 1924 to the performance in the School Certificate examinations 4 or 5 years later. Somewhat surprisingly,in view of the limited use of mental tests and varying age allowances, he obtained for one centre with 625 pupils a rough average correlation of 0.40. (p.179). He noted an article by A.Donald Amos, in 1931, and said "on the whole... a combination of intelligence tests plus the written entrance examination gives the best forecast of what a pupil will do, not only in his first and second, but even in his fourth year at school". (Valentine,C.W., 1932.p.98) In the 1941 West Riding Report Special Place Examinations (London) reference was made to an enquiry in Yorkshire in 1937 where the object was to find the relative value of two types of examination:- (i) Moray House Standardised Intelligence Tests together with unstandardised tests

in English and Arithmetic and (ii) A battery of Standardised Tests (Moray House) in English, Arithmetic and Intelligence. The results of these tests were compared with the order of merit which the Primary School Head Teachers estimated before the tests were given. For unsuccessful pupils it was found that the multiple correlations of both types of examination were almost equal, but for successful pupils the multiple correlations of the standardised tests exceeded the unstandardised tests. The highest value was obtained from the Arithmetic Tests (in both cases) and English gave the least predictive values. Another enquiry undertaken in the West Riding in 1935 compared an Intelligence Test plus traditional English and Arithmetic Tests with "internal" assessments obtained in Secondary Schools after two and three years of Secondary work. The standardised Intelligence Test was found to be significantly a better predictor than the other two. When the three tests were considered as a team, the best prediction was obtained with relative weighting of the tests (Intelligence:English: Arithmetic) of 2: 1: 1. (Emmett,W.G., 1942). T.E.Stubbins (1940) reports on two enquiries he undertook. Dealing with one sample of 154 and another of 148 children he used the results obtained in nine School Certificate subjects (English Language, English Literature, History, Geography, German, French, Mathematics, Physics, and Chemistry) as the criterion of success. He averaged the School Certificate marks and used the weighted average in correlations with predictions based upon Arithmetic, English and Intelligence Tests. He found that the English mark was the best predictor with the Intelligence Score second. English was the best predictor of School Certificate marks in that subject and the Intelligence Score was the best predictor of success in Mathematics and Physics - better than the Arithmetic score. The correlations were all of the

order of 0.2 and are given in more detail later (page22).

A major contribution was reported by Professor William McClelland in 1942. To use Professor James Drever (Senior)'s words this was "the final Report of the Scottish Committee.... in the International Examination Inquiry...representing the completion of the tasks undertaken in 1931 by the Scottish Delegation". This was a most comprehensive work and showed that the best battery for predicting secondary school success was made up of a standardised Intelligence Test together with the combined English and Arithmetic marks in the qualifying examination of the traditional type plus the primary school teachers' estimates of pupils' attainments in English and Arithmetic, suitably scaled. The scaling of junior school teachers' estimates was an important feature of this research. McClelland concluded that the estimates for English and Arithmetic had to be scaled separately, on the results of uniform examinations in English and Arithmetic respectively. The scaling method, (K.J.Holzinger,1928.), used the formula

$$T_S = M_Q + \frac{\sigma_Q}{\sigma_T} (T - M_T),$$

where T_S is the Teachers' scaled estimate for the subject,

M_Q is the Mean of the marks obtained in the qualifying examination,

T is the Teachers' raw estimate in the subject,

M_T is the Mean of the Teachers' raw estimates,

σ_Q is the standard deviation of the marks in the Qualifying examination in the subject, and

σ_T is the standard deviation of the Teachers' estimates.

Finally, the Teachers' scaled estimates for English and Arithmetic were added together to give a single score for each child.(pp 40-47).

3,229 pupils were considered. It should be remembered that transfer was at the age of 12, as the research related to Scotland. After careful consideration, the Tests were weighted equally.(p.83) Follow up cards were devised for the purpose of securing a general estimate of the pupils' progress at the conclusion of the first three years of the post primary course, or when he left school.(p.7) It was proposed to follow up the senior secondary pupils for five years, but this project had to be abandoned because of war conditions. The research was notable too, for the very great care taken to try to ensure that the secondary school examination marks, on which the follow up criterion was based, would be comparable.(Pupils in different classes all took the same terminal examination. Teachers marked the same questions throughout the school age group, rather than all the questions in a set of papers.) Even so, there were differences between schools and in Chapter 6, Professor McClelland gave a detailed description of the methods used in combining correlations. One possible method would have been to calculate separate correlations for each school and then to use Fisher's z transformation. This might have been in error due to differences in ability levels between the children in the different schools. The theory behind the method finally chosen was to scale the marks obtained in the secondary schools on the qualifying examination marks, so that the marks between schools were comparable, and then to carry out a single correlation procedure. The "best battery" was found to be a combination of IQ + marks from the traditional type Qualifying examination in English and Arithmetic plus junior teachers' scaled estimates. The correlation found was 0.804. (Incidentally, another section, pp 20 - 22, was devoted to showing that Intelligence Quotients could be used instead of scores obtained from intelligence tests.).

The research also provided information about the predictive power of objective type standardised tests for English and Arithmetic. A battery of IQ + Standardised English and Arithmetic + Teachers' scaled estimates, gave a correlation of 0.786. (A correlation of 0.898 had been found between the Qualifying Examination traditional type tests in English and Arithmetic and standardised testsin those subjects. p.22)
A battery of IQ + Standardised tests in English and Arithmetic gave a correlation of 0.738, which compares with 0.784 for IQ + the traditional type of qualifying examination. With IQ alone, the correlation found was 0.690. (p.166).

Follow up enquiries considering children transferring to Secondary Modern Schools are rare. H.S.N.McFarland (1948) described an enquiry relating to three such schools in Darlington. Seven predicting variables were considered: A Moray House "Intelligence" Test, a Moray House English Test, a Moray House Arithmetic Test, Teachers' ratings for English, Arithmetic, Other Subjects, and also Teachers' Estimates on a carefully defined scale of "character". The Secondary School Head Teachers were asked to provide a single assessment for each child and these were checked to ensure that the scores were well distributed. The Primary School Teachers' estimates were scaled against the appropriate subject quotients. The Secondary School assessment was made after two years of the secondary course. The intercorrelations of the seven predicting variables were calculated separately for each secondary school and then the three batteries of correlations were pooled. The pooled intercorrelations were used to obtain the regression coefficients of the predicting variables on the criterion. Aitken's method of pivotal condensation was used. The description given in the Thesis is very detailed. I shall give only a few of the many findings. In table 4 multiple correlations for the different possible combinations of tests in

predicting batteries are given. For IQ alone the correlation was 0.707; for IQ + EQ + AQ 0.747; for IQ + EQ + AQ + Teachers' estimates in English and Arithmetic 0.769, and for all 7 predicting measures 0.781.

As the multiple correlations were best weighted according to their regression coefficients, the correlation for the unweighted sums of all the predicting variables was also calculated for each school. Pooling these three correlations (by sums of squares and products) the single value of 0.736 was obtained - compared with the 0.781 mentioned above. Next to the Intelligence Quotient, the Character Rating was found to have the most significant regression on the criterion, but when the Character Ratings for 16 of the larger primary schools were correlated with their appropriate scaled Teachers' estimates, it was found that most of the correlations were greater than 0.8 and this suggested that the Character Rating might be unduly influenced by academic success. Incidentally, I would mention here that H.J.Hallworth (1964) also concluded that they were greatly influenced by academic performance.

Another notable enquiry was that by Dr.Rutter, who considered all the pupils entering a grammar school in the years 1934 - 1943, except 1937. He compared the actual marks obtained by them in the qualifying examinations, with the actual marks obtained by them in three separate School Certificate subjects taken five years later. It was a large mixed grammar school in an urban coal-mining area, with an intake of 96; some 10% of the age group. Previous investigators, notably Prof. Valentine, had used a single criterion of grammar school success, usually an estimated order of merit reversed and treated as marks. Rutter used the composite aggregate marks obtained in English, Geography and Mathematics. The qualifying entrance examination was originally unstandardised tests in English and Arithmetic weighted 3:2 with an age

allowance of 1% of the actual marks obtained for each complete month below 12 years of age. This became 2% in 1938. An unstandardised Intelligence Test was added from 1937 - 1939. From 1940 - 1943 a single Standardised Intelligence Test was used. (M.H.) The criterion was the N.U.J.M.B. examination considering only those pupils who sat the 4 subjects (average for English and English Literature) taking the examination for the first time. He used a "pooling square" (G.Thomson,1948) with both qualifying tests and School Certificate subjects weighted in any given way and secondly, weighting the School Certificate subjects 1:1:1 and 1:2:1. The cases of maximum correlation were calculated using a method described by Professor E.A.Peel,(1948). He was, of course, concerned only with a sample of selected pupils, containing approximately 10% of the population (not corrected for the whole population). For the unstandardised English and Arithmetic Tests in the years 1934 - 36 the correlation ranged from 0.50 to 0.57. This was raised when an unstandardised Intelligence Test was used (1937 - 39) to a range of 0.65 to 0.71. For the years 1940 - 43, using a single standardised Intelligence Test, the correlations were 0.63, 0.43, 0.62, and 0.68 respectively. He concluded that the predictive value of the entrance examination was considerably improved by adding an Intelligence Test and also that the single Moray House Intelligence Test gave prediction almost as good as the combination of unstandardised tests. For the years 1940 - 43, the School Certificate tests correlated highly among themselves (English showed the closest relation with the Intelligence Test). The multiple correlation of the School Certificate subjects as a team with the Intelligence Test was high, reaching 0.711 in 1943 if weighting for School Certificate was English 12: Mathematics 5: Geography -1: the latter being due to the high intercorrelations of the School Certificate examinations. (Rutter,D.1950)

Professor E.A.Peel and Dr.Rutter then carried out a further survey on the predictive value of the entrance examination as judged by the School Certificate Examination. For this they used 5 Grammar schools in Durham County out of 20, those taking the N.U.J.M.B. examinations, whose results were issued as marks. In 1944 423 pupils were admitted to the schools on the basis of standardised tests in Intelligence, English and Arithmetic and 279 of these sat School Certificate in 1949. For the criterion they used English Language, English Literature, French and Mathematics as one group and Physics, Mathematics, Biology and Geography as a second group. This reduced their number of candidates to 234 for the first group and 52 for the second. They found that with this series of Entrance Tests the Standardised IQ score was the best SINGLE predictor of achievement at School Certificate,whether one considers either predominantly Arts or Science. (approx. 0.4). The English Test at entry was a more efficient predictor of English Language, English Literature and French in School Certificate than was the Arithmetic Test as a predictor of Mathematics success. (0.48 compared with 0.39, a difference almost twice the standard error of the 0.48 correlation coefficient.)

The correlations over the period of 5 years between Intelligence and English and the School Certificate subjects were of a similar magnitude, 0.48, 0.48, 0.46, 0.44 and 0.43, to the correlations between the School Certificate subjects, English Language, English Literature and French: 0.59, 0.54 and 0.43. This indicated the efficiency of an order of merit based on the 1944 entrance tests as a forecast of an order of merit based 5 years later on the School Certificate subjects. School Certificate Mathematics and Physics were better predicted than Geography and Biology by tests of Intelligence and Arithmetic.

As a team the battery of entry tests gave better predictions of the core subjects, English Language, English Literature, French and Mathematics, than of the predominantly science subjects. The best team weights were IQ: English: AQ, at 1.0: 0.9: 0.3, to give a best prediction of 0.585 for the "core" subjects and weighted 1.0: 0.5: 0.7, to give a best prediction of 0.487 for the science subjects. If the weights of the School Certificate subjects were also varied to yield the maximum possible correlation between the two "teams", maximum correlations of 0.597 for prediction of English Language, English Literature, French and Mathematics; and of 0.554 for the Science subjects, were obtained. Simple weighting of the entry tests and the School Certificate subjects led to a prediction of 0.585 and 0.487, significantly close to the maxima quoted above. (Peel E.A. and Rutter D. 1951.)

A major contribution to information on the role which could be played by Teachers' Estimates was made when G.Bosomworth (1953) took as the criterion of success the child's order of merit in form examinations during the first two years of grammar school life. He said he considered that many extraneous influences affect a child during his school life "up to the School Certificate stage". At 11-plus he considered the grading examination marks and also Teachers' estimates on a "T" score. This was based on the order of merit in various Primary School subjects (Oral and written English, Mental and written Arithmetic, Nature Study and Social Studies) each given a mean of 50 and standard deviation of 10. This average "T" score did not allow for differences in ability between school groups and the County as a whole, not did it allow for differences in age. He then obtained a scaled "T" score by finding the mean grading examination mark for each school group and increasing or decreasing each child's "T" score (average) by the

corresponding fraction of the standard deviation. An age allowance was also made. He found comparatively equal correlations between Teachers' ratings as predictors and the grading examination as the predictor. The grading examination mentioned consisted of standardised objective tests of ability and attainment. The best prediction was obtained by adding scaled Teachers' Estimates, the Intelligence Test and the English Test, with roughly equal weights, (0.332 and 0.342 respectively) and adding the Arithmetic Test with a good deal less weight (0.058). The adding of scaled Teachers' Estimates clearly improved prediction, since the multiple correlation was raised from 0.844 to 0.918.

W.G.Emmett and F.S.Wilmut reported in 1952 on enquiries following up 90 boys and 38 girls from the 1941 entry and 112 boys and 41 girls in the 1942 entry to Grammar Schools in Huddersfield. The "entrance" batteries were Moray House Tests in Intelligence, English, and Arithmetic, given to the whole age group. The secondary criterion used was the marks obtained subject by subject in School Certificate Examination. With, English Language for example the correlations were IQ: 0.505; EQ : 0.498; AQ : 0.204. With Mathematics they were IQ: 0.514; EQ : 0.299; and AQ : 0.429. All the main subjects of the Secondary School curriculum were considered.

Later Dr.Emmett contributed a further article based upon a study of 985 children who transferred to 12 Grammar Schools and 939 who transferred to 13 Modern Schools, following objective tests given in the Primary Schools in March, June and November of 1947, and compared with the County Transfer examination taken in February 1948. The children in the Secondary Schools were assessed in July 1951. Correlations as follows were found:-

	E + A in School Tests	In County Exam.
Grammar School Children	0.811	0.830
Modern School Children	0.848	0.853

It will be noted that the "traditional" papers gave slightly higher correlations. The criterion of secondary success was left to the Heads of the Secondary Schools and in the Grammar Schools covered some ten subjects. The Modern School Heads used English, Social Studies, Mathematics, Science, Handicrafts, Needlecraft, Art and Physical Education. Within each school an order of merit was drawn up and converted to a standardised score with a normal distribution using specially prepared tables. The best single predictor was found to be the Intelligence Test. For the Grammar Schools the figures were :-

	IQ.	English EQ.	Arithmetic AQ	Multiple Correlation (best weights)
Using School Tests:	0.792	0.721	0.735	0.834
Using County Tests:	0.786	0.740	0.768	0.831

These, Dr.Emmett noted (p.94) were high and he referred back to his earlier study (Emmett, 1942) relating to 765 children who entered Secondary Schools in the West Riding in 1933 and 1934. In that enquiry, after 3 years of Secondary Schooling the multiple correlation with the entrance tests was found to be 0.764 (p.33). It was noted that the criterion of success in the Secondary School was somewhat less well defined for the earlier enquiry (though pages 20 and 21 of the book provide ample evidence of the very great care taken by the Secondary School teachers).

Professor Jack Wrigley (1955) gave a most valuable commentary on previous investigations and reported on his own findings.
First, he made a very important point. "Any criterion of success in school is likely to have some limitations...(we) are merely using the best estimates available." He ignored W.McClelland's findings in Scotland beaause the correlations obtained were not exactly comparable. For the 461 children in Scotland the correlations obtained were Arithmetic 0.671; English 0.699 and Intelligence 0.70. Wrigley

argued that the selection for senior schools in Scotland was not very stringent at the time and so the follow up groups were more variable in ability than the groups in England. In addition a correction had been made for restriction of the range of ability in the second and third year groups.

He also ignored two early researches: Amos A.D. (1931) and Collier J.W. (1933), because the Intelligence Tests were not used in the selection process itself. Sandon, F.(1936) had shown "that it is not valid to compare a follow up correlation obtained from a test which was not used for actual selection purposes with one which was so used. There is a tendency for the test which was not used to yield a higher correlation than it should, just because it had a non-selective function". Actually Sandon, in a most illuminating article dealing with the necessary errors which any selective process must make, coined a phrase to warn all students of "selected" groups with non-normal" distributions, "the 'other' test tends to be better!".

Wrigley's own enquiries were based (a) upon work in a Northern Grammar School and (b) upon a study in Northern Ireland. In Lancashire he studied 227 children, firstly considering their terminal examination marks and Teachers' estimates of their abilities after 1, 2 and 3 years in the school and secondly by taking their separate average marks in at least 4 subjects in G.C.E. "O" level examinations in 1951, 1952 and 1953 respectively. The correlation coefficients were low (for Arithmetic 0.261 with Teachers' Estimates and 0.254 with School Certificate, for Intelligence Quotients 0.155 with Teachers' Estimates and 0.211 with School Certificate).

In Northern Ireland he considered 472 children who sat the 1948 Qualifying Examination in County Antrim and who in 1951 sat the Junior Certificate (for which he averaged the marks in 8 subjects, first given the same means and standard deviation) and later the 320 who went on to take the Senior Certificate (for which he first equated

the averages of the subjects). In contrast to the Lancashire experiment the Northern Irish Arithmetic and English papers were of the "traditional" type.

In view of the many enquiries I have mentioned there may be some value in reproducing an abridged version of Professor Wrigley's summary Tables. He was concerned really with the ORDER in which Arithmetic, Englishcor Intelligence Tests were most efficient predictors. These are shown in parenthesis ( ) below:-

Name of Researcher	Arithmetic	English	Intelligence Tests	Population size in enquiry
Stubbins 1	0.174(3)	0.232(1)	o.202(2)	154
Stubbins 2	0.133(3)	0.265(1)	0.243(2)	148
Emmett	0.330(3)	0.335(2)	0.454(1)	765
Peel and Rutter	0.118(3)	0.392(2)	0.417(1)	234
Emmett-Wilmut 1	0.424(3)	0.476(2)	0.499(1)	128
Emmett-Wilmut 2	0.408(3)	0.488(2)	0.533(1)	153
Wrigley Lancs.	0.254(1)	0.191(3)	0.211(2)	227
Wrigley N.I.	0.543(1)	0.489(3)	0.516(2)	472
"Average"	0.340(3)	0.369(2)	0.423(1)	2281

Noting the remarkable power of the Intelligence Tests, since Arithmetic and English are always part of the criterion for success, he concluded that the generally accepted view of the "order of merit" of the predicting powers of the constituent parts of the then "usual" eleven-plus examination was a sound one.

More recently, and again using "traditional" type papers in English and Arithmetic, S.C.Richardson (1956) reported on a follow up survey relating to 313 children who entered four Plymouth Grammar Schools in 1949. At the end of the first and second years of the Grammar School course, the following measures were obtained: (i) the end of year examination marks for the different subjects and (ii) marks

for everyday work in the different subjects, but not under examination conditions. The distributions of the marks were standardised and then weighted according to the Head Teachers' estimates of the value of the subject. These marks were added to form the criterion. They were then correlated with different combinations of the tests used in selection and also with Primary Teachers' estimates (in rank order and not used in the selection process.). The estimates were scaled on the total examination results to eliminate differences between schools. A correction for selectivity was applied. High correlations were found (the attainment tests were of the older 'traditionsl' type). Taking the qualifying examination as a whole, English + Arithmetic + Intelligence Quotient, the correlations were 0.863 with the first year criterion (n = 313) and 0.846 for the 2nd.year (n = 286). For IQ alone against the 1st.year criterion the correlation was 0.826. The Primary Teachers' estimates gave the best single predictor correlations (0.835 with the 1st.year and 0.791 with the 2nd.year). Note,however, the point made earlier by Sandon - the Teachers' Estimates had not been used in the process of selection.

A major research project was that carried out by A.Yates and D.A. Pidgeon and the National Foundation, which is the subject of "Admission to Grammar Schools" published in 1957. As the then Director, Professor Morris noted "The....report represents the climax of the large scale programme of research launched in 1951.". Approximately 1,200 children in Twickenham of the 1,500 who took the Authority's 1951 entrance examination, and a similar number in 1952, formed the basis for the survey. It would be impossible to summarise this work adequately in a few pages and I have, therefore, to be highly selective in the points I mention.

Two possible methods of assessing success in the Secondary Schools were considered - (a) that of obtaining a Head's assessment and scaling

it against the score obtained in standardised tests in the secondary school and (b) using examination marks from Secondary School internal examinations and scaling these against scores obtained in the entrance examinations. It was found that there was fairly close agreement, higher correlations being obtained for grammar school children when the 'eleven-plus" scores were used for scaling, while the use of objective tests taken in the secondary schools was preferable for prediction within the modern schools. (pp.75-76)

They suggested a method of scaling Primary Heads' assessments, by allocating scores obtained from Verbal Reasoning Tests in descending order to the children from the school, listed in order of merit.(pp 77-93). In their enquiry age allowances were not made in the Primary Heads' assessments, but they describe possible methods of making such allowances. (pp.64 and 92). A method of "within schools" analysis was used, and the results then pooled to yield an overall indication of the predictive values.(pp.61-74). Using 'simple' unboosted correlations, within schools and within sex they found for example:-

Predictor	1951 3rd.year group	1952 2nd.year group
Primary Head's Assessment scaled by their method	0.748	0.761
Verbal Reasoning Test	0.704	0.726
Standardised Arithmetic Test	0.659	0.676
Standardised English Test	0.622	0.683 (p.65)

With "boosted correlations within schools and within sex they found:

Predictor	1951 2nd.year	1951 3rd.year	1952 2nd.year
Primary Head's Assessment(F)*	0.909	0.893	0.882
Verbal Reasoning Q.	0.893	0.875	0.868
Arithmetic Q.	0.862	0.853	0.846
English Q.	0.859	0.838	0.845

*National Foundation's Method

There was "close agreement between the boosted correlations and those derived from the analysis ignoring schools" (p.66) - suggesting "that the effects on their pupils' progress of differences between the secondary schools involved in this inquiry were small".
In short the best SINGLE predictor was the Primary Heads' Assessment scaled by the National Foundation's method of referring to the scores obtained by the children in a standardised test of verbal intelligence. An exhaustive analysis was made of all the possible combinations of the 'single' predictors used (which included tests of spatial ability) and (p.69) "the best sequence would seem to be P.H.A.(F), V, E 2., with A. and one of the two spatial tests making equal claims for fourth place. If the Arithmetic Test was the one selected the 'boosted' correlation was found to be 0.929.(p.71).
Using a 'weighted' battery a 'boosted' correlation of 0.931 could be obtained from "the best" three tests or assessments by taking 4:P.H.A.(F), 2: E 2., and 1: Sp.1 (p.72).
Excluding Teachers' Assessments, V + E 2 + A gave a 'boosted' correlation 0.921. The resultant gain in prediction from various weighting of V, E 1 and A was found to be negligible. (p.73).
(The E.1 mentioned was a standardised English test of the objective 'completion, cross-out, tick or underline' type, while E.2 was a standardised English test allowing more freedom in the answers.)

Misallocations - Errors. (pp.144-145). Taking a Validity Coefficient of 0.85 and a 20% allocation to Grammar Courses it was calculated that for every 1,000 children 122 would be wrongly allocated (61 to Grammar and 61 to Modern schools). (Sandon had calculated that with selection at $2\frac{1}{2}$ standard deviations above the mean there would be a 1 in 3 error in the total number admitted to Grammar School with 0.95 valid correlation; - this was reduced to less than 4% error in admissions to Grammar by lowering intake mark to 2 standard deviations above mean but, as he said, "what a cost"; that of increasing the "Grammar" provision by

over $3\frac{1}{2}$ times the original number of places available.)

Yates and Pidgeon went on to show that their calculation implied that approximately 78,000 children would be "wrongly" allocated in any one year, of whom only 12,800 were likely to be re-allocated to grammar schools and 3,200 to change from grammar to modern schools (hence the 60,000 in the "wrong" schools quoted earlier). They also pointed out that some of those "wrongly" allocated would benefit, and that other avenues to G.C.E. were open besides grammar courses.

They proceeded to show,(p148/9) and to check in Twickenham, that even with a "borderzone" of a size determined statistically from the estimated validity coefficients of the Tests used, "even the best methods are likely to involve "wrong" allocations in the case of about 10% of the candidates". From the possible "limits of improvement" (p.192) they argued "it would seem desirable to make the type of school to which a child is allocated less decisive in determining the kind of course he will ultimately follow..."

On the general question of the use of selective devices in the future they comment (pp.92-93) "...no matter what system is introduced, some process of allocation will still be necessary.... The various techniques of assessment that have been developed within a tri-partite system would serve a useful purpose in such a scheme of educational guidance.....they have not been tried and found wanting.....rather....they have been expected to perform a virtually impossible task".

Earlier they had concluded "The alternatives with which the authorities are faced are either to accept the competitive situation... or so to re-arrange their system of secondary education as to disguise, blur or remove the distinction between grammar school and other types of secondary school courses. Unless this re-organisation is so radical as to remove the distinction altogether by, for example, providing a common curriculum for all secondary school children it is unlikely to solve....

the problem created by the excess of the demand for grammar school places over the supply.".

Also in 1957 there was published a most comprehensive and authoritative book covering every aspect of Secondary School Selection. This was the main title of the book, which was edited by Professor P.E. Vernon and was an inquiry by members of the British Psychological Society. "The aim is to provide a review of the relevant information, so that interested persons may be able to distinguish actual research findings from statements associated with supposition, prejudice and self-interest, and avoid some of the confusions which arise when questions which are fundamentally economic or administrative are addressed to psychologists." (p.8). One implication is that had there been enough money to provide very many more buildings, and had enough teachers been available in the immediate post-war period, the whole question of "the eleven-plus examination" might very well never have received so much prominence. The book gives complete coverage of the examination and selection process at the time and the factors leading up to its development. Details are given of the researches which I have briefly outlined. Chapter X summarises the inquiry and lists 32 recommendations.

In view of the continuing likelihood of some form of selection procedure being adopted in many areas for some time to come it is worth re-stating one or two of the Psychologists' conclusions (with which Yates and Pidgeon's contemporary findings fully agreed):-

"12. There is little to choose between the overall validities of various combinations of tests, examinations and estimates; but the intelligence test is so consistently successful that it should not be dispensed with...

"14. Even with these high validities, predictions can never approach perfection...

"15. The inevitable imperfections even of highly valid predictions

should be borne in mind in any plans for re-organising the secondary educational system. It should also be realised that the proportion of the population suited to an academic course leading to good G.C.E. results is quite small....and... depends greatly on the quality of teaching in, and the 'morale' of, the particular school.
"16. It would be preferable to abandon the terms "intelligence" test and 'IQ" in 11+ selection, and to substitute, e.g. 'general educability' or 'academic (or other) aptitude' tests. Psychologists should not claim that they are measuring purely innate ability, in contrast to acquired attainments." (pp.172 - 173)

Thus there was the recognition that the eleven-plus examination was a highly efficient (but not perfect) indication of likely performance of children in secondary schools. However, many other factors, not "examined" by the selection procedure also affect school-work,and from 1957 the emphasis changed.

Researches by Professor E.Fraser in Aberdeen, and by Professor S. Wiseman in Manchester showed the effect of the child's home and neighbourhood. Professor H.J.Eysenck and P.O.White, Dennis Child, J.Rushton, W.J.Hallworth among others, turned attention to personality ratings and their link with measures of intelligence and scholastic performance. The effects of the internal organisation of schools, the age of transfer and the arrangements made by secondary schools to smooth the passage of new entrants, came under scrutiny. Professor Nisbet in Aberdeen is just publishing an account of this latter work.

All these influences affect correlations obtained between assessments of junior school work and that years later in secondary schools. I have therefore tried to give some account of them in a later Section. (Section 5).

"Contrary Imaginations" and "Frames of Mind" by Professor Liam Hudson (1965 and 1968) reflect changing attitudes to "measures" of "intelligence". So do works like "Creativity and Intelligence" by

J.W.Getzels and P.W.Jackson (1962) and the recent lengthy and detailed articles in the"Harvard Educational Review" (Vol.39, Nos. 1 and 2, Winter 1968 and Spring 1969) in which Professor A.R.Jensen tried to show that there was a limit to the extent to which IQ and Scholastic Attainments could be "boosted" and was "taken to task" by psychologists of the standing of L.J.Cronbach, J.McV. Hunt and D.Elkind.

Nevertheless, before leaving the topic of the use of "intelligence" tests and measures of attainment to "select" for admission to academic courses two points are worth noting.

With the general move towards equality of treatment, the problem of how to deal with "gifted", very able, children becomes acute.

Professor N.R.Tempest in Liverpool is experimenting with "material which will challenge able children" in a primary school course and "which may be used by a teacher who has one or two such children in a class of forty". He has used "intelligence" tests (W.I.S.C. for seven year old children), but also gave a "creativity" test and took a reading age and teacher nominations into account,when deciding which children should be included in the class of 15 which he has drawn together for the purpose of the experiment. (Information from a letter addressed to me, dated 24th.July 1969.)

The second point, and one which is receiving increasing emphasis, in view of the "competition" and increasing cost relating to higher education, concerns enquiries conducted to try to decide which measures correlate with later success at this level. Many years ago Professor F.V.Smith (1946) reported that the aggregate obtained in the "Leaving Certificate" examination was more significantly associated with success in the first year examinations than was the "IQ" measure derived from the OTIS Higher Test. (0.30 compared with 0.07). More recently Dr.H. Hohne has reported (duplicated accounts now unfortunately out of print) on studies in the 1950's related to prediction of success in the Arts and Science faculties of the University of Melbourne. However, as Dr.

W.C.Radford in a letter to me dated 25th.July,1969, points out "At the present time in Australia some work is being done with tests..." (including Scholastic Aptitude Tests)..."to see whether the current kinds of examinations used at the end of secondary school for purposes of selection for tertiary education can be supplemented or replaced. It was only last year that the first series of tests were used and we have obviously not had time to assess their likely effectiveness."

In a recent inquiry at University level, L.A.Marascuillo and G.Gill (1967) examined the value of the Miller Analogies Test and the Undergraduate Grade Point Average for 106 successful and 64 unsuccessful Doctoral students. Neither measures were very good at distinguishing between successful and unsuccessful students.... "The variables which do discrimate seem to be related to a factor of commitment to scholarly work".

In this country Professor F.W.Warburton (1963) concluded that factors to which success in a University Department of Education were largely UNRELATED include "Verbal and Non Verbal Ability, Extraversion, Arts versus Science interests..." and went on "it would appear best in selecting potential teachers to concentrate on attainment and general culture rather than ability..."

To conclude this section, reference should be made to a survey reported by N.France (1964). This gave details of a "follow up" in Kent of a 6.5% sample of all secondary school children. This "enabled a comparison to be made over a range of $6\frac{1}{2}$ years between scores on tests of attainment and intelligence, of reasoning ability and mechanical skills and of tests timed and untimed". "Whether a reasoning test is timed or untimed seems to have little significant effect on its predictivity." (Accuracy versus Speed ?)

"The most predictive tests are tests of reasoning. Whether the content is basically literate or numerate makes little difference provided that it is predominantly in a verbal form." (p.30).
Many other tests were used - Mathematics, English and "Intelligence" at 14 as well as at 11 years of age. The Secondary Heads' Assessment at the age of 14 was scaled with a mean of 100 and a standard deviation of 15 points. It was found that the correlations were always higher for tests at 11 and tests at 14 (c.f. Yates and Pidgeon relating to Grammar Schools) than with tests at 11 and Heads' Assessments at 14, but the ORDER of forecasting efficiency was always the same - the correlation between "IQ" at 11 and "IQ" at 14 was 0.84 compared with 0.79 with Heads' Assessments. Tests of verbal reasoning proved to be more reliable than tests of less verbal skills. One important finding was the correlation of "IQ" scores after intervals of $6\frac{1}{2}$, $5\frac{1}{2}$, $4\frac{1}{2}$ and 4 years - 0.68, 0.79, 0.79 and 0.84 respectively. (France, 1964).

It is submitted that the work mentioned in this section demonstrates beyond doubt the "efficiency" and importance of standardised tests of Verbal Reasoning and English and Arithmetic in relation to probable performance in later years.

References in Section 2(b)

Amos A.D.	1931:	Examinations and Intelligence Test Forecasts of School Achievement -	British Journal of Educational Psychology. Vol.I (p.73.)
Bosomworth G.	1953:	Use of Teachers' Estimates in the Allocation of children to Secondary Schools -	Durham Research Review. Vol.4 (p.50)
Collier J.W.	1933:	The Predictive Value of Intelligence Tests for Secondary Education: -	British Journal of Educational Psychology. Vol.3 (p.65)
Emmett W.G.	1942:	An Inquiry into the Prediction of Secondary School Success. -	University of London Press. 1942. reprinted 1945.
Emmett W.G.	1954:	Secondary Modern and Grammar School Performance Predicted by Tests given in Primary Schools -	British Journal of Educational Psychology. Vol.21 (pp.126-134)
Emmett W.G. and F.S.Wilmut	1952:	The Prediction of School Certificate Performance in Specific Subjects. -	British Journal of Educational Psychology.Vol.22 (pp.52-62)
France N.	1964:	The use of group tests of ability and attainment; a follow up study from primary to secondary schools. -	British Journal of Educational Psychology. Vol.34 (pp.19-33)
Getzels J.W. and Jackson P.W.	1962:	Creativity and Intelligence -	Wiley and Sons, New York.
Hallworth H.J.	1964:	Personality ratings of adolescents; a study in a Comprehensive school. -	British Journal of Educational Psychology.Vol.34 (pp.171-177)
Heim A.W.	1954:	The Appraisal of Intelligence -	Methuen,London. (p.111)

/continued

References Section 2(b) continued

Holzinger K.J.	1928:	Statistical Methods for Students in Education.	Ginn & Co., New York. (p.121)
Hudson L.	1965: 1968:	Contrary Imaginations and Frames of Mind	Methuen, London
McClelland W.	1942:	Selection for Secondary Education	University of London Press.
McFarland H.S.N.	1948:	The Predictive Value of Standard Tests and Teachers' Estimates	B.Ed.Thesis: University of Edinburgh.
Marascuillo L.A. and Gill G.	1967:	Measurable differences between successful and unsuccessful doctoral students.	Californian Journal of Educational Research.Vol.18 Part 2(pp.65-70)
Peel E.A.	1948:	Prediction of a Complex Criterion.	British Journal of Statistics, Psychology Section No.1. part 2. (pp.84-87)
Peel E.A. and Rutter D.	1951:	The Predictive Value of the Entrance Examination as Judged by the School Certificate Examination.	British Journal of Educational Psychology Vol.21. (pp.30.35)
Richardson S.C.	1956:	Some Evidence Relating to the Validity of Selection for Grammar Schools.	British Journal of Educational Psychology.Vol.26 (pp.15-24)
Rutter D.	1950:	An Inquiry into the Predictive Value of the Grammar School Entrance Examinations.	Durham Research Review. Vol.I (pp.3-11)
Sandon F.	1936:	Selection by a nearly perfect Examination.	Annals of Eugenics (pp. 65-76)
Stubbins T.E.	1940:	The Prognostic value of the subjects of a Secondary School Entrance Examination.	British Journal of Educational Psychology.Vol 10 (pp.16-24)

/continued

References Section 2(b) continued

Smith F.V.	1946:	Predicting Success at a Tertiary Level -	The Forum of Education.Vol.4 Nos.2 and 3 Sydney (pp.12-25)
Thomson. Sir G.H.	1948:	Factorial Analysis of Human Ability.(Chap.6)	University of London Press
Valentine C.W.	1932:	The Reliability of Examinations. -	University of London Press.(p.31)
Vernon P.E. Editor	1957:	Secondary School Selection. -	Methuen: London.
Warburton F.W.	1963:	Predicting Student Performance in a University Department of Education.	British Journal of Educational Psychology.Vol.33 (pp.68-79)
Wrigley J.	1955:	The Relative Efficiency of Intelligence and Attainment Tests as Predictors of Success in Grammar Schools. -	British Journal of Educational Psychology.Vol.25 (pp.107-116)
Yates A. and Pidgeon D.A.	1957:	Admission to Grammar Schools. -	National Foundation for Educational Research: Newnes, London.

2(c) The Procedure used in 1966 to Allocate Children to Secondary Schools.

Only those children whose parents wish them to be considered for admission to Grammar Schools take the examination. All the children in the County in the appropriate age group are eligible and may be entered for the examination at the request of parents, even if Heads of Primary Schools consider them unsuitable for admission to Grammar Schools.

The overall maintained school population of appropriate age in 1966 was 1230 boys and 1163 girls, a total of 2393. Of these 748 boys and 805 girls were entered for the examination. In addition there are many Independent Schools in the area. In these the relevant age groups were 154 boys and 116 girls, of whom 148 boys and 111 girls were entered for the examination. The total age groups were therefore:

Boys	Girls	Total
1384	1279	2663

The number of candidates were:

Boys	Girls	Total
896	916	1812

Approximately 30% of these were deemed to have "qualified" for places at Grammar Schools; i.e. 413 boys (29.8%) and 372 girls (29.1%). By no means all of these accepted places, many transferring to Independent Schools.

Six elements entered into the assessments: 2 Verbal Reasoning Tests, Standardised Tests of English and Arithmetic, an Essay and a "Teachers' Ranking" Score. The written examinations were held a month apart, one early in February (2nd.) and the other in March (9th.). A Practice Test (Standardised Verbal Reasoning, but not marked) was administered some three weeks before the first day of written examination. The Verbal Reasoning Tests used were Moray House No.77 and N.F.E.R.Verbal Test 15A. For English, the National

Foundation's English Test 15 was used, and for Arithmetic the National Foundation's Arithmetic Test 15E.

For the Essay 45 minutes was allowed and a choice of five subjects were given.

Bearing in mind the work of Professor Wiseman (1949) the Essays were marked by four independent judges and the scores totalled. The totals were then standardised using the English Test and an age allowance given.

For the Primary Teachers' Ranking a method outlined by A.Yates and D.A.Pidgeon (1957)was used. Heads of Primary Schools were asked, ignoring differences in age, to draw up a list of candidates in order of merit. These were then matched (with an age allowance calculated month by month) against the quotients obtained from both of the Verbal Reasoning Tests. Additional weight was given by adding half as much again of the score so obtained.

Boys and Girls were assessed separately and orders of merit drawn up based upon the total sum score obtained for each child from:-

		Weight
1.	Verbal Reasoning Test 1. (M.H.77)	1
2.	Verbal Reasoning Test 2. (N.F.E.R.15A)	1
3.	English (N.F.E.R. English 15)	1
4.	Essay	1
5.	Arithmetic (N.F.E.R.Arithmetic 15E)	2
6.	Teachers' Ranking(based on V.R.1 + V.R.2 + $\frac{1}{2}$(V.R.1 + V.R.2))	3

These orders of merit were considered by a Selection Panel consisting of Teachers from Primary and Secondary Schools and Elected Representatives. There was no additional "Borderline Procedure", but a Review Panel met to consider individual cases of hardship or special concern. Before the examinations, parents of candidates were asked to list the grammar schools available in their order of preference. The allocations to these schools were

based largely upon the order of merit lists, though account was also taken of distances to be travelled by children on journeys to and from school.

As noted, approximately 30% of the age group were deemed to have qualified for "grammar" school courses including places at 4 city grammar schools, 2 rural grammar schools, one bi-lateral school, direct grant and denominational schools and grammar schools maintained by other local authorities within easy reach of certain rural areas of the County.

References in Section 2(c)

Wiseman S.	1949:	The Marking of English Compositions in Grammar School Selection.	- British Journal of Educational Psychology Vol.19(pp.200 - 209)
Yates A. and Pidgeon D.A.	1957:	Admission to Grammar Schools.	- N.F.E.R., Newnes, London. (pp.88 and 92)

2(d) A Review of the Relationship between Placing in the Order of Merit Lists in 1966 and Heads' Ranking in 1969.

(1) General Remarks

The work described in this Section is not as detailed as might be desirable for five main reasons:-

(a) Ideally, for a full "follow up" a much more comprehensive battery of Secondary Tests would be required in place of the single Secondary Verbal Reasoning Test used to standardise assessments obtained from the Secondary Schools.

(b) The 1969 testing and assessing programme relied heavily upon the goodwill of colleagues in a busy summer term when the general predictive value of the "eleven-plus" examination is already well documented (references in Section 2(b)) and re-organisation on Comprehensive lines is in prospect.

(c) There is really also a need for more sophisticated calculating aids than the simple "muldivo" hand operated machine available and, indeed, for more time than is immediately available to extract more detailed information.

(d) Fewer schools than expected actually assess their third year children by means of examinations and rarely by one common to all members of the age group. In several cases assessments were available on an 'A', 'B', 'C' grading scale only, which were of little use for the present enquiry.

(e) Information was available only about relatively small numbers of children in several different schools. Evidently population mobility in the area is quite high. In addition, children who were absent from the main 1966 examination and who were then assessed on "supplementary" tests were ignored, thus further reducing numbers. "Pooling" correlations obtained from such groups would be artificial, especially as it became clear that the distributions of secondary assessments were often not "normal" nor, when compared one with

another, homoscedastic. The measures taken to overcome these difficulties are described below, but the findings must be qualified in the light of alterations made to "raw" distributions of scores to enable comparisons to be made.

Multiple correlations to find "best weights" have not yet been calculated, though in this connection one may perhaps be permitted to cite Yates and Pidgeon (1957 p.73);- "proper weights.... cannot be made available for the children to whom they apply, and it does not necessarily follow that the weights appropriate for the test scores obtained by one group of children will be equally effective when applied to the scores of a different group".

Turning then to the general procedure, the summary sheets relating to all 1812 candidates were available as were nearly all the actual examination scripts (though several were for "supplementary" tests). Allowing for the choice of schools (and preferences between Secondary Modern schools were also taken into account) as many of the 1966 candidates as could be traced were noted. The Heads of the 27 main Secondary Schools involved were asked to give for each of the original candidates now in the 3rd. year of secondary schooling:-

(i) a rank in order of merit list relating to overall suitability for academic education.

(ii)a similar rank placing relating to English.

(iii) a similar rank placing relating to Mathematics.

These rankings were to be based on examination performance

relating to the whole third year age group in the Secondary school, though Heads were permitted to adjust the rank order if the examination position was not typical.

In order to standardise these rank orders, Heads were also asked to administer a Secondary Verbal Reasoning Test to all children in the third year age group. For this purpose some 1600 of the National Foundation's Secondary Verbal Test No.2 were used and most of these were marked by teachers in the schools concerned. The quotients obtained were used for scaling purposes for each of the three ranking orders given: Overall suitability; English, and Mathematics. The scaling was carried out as suggested by A.Yates and D.A.Pidgeon (1957); the quotients relating to the age group were arranged in order and 'matched' against the ranked assessments. As will be mentioned, in a few cases corrections for age were made, by calculating the average of the children in the school age group, converting the quotients to raw scores and then re-calculating quotients for individual children based upon the "raw score" and actual age of the child concerned. It was found that correlation coefficients corrected for age were almost the same as those which had not been so adjusted.

The cases of boys and girls were considered separately. Where numbers under review in a particular Secondary School were small (20 or less) non parametric correlation coefficients, Spearman Rank Order Coefficients, were calculated.

Where product-moment correlations were calculated a serious problem was encountered. Even after the scaling of the assessments using the standardised tests, it was found that many of the distributions of scores "within schools" were skewed. This was due, no doubt, to the effects of restriction of range by selection and also by the "optional" nature of the "eleven-plus" examination.

Simple correlation coefficients were calculated and are given

for information, but they are probably of little value as they would have been affected by the shape of the distribution curves.

Some method had to be found for "normalising" distributions and equating means and standard deviations if comparisons were to be made "between schools". Expressing measures as "standard scores", i.e. in terms of proportions of the appropriate standard deviation, would not affect the 'shape' of the distribution and so errors would remain in the correlation calculations. The method finally adopted was to convert all the quotients to an adaptation of the 'T' scale suggested by W.A. McCall (1939) and named after E.L.Thorndike. The method is described by J.P.Guilford (1965). The original distribution is converted (by graphical methods) to one based upon a normal curve whose tails extend from $+5\sigma$ to -5σ, with a mean of 50. The scale unit becomes 0.1 and cumulative proportions are found to determine the appropriate 'T' score. In practice virtually all scores fall in the range $\pm 3\sigma$. The standard deviation of the 'T' scores is 10. Adapting the scale by taking the mean of 50 to become a new mean of zero, T' scores can be measured as deviations from this mean. The normal Pearson product moment formula

$$r_{xy} = \frac{\sum xy}{N\sigma_x\sigma_y} \quad \text{becomes} \quad \frac{\sum T'_1 T'_2}{100N}$$

As the conversions to T scores were based upon graphs and integer measures, it was possible that the mean would not be exactly 50, nor the standard deviation 10. This aspect was therefore examined and adjustments made to the scaling to ensure that means and standard deviations differed little from these figures. Even so, in practice to ensure accuracy the actual formula used was

$$r_{xy} = \frac{\sum T_1 \; T_2}{\sqrt{(\; T_1^{\,2})(\; T_2^{\,2})}}$$

where T_1 and T_2 are deviations from the True mean.

In all it proved possible to study "follow up" data relating to 553 children, but as the numbers vary for different parts of the enquiry the appropriate figures are quoted for each section.

References in Section 2(d) (1)

Guilford J.P. 1965: Fundamental Statistics in Psychology and Education. - McGraw-Hill, New York. (pp.518-524)

McCall W.A. 1939: "Measurement", MacMillan, New York. Chap.22. quoted by H.E.Garrett "Statistics in Psychology and Education", Longman, Green & Co. New York (1958) p.314.

Yates A. and Pidgeon D.A. 1957: Admission to Grammar Schools. - N.F.E.R. and Newnes, London. (pp.88,89 & 93)

2(d)(ii) A Review of the Relationship between Placing in the Order of Merit list in 1966 and Heads' Ranking in 1969 : Overall Assessment.

It will be recalled that Secondary School Heads had been asked to provide a rank order relating to overall suitability for an academic course. On the basis of N.F.E.R.Secondary Verbal Test 2, given to the whole age group of children in the schools concerned, this was converted, using the method described in the previous section, to a quotient score used as "follow up" criterion. These scores were compared with the total marks gained in the 1966 "11+" examination. No "boosting" of correlation coefficients, to allow for selection was made, but where appropriate both distributions of scores were converted to 'T' scales as described. The following correlation coefficients were calculated:-

Table I

School (no. in sample: sex 'b' or 'g')	Product moment correlation	
	Direct	Using T' scales
B. (29b) Modern	0.739	0.778
K. (20g) Modern	0.853	0.714
L. (26g) Modern	0.510	0.468*
M. (41g) Modern	0.660	0 .643
O. (34g) Grammar	0.509 or 0.516 when corrected for age.	0.504

All the above correlations are significant if r is tested against a null hypothesis using (N - 2) degrees of freedom (Garrett 1958). The coefficient marked * is significant at the 0.05 level, the remainder at the 0.01 level of confidence.

Except for the one marked *, the following product moment correlations are not significant:-

School	r direct	r using T'
P. (21g) Grammar	0.573*	0.073
Q. (71g) Grammar	0.069	0.041

In each case the most likely reason for non-significance is homogeneity of the sample. Approximately 70% of the students have Secondary Verbal Reasoning Quotients within a narrow 12 point group of scores. In each case the same high proportion have original "order of merit" total scores within a group of 40 points.

For the smaller samples rank order correlation coefficients were calculated. These were tested for significance against the null hypothesis (Siegel.S. 1956). Details are given in Table 2 below:-

Table 2

School (no. in sample : sex)	r_s	significance level
E (13b) Grammar	0.379	not sig.
G.(10b) Modern	0.770	0.01
G.(11g) Modern	0.902	0.01
H.(25m) Modern	0.693	0.01
I.(16m) Modern	0.719	0.01
J.(12m) Modern	0.573	0.05

The last topic in this Section is to try to obtain some overall indication of the relationship over the whole range. It was notably difficult to obtain "overall" secondary assessments especially for boys. In the 6 schools listed above containing boys, information was available for only 74 in all. Combining these small samples is probably unreliable, but not to do so would really only leave school B mentioned in Table 1. Of course, all the boys and girls in the age groups in the schools mentioned sat the same Secondary Verbal Test, so standardising assessments on that test and ignoring schools should give some indication of the

overall relationship between the original mark list and later assessments. The product moment correlation calculated on this basis for the 74 boys in Grammar and Modern schools is

$r = 0.678(4)$ which is significant at the 0.01 level.

A similar calculation for all 245 girls concerned in Grammar and Modern schools gave

$r = 0.774$ also significant at the 0.01 level.

If the "within schools" analysis for girls is modified using the 'T' scale mentioned, to equalise standard deviations and means and normalise distributions a figure of

$r = 0.341$ for 213 girls is obtained (significant at 0.01 level).

This figure may be compared with $r = 0.574$ for 121 girls in 3 Modern schools and one Grammar school, whose scores had also been 'normalised' (also significant at the 0.01 level of confidence).

One may conclude that even after 3 years of Secondary schooling there is a significant relationship between ranking in the original "11+" examinations and assessment of performance in Secondary Modern and Grammar Schools.

References in Section 2 (d)(ii)

Garrett H.E.	(1958)	"Statistics in Psychology - and Education"	Longmans, Green and Co. New York (pp.200-201)
Siegel S.	(1956)	"Nonparametric Statistics - for the Behavioural Sciences"	McGraw Hill. New York. (pp.210-284)

2.(d)(iii). A Review of Placing in Heads' Order of Merit Ranking for Overall Performance in 1969 and the Relationship with Quotients obtained from the 1966 Verbal Reasoning Test No.15A.

This section deals with a similar assessment to that dealt with in 2(d)(ii), except that the 1966 measure used is the National Foundation's Verbal Reasoning Test 15A instead of the final order of merit mark list.

The tables therefore are directly comparable with Tables I and 2 above. They are:-

Table 3. ("within" schools)

Product Moment correlation coefficients between Heads' Ranking for Overall Suitability for an Academic Course in 1969 and Assessment by N.F.E.R.Verbal Test 15A in 1966. (Significance levels are shown in parenthesis; 'N.S" means not significant.)

School (no.in Sample: Sex)	r	r,based on T scales.
B. (29b) Modern	0.696 (.01)	0.704 (.01)
K. (20g) Modern	0.799 (.01)	0.683 (.01)
L. (26g) Modern	0.33 (N.S.)	0.213 (N.S.)
M (41g) Modern	0.683 (.01)	0.592 (.01)
O. (34g) Grammar	0.320 (N.S)	0.219 (N.S)
P. (21g) Grammar	-0.0167(N.S)	-0.096 (N.S)
Q. (71g) Grammar	-0.016 (N.S)	-0.100 (N.S)

The "direct" calculation of r for school M is marginally higher than that based upon the total "11+" examination marks shown in Table I. The remainder of the significant coefficients are substantially less.

Table 4. ("within" schools)

Rank order correlation coefficients between 1969 Heads' assessments for overall suitability at 1966 N.F.E.R. Verbal Reasoning

15A ranking.

School (No. in sample: sex.)	r_s	significance level
E. (13b) Grammar	0.309	not Significant
G. (10b) Modern	0.803	.01
G. (11g) Modern	0.863	.01
H. (25m) Modern	0.472	.05
I. (16m) Modern	0.51	.05
J. (12m) Modern	0,43	Not significant

If one wishes to obtain an assessment "excluding schools" the problems of pooling mentioned in the previous Section arise again. As far as boys are concerned, there is really only a small sample from a single school, B, giving a product moment correlation of the order of 0.70.

In the case of girls, the correlations can be pooled using the T scales and for schools K,L,M,O,P and Q this gives a figure of

r = 0.187 for 213 girls, significant at the .01 level of confidence.

This is to be compared with the figure of 0.341 obtained for the same girls at the end of the last Section for total "11+" marks.

2(d)(iv).A Review of Placing in Secondary Assessments for Mathematics in 1969 and Quotients obtained from the 1966 N.F.E.R. Arithmetic Test 15E .

Data from more schools was available for 1969 Assessments for Mathematics. As before the rankings for Secondary Mathematics were scaled using the Secondary Verbal 2 Test Quotients. These were then compared with Quotients obtained in the 1966 "11+" Arithmetic Test. The findings are given below:-

Table 5 ("within"schools)

Product moment correlation coefficients between Secondary Assessments for Mathematics in 1969 and Assessment by N.F.E.R. Arithmetic Test 15E taken in 1966.(Significance levels are shown in parenthesis; "N.S." means not significant.)

School(no.in sample: sex)	r	r,based on T scales
Boys		
A. (38b) Modern	0.725 (.01)	0.673 (.01)
B. (29b) Modern	0.839 (.01)	0.900 (.01)
C. (70b) Grammar	0.443 (.01)	0.412 (.01)
D. (59b) Grammar	0.111 (N.S)	0.115 (N.S)
Girls		
K. (20g) Modern	0.650 (.01)	0.840 (.01)
L. (26g) Modern	0.462 (.05)	0.460 (.05)
M. (41g) Modern	0.769 (.01)	0.650 (.01)
N. (32g) Modern	0.199 (N.S)	0.716 (.01)
O. (34g) Grammar	0.342 (.05)	0.422 (.01)
P. (21g) Grammar	0.241 (N.S)	0.384 (N.S.)
Q. (71g) Grammar	0.211 (N.S)	0.169 (N.S)

Table 6 ("within" schools)

Rank order correlation coefficients between Secondary Mathematics Assessments 1969 and 1966 Arithmetic Scores.

School (no.in sample: sex)	r_s	significance levels
E. (13b) Grammar	0.286	Not significant
F (16g) Modern	0.300	Not significant
F (9b) Modern	0.896	.01
G. (10b) Modern	0.624	.05
G. (11g) Modern	0.768	.01
H. (25m) Modern	0.385	.05
I. (16m) Modern	0.812	.01
J. (12m) Modern	0.614	.05

For overall assessment ignoring schools, again two possibilities exist. One is to use the common assessment of the Secondary Verbal 2 scaling, the other to "pool" scores from the T scores. Tables 7 and 8 give details.

Table 7

Product moment correlations calculated using T scales to pool data from various schools. 1969 Mathematics Assessments compared with 1966 scores on N.F.E.R. Arithmetic Test 15E.

Boys	r	significance level
Schools A,B,C, and D 196 boys.	0.449	.01
Girls		
Schools K,L,M,N,O,P and Q. 245 girls	0.457	.01

Table 8

Product moment correlations calculated on samples drawn from the total population considered, using the scaling on

Table 8 continued

Secondary Verbal Test 2, ignoring differences in variances and means in different Secondary schools.

	r	significance level
Boys 111 cases	0.718	.01
Girls 112 cases	0.773	.01

Of these, the lower values found in Table 7 would seem to be more reliable. They can be compared with the 0.778 for 29 boys and 0.341 for 213 girls found on the basis of the overall mark list, 0.704 and 0.187 for the same boys and girls respectively on the basis of the Verbal Reasoning Test alone.

2(d)(v). A Review of Placing in Secondary Assessments for ENGLISH in 1969 and Quotients obtained from the 1966 N.F.E.R. ENGLISH Test 15.

This section is similar to the preceding ones. The 1969 English Assessments were scaled using the Secondary Verbal Test 2 scores. The comparison is with the ENGLISH Test used in the Selection Process in 1966. The findings are given below:

Table 9 ("within schools)

Product moment correlation coefficients between Secondary Assessments for ENGLISH in 1969 and N.F.E.R. ENGLISH Test 15 in 1966. (Significance levels shown in parenthesis; "N.S" means not significant.)

	School (no.in sample;sex.)	r	r,based on T Scales
Boys	A (38b) Modern	0.744 (.01)	0.764 (.01)
	B (29b) Modern	0.657 (.01)	0.648 (.01)
	C (70b) Grammar	0.359 (.01)	0.379 (.01)
Girls	K (20g) Modern	0.183) (N.S)	0.520 (.05)
	L (26g) Modern	0.082 (N.S)	0.325 (N.S)
	M (41g) Modern	0.640 (.01)	0.539 (.01)
	N.(32g) Modern	0.612 (.01)	0.555 (.01)
	O (34g) Grammar	0.429 (.05)	0.366 (.05)
	P (21g) Grammar	0.540 (.01)	0.497 (.05)
	Q (71g) Grammar	0.131 (N.S)	0.263 (.05)

Table 10 ("within" schools)

Rank order correlation coefficients between Secondary English Assessments in 1969 and 1966 English Scores.

Table 10 continued

School(no.in sample; sex)	r_s	significance levels
E (13b) Grammar	o.492	N.S
F (16g) Modern	0.659	.01
F (9b) Modern	0.683	.05
G (10b) Modern	0.696	.05
G (11g) Modern	0.873	.01
H (25m) Modern	0.631	.01
I (16b) Modern	0.672	.01
J (12m) Modern	0.792	.01

As with Arithmetic/Mathematics, two possible ways of obtaining some "overall" picture exist. One uses the common assessment of the Secondary Verbal 2 scaling, the other "pooling" scores using the T scales. Tables 11 and 12 give details.

Table 11

Product moment correlations calculated using T scales to pool data from various schools. 1969 English Assessments compared with 1966 scores on N.F.E.R. English Test 15.

Boys	r	significance level
Schools A,B and C 137 boys	0.547	.01
Girls		
Schools K,L,M,N,O,P and Q. 245 girls	0.411	.01

Table 12

Product moment correlations calculated on samples drawn from the total population considered, using the scaling on Secondary Verbal Test 2, ignoring differences in variances and means in different Secondary schools. 1969 English Assessments compared with 1966 English Quotients.

Table 12 continued

		r	significance level
Boys	111 cases	0.658	.01
Girls	112 cases	0.786	.01

Again the lower values in Table 11 would seem to provide safer bases for judgments. Table 13 summarises the T scale assessments found in the previous sections so that they can be compared.

TABLE 13

Summary of product moment correlation coefficients using T scales to equalise means and standard deviations, relating to direct relationships between Secondary assessments and "11+" assessments:

Assessments Compared	Sample Size	Pearson 'r' Coefficients Boys	Girls	Significance Level
Secondary "Overall" suitability for academic course and (a)"11+" Total marks	29	0.778		.01
	213		0.341	.01
(b) V.R.15A Quotient	29	0.704		.01
	213		0.187	.01
Secondary Mathematics Ranking and"11+" Arithmetic	196	0.449		.01
	245		0.457	.01
Secondary English Ranking and "11+" English	137	0.547		.01
	245		0.411	.01

All the correlations are significant at the .01 level of confidence. In the case of girls there is little to choose between the sizes of the coefficients. In the case of boys the sample sizes vary so much that little comparison can usefully be made.

2(d)(vi). A Note of the Relationship between Scores obtained from the 1966 Essay and Subsequent Secondary Assessments for ENGLISH.

It may be of interest to have a brief note on the relationship between the 1966 "11+" Essay scores and Secondary Assessments for ENGLISH. It will be recalled that the Essays were marked by 4 independent judges and the total score obtained standardised using the ENGLISH Test as a basis. "Within Schools" correlations are quoted below:-

Table 14

"Within Schools" correlation coefficients calculated between ranking based on the "11+" Essay Scores in 1966 and Secondary English Assessments in 1969. (Significance levels are shown in parenthesis; "N.S" means "not significant".)

School (sample size;sex)	Product moment correlation:r	Rank Order correlation r_s
G (10b) Modern		0.636 (.05)
G (11g) Modern		0.293 (N.S)
H (25m) Modern		0.535 (.01)
I (16m) Modern		0.679 (.01)
J (12m) Modern		0.827 (.01)
M (41g) Modern	0.399 (.01)	
O (34g) Grammar	0.250 (N.S)	
P (21g) Grammar	0.330 (N.S)	

As far as the Secondary Modern Schools are concerned it can be seen that the relationship between Essay Scores and subsequent performance in English is generally highly significant. The samples were really much too limited to draw any valid conclusions as far as Grammar Schools are concerned.

2(e) A Note on the Stability of Verbal Reasoning Quotients.

Verbal Reasoning Quotients are not to be confused with "Intelligence", This may seem a trite comment, but the point has certainly been hammered home, even in the "popular press" recently. Articles like "The Expanding Mind" (Pidgeon 1969a) and "Student....(G.C.E.'O' level)....Choice linked with Social Class, I.Q" (Miles 1969), add fuel to the controversy which has a central role in the change to comprehensive organisation of Secondary schooling and also the "streaming/non streaming" discussion. "The use of intelligence tests has been dominated by the idea that intelligence was an inherited characteristic of the mind that could be reasonably accurately measured." (Pidgeon 1969b). He goes on to propound "A changed view". "By "intelligence" Mr.Pidgeon here evidently means what he and other writers have called "potential ability". But, as was stated long ago, "potential ability" is in this respect on the same footing as "potential energy". Neither can be directly measured; but (with varying degrees of accuracy) their amount can be estimated. That such estimates are reasonably exact in the case of "intelligence" has now, I should have thought, been fully demonstrated." (Burt 1969a).

"These measures are only signposts - indicating essential principles on which the growth of the intellect depends. We are affected by social context as well as innate factors." (Hudson 1968).

"A person's test score reflects a number of different factors... ...The major...variables (are)...the individual's inherited potential, both in terms of (1) general intelligence and (2) specific capacities for training..., plus the environment in which the organism has developed." (Goslin,1963).

"The clinical evidence has indicated, in effect, that there are two components in intelligence test performance...(A) innate potential

and ....(B) the functioning of a brain in which development has gone on... two different meanings of 'intelligence'." (Hebb,1949).

The method of measurement is important too. David Elkind noting Piaget's work, stresses differences in terms of qualitative as well as quantitative improvement in follow up studies of "intelligence" in children. (Elkind, 1969).

Professor Jensen's view that less than 20% of the variations in intellectual ability are due to environmental factors has recently been upheld by Sir Cyril Burt (1969b).

Professor Bloom's (1964) enquiries in the United States give evidence of the considerable consistency of intelligence test scores. As Jensen (1968) says, "Beyond the age of 10, regardless of the interval between the tests, the obtained test - re-test correlations fall in the range between the test's reliability and the square of the reliability".

In this country, N.France's major enquiry in Kent reported in 1964 and mentioned earlier at the end of Section 2(b), showed correlation of I.Q. scores over the years to be of the order of 0.77. The findings of the present survey, given below, are of the same magnitude. It must be appreciated that correlations of this order allow for considerable variation in individual cases. As Michael Armstrong has said, quoting Professor P.E.Vernon, "One should never think of a child's I.Q. (or other test result) as accurate to one per cent. Rather an I.Q. of say, 95, should be thought of as a kind of region or general level...".(Armstrong,1969).

Turning then to information available from the present survey, it is possible to provide direct evidence from Verbal Reasoning Tests prepared by the same Research Organisation.

If one compares Verbal Reasoning Quotients obtained from the

National Foundation's Test 15A in 1966 and Quotients obtained from the same children in 1969 using N.F.E.R.Secondary Verbal Test 2, one obtains the following product moment correlations:

Table 15

Quotients obtained from Secondary Verbal Test 2 (1969) compared with Verbal Reasoning Test 15A in 1966.

	r	Significance level
Boys (no. 111)	0.713	.01
Girls (no. 112)	0.874	.01

The average drop in Verbal Reasoning Quotients over the intervening 3 years is:

Boys: 4.74 points
Girls: 5.14 points.

References in Section 2(e)

Armstrong M. (1969) : "Verdict on the Facts" Advisory Centre for Education, Cambridge.(p.36)

Bloom B. (1964) : "Stability and Change in Human Characteristics" New York, Wiley. (quoted by Jensen. 1968)

Burt. Sir Cyril. (1969a): Letter to the Editor of "Educational Research News", no.7. Sept.1969. N.F.E.R.,Slough, Bucks.

Burt. Sir Cyril. (1969b): 'Recent Studies of Abilities and Attainments" in the Journal and Newsletter of the Association of Educational Psychologists (quoted by Sir William Alexander in "Education" Vol.134. No.12. Sept.1969. p.1142)

Elkind. D. (1969) : Harvard Educational Review. Vol.39.No.2. Spring 1969. (p.327)

France N. (1964) : "The use of group tests of ability and attainment; a follow up study from Primary to Secondary School." British Journal of Educational Psychology. Vol. 34.(pp.19 - 33)

Goslin D.A. (1963) "The Search for Ability: Standardised Testing in Social Perspective". Russell Sage Foundation, New York.(pp.151-152).(quoted David E.Lavin "The Prediction of Academic Performance." Science Editions. John Wiley and Son, New York.1967(p.48)

Hebb. D.O. (1949) : "The Organisation of Behaviour". John Wiley, New York.(chap.11), reprinted in "Intelligence and Ability." Stephen Wiseman,Editor. Penguin no.U.P.S.5,1967.Harmondsworth,Middx.(pp.250-251)

Hudson L. (1968) : "Frames of Mind" Methuen,London (p.88)

Jenson A.R. (1968) : "How much can we boost I.Q. and Scholastic Achievement" Harvard Educational Review, Winter 1968. Vol.39.No.1 (p.60)

Miles. H.B. (1969) : "Student choice linked with Social Class I.Q." Article in "The Times" Aug.25th.1969(p.3.)

Pidgeon D. (1969a): "The Expanding Mind" Article in "The Sunday Times Magazine".March 2nd.,1969 (pp.38-41)

Pidgeon D. (1969b): "Intelligence: A Changed View" Article in "Educational Research News" N.F.E.R., Slough, Bucks. no.6.May 1969.

3. Study of the relationship between Accuracy and Speed in the 1966 Tests and Subsequent Performance in Secondary School.

(a) Introduction

It has been made clear in the General Introduction that one of the chief purposes of this enquiry was to see if there was any simple and readily observable relationship between assessed Accuracy or assessed Speed in attempting the 1966 papers and subsequent ranking in Secondary Schools.

There are really three main features, all interacting with each other and with general ability attainment. These are:-

(a) Accuracy of work, which is most conveniently assessed by considering the proportion of correct responses out of those attempted,

(b) Speed of work, which could be assessed either as the proportion of questions attempted out of the maximum possible number, or measured by noting the "last number attempted" i.e. how far through the paper the child apparently progressed. In this study, the first of these possibilities has been chosen since it ties up most readily with the assessment of accuracy. However, in a few cases, as a separate measure, the other possibility has been used as indicated in (c) below -

(c) Tendency to "guess" or "confidence". One may reasonably assume that if a child reads a question and does not mark the paper in an attempt to answer it, he was not even confident enough to "guess". (There may, of course, also be an element of guesswork in incorrect answers, but this could probably be accepted as a genuine constituent of inaccuracy.) Preparedness to "risk an answer" may therefore be assessed by considering the number of questions answered as a proportion of those read - i.e. compared with the number of the last question attempted. One cannot be sure, of course,

that every question "in the middle of the paper" has been read. Some error is therefore bound to enter into such a measure, particularly if the last question on the paper is an "easy" one, in order that the child shall finish in a happier state of mind. Additionally, of course, where questions are printed in "blocks" most of a "block" may be omitted if the first one or two are found difficult. Despite these reservations, it was considered that such a measure might provide useful information.

In the review of previous literature which follows, an attempt has been made to distinguish the three elements mentioned above.

When one is considering performance on a timed test and intending to use the number of the question finally reached as a measure, thought must be given to the relative difficulty of items in the early part of the test compared with later ones. Accordingly I wrote to both Moray House and the National Foundation to seek information relating to the relative gradients of difficulty of the tests used in 1966. As they were nearly all N.F.E.R.Tests, these were a major concern. Thanks to the kindness of the Director and a Senior Research Officer,Mrs.James, I was able to spend an afternoon at the Headquarters of the National Foundation studying the construction of the tests which had been used in 1966 and the relevant information relating to the facility and discrimination of each item used was made available. While a steady increase in difficulty was possible in the Arithmetic Test, which is timed in separate sections, the bringing together of groups of items with the same underlying principles made this more difficult to achieve in the English and Verbal Reasoning Tests. The progression along the gradients of difficulty here occurred within each block of questions, so that the first questions in a new block were sometimes considerably easier than the last questions in the preceding one.

This, of course, does not affect calculations when the test is considered as a whole. Professor Liam Hudson in his study, mentioned earlier, of students' choice of Arts/Science specialisation made a good deal of use of accuracy on the first half of the "Intelligence" Test compared with that achieved on the remainder. Of course, the halves of the test he used were timed separately, but an inspection of the pattern in the Verbal Reasoning Tests used in the present enquiry (where the halves were not timed separately) revealed that in general the children were far more accurate when answering the first 50 questions. This coupled with the fact that questions were far more frequently omitted in the second half of the paper could be construed as supporting the views (a) that generally the later questions were more difficult and (b) the very vast majority of the children worked steadily through the paper reading each question before attempting the next. This latter assumption is relevant to the point made earlier, that using the number of the last question attempted as one of the measures of speed for a "confidence" calculation depends upon the belief that all the previous questions have been read and considered by the child.

3. The Relationship between Accuracy and Speed and Subsequent Performance

(b) Review of Previous Literature

Almost as soon as group tests of ability were developed (principally during the first world war), the popular saying "slow but sure" was used to question the findings. G.M.Ruch and W.Koerth (1923) tested subjects using the Army Alpha Form 7, checking answers; first allowing the standard time limit, secondly allowing double that time,and finally with unlimited time. The correlations showed virtually no change in the order with respect to intelligence whether the tests were timed or not; they were 0.966 with twice the "standard" time and 0.945 with unlimited time. In this experiment though, the groups contained a wide range of ability and it was not really very likely that the less able would have been capable of improving their scores to such an extent as to overtake the more able.

In 1924 Bernstein reported on the first major study of this problem as it affected children. He used a large number of specially constructed tests with 3rd. year boys at a London Central School. As he was a teacher at the school he was, as he said, "in the years 1920-1921 able to give ample time to the study and also to observe the boys in the course of their ordinary school work". He gave "Leisure" and "Haste" tests - in the latter there was more work to be done than could be managed by all but the strongest subjects. His results showed the two sets of scores to be correlated equally with Teachers' estimates of intelligence and of "slowness". The crucial measure, the difference between the"Leisure" and "Haste" scores, L - H, showed scarcely any relationship with the other measures. There was no correlation either with perseveration, so this did not appear to be a factor in slowness. He therefore rejected the view that there was a separate special ability for speed in intelligent acts.

These findings were queried by J.D.Sutherland (1934). He noted that "speed of work" was supposed to have little effect in the "leisurely" tests, but the time allowed for both kinds was limited to 30 seconds, there being fewer more difficult items in the leisure series. As Sutherland pointed out, "the mere knowledge that time was short would tend to destroy any leisurely atmosphere". In any case there seems to have been far too much to do (in one of the "leisure" tests about 80 words had to be read and then 8 written to complete the item, so it seems unlikely that 11 - 14 year olds would reach their ceiling and exhaust their responses in the time allowed). Sutherland gave some of the leisure tests to "normal" children and not one of them completed the items within the given time. Extra time enabled all of them nearly to double their scores. Sutherland therefore rejected the evidence as being neither for, nor against, a speed factor. He himself tested 170 Borstal Boys (i) to obtain the level of intelligence under conditions of ample time and (ii) to find the time taken for a number of problems. With the performance tests, the times taken to do a few tests of moderate difficulty (solved by nearly all the group) gave a good measure of intelligence. He found that, if making maximum effort, the time taken to solve the problem seemed to be entirely governed by the number of steps required. (This view has since been upheld in enquiries relating to mental problems.) Sutherland quoted Spearman ("Abilities of Man" 1927, MacMillan, London) "Some people exhibit a preference for speed rather than accuracy in the uncontrolled situations of everyday life.". Commenting on the fact that previous work showed that increasing the time allowances in time limit tests of intelligence did not alter the rankings for intelligence, Sutherland went on "This indicates that the rate of work in these tests is determined by the level of intelligence - yet popular opinion is almost unanimous in asserting that a separate ability for speed exists. Probably in everyday life people do exhibit a speed factor because they have natural speeds at which they prefer to work.".

"In these (intelligence) tests the natural speeds are overruled by the condition of working at maximum pressure. In addition, naive defence mechanisms may preserve the belief in a speed factor."

The suggestion that speed of working and "intelligence" were closely linked came under scrutiny again when Slater (1938) studied 450 children in a Willesden school in 1934 - 35. He described a method for measuring speed rates by which the children worked at their own pace and recorded for themselves the time they spent on each problem. The speed rate did not show any close relation with measures of intelligence obtained either from Verbal or Non-Verbal Tests, whether the time for the test was limited or not. Among the group tested, the children when left free to set their own pace tended to adopt consistent speed rates and adhered to these when working on problems of different degrees of difficulty and different types. Their speed rates could not be considered to depend only upon the amount of their general intelligence.

More recent work was reported by Aubrey J.Yates (1966). He studied 95 first year Engineering students in Western Australia, given the progressive matrices (1947) test with initial and final time limits; the Nufferno Level GL/2C/46 test with a time limit and an objective (mathematical)test with initial and final time limits. The Matrices revealed a slow working group handicapped by the time limit. On the Level test, this group completed fewer items within the time limit, but obtained as high a Level score as the faster working students. The group was also shown to be handicapped by the time limit placed on the objective test. He concluded that a Test like the Progressive Matrices, which confounds level and speed of intellectual performance is likely, if given with limited time available,to underestimate the intellectual capacity of a small but significant group of students. Thinking that there might be some link with personality - Introversion/Extraversion - he carried out

a similar experiment with a group of 9 year old children at a school in Perth. Yates had noted the suggestion (H.J.Eysenck 1947) that Extraverts may work quickly but inaccurately while Introverts might be slow and accurate. Yates therefore also used the Junior Maudesley Personality Inventory. He was testing the theory that on the Matrices a slow but accurate group of children would be identified who would benefit markedly by the allowance of more time. As he expected, this happened, 13 out of 100 children forming such a group. These were also handicapped on the timed arithmetic test, but markedly improved their scores (77% now "passing") when the time limit was removed. He noted no significant differences relating to either neuroticism nor extraversion measures.

On this complex problem, Professor P.E.Vernon, (1950), commented; "These findings (about Speed, Power and 'g') raise extremely difficult problems for intelligence and psychological testing, although, in view of the high correlation of most time-limit with power scores, they do not justify the criticism often made of tests for children. Allowing that level and speed can be distinguished under certain conditions, should we regard them as additional sub-factors constituting as it were another dimension to our Figure 2 ? Or should we regard level scores as the most appropriate measure of any general group factor, speed being a distinctive factor ? If so, is speed general to a great variety of tests, or largely specific, or can it be reduced to certain other better established factors W,P,M, or to ease versus difficulty ?

"In the writer's view, the solution should depend more on practical considerations than on Factor Analysis alone, although further intensive investigation of practice, difficulty, carefulness, speed and other factors is urgently needed. For the time being we shall obviously continue chiefly to use time-limit tests, for the sake of convenience..... "Thus in 11+ selection, the grammar schools presumably require a modicum of speed-at-work in addition to power."

Eysenck (1966) has cited evidence that while speed of reaction in simple tasks is not related to general intelligence, speed of information processing, correlates significantly with I.Q. measures. However, even this aspect has complications, since it has been shown, at least among older students, that some are fast "starters" while others only get into their stride after considerable delay. In contrast, a third group are consistently fluent while some are constantly inhibited. (Hudson 1968). That personality traits were likely to affect both speed and accuracy and hence raw scores in Verbal Reasoning Tests has been suggested by other enquiries. P. Ley and others (1966) proposed that test scores should be studied at different intervals of time as the test progressed since it seemed likely that extroverts "start fast" and build up inhibition, leading to subsequent errors or omissions. In addition, correlations of the order of 0.6 between accuracy and measures of emotional stability have been found. (Hudson L. 1961).

That the same test could act in different ways was pointed out by A.W.Heim. "A comparison of 4 sets of progress curves revealed that the same test, given with the same instructions, time limit etc., could be for one group principally a test of speed - in the sense that those attempting most questions automatically gain the highest scores - and for another group,primarily a test of power - in the sense that some of those attempting many questions obtain lower scores than some of those attempting few questions."(Heim 1954).

In 1956, Duncan Howie concluded: "Accuracy measures (low error scores) apart from saturation in an ability factor identifiable with Vernon's V.Ed. showed positive saturations in a speed factor and defined no further factor. Thus with ability constant, accuracy appeared as the converse of speed. The speed factor appeared as a rather broad factor in a variety of test media, indicating a difference trait of TEMPO in functioning."
Howie gave tests to 158 boys in schools in the Armidale District, whose mean age was 12 years,2.3 months, with Standard Deviation of 5.35 months. He found a reasoning factor underlying error scores in certain tests and taken as identifiable with the factor of "carefulness" indicated by Fruchter. (see page 71 below). In another investigation (the one reported in the article listed) on a group of children of similar age and scholastic grade he found evidence of a speed factor in "habitual and routine operations".
"It is a not unlikely possibility that, where the level of ability in a task is constant, a slower tempo may permit more careful consideration of what is required in the test items, and a faster tempo, less." He was merely asking if the accurate worker tended to work more slowly (keeping the ability level constant) and he found (using accuracy scores in Word Knowledge, Spelling Recognition, Thurstone's Letter Series and Pedigrees, Addition and Subtraction and Number Checking) that they did. "This seems quite simply to be a degree of error

possibilities in relation to output...." (Howie D. 1956)

Before turning to give greater consideration to the question of accuracy it is well worth noting a point made by T.P.Lele in the course of an enquiry in Dundee in 1952. Having studied 177 children in the first years of 3 secondary schools he found clear indication that by extending the time limit there is no appreciable change in the rankings of the individuals (for "intelligence"). In this case there was an ability range of 48 points, but, on a restricted ability sample, the correlation tended to decrease as the difference between the abilities becomes smaller and he noted the danger then, of relying upon "IQ" measures obtained from speed-timed tests when comparing pupils of almost similar levels of ability. (As, for example, at the "borderline" stage of a selection procedure.)

The possibility that a factor of Carefulness could be isolated and identified was investigated by Benjamin Fruchter. He constructed 4 tests - pencil and paper, similar to navigators' work, plotting and reading scales. In 1945 he tested 354 aviation students. The directions did not stress either speed or accuracy to the exclusion of the other. Two scores were obtained from the same administration of each test - the number of correct responses and the number of errors. Error scores had considerable range and variability. It was noted that when correct and error scores were correlated the coefficients were relatively low, indicating considerable independence of the two scores. He suggested that the error scores might be fundamentally different from the correct response and should be treated separately in the analysis. Factor analysis was used with a number of reference tests. "Right" and"wrong" scores were treated separately (otherwise they would be experimentally dependent, from same set of responses). Two matrices were prepared,one containing the 4 "rights" scores of the carefulness tests against the 11 reference tests, the other the "wrongs" scores similarly. Each matrix was factored by the centroid method. The six factors extracted were rotated graphically to "psychologically meaningful

positions". Inspection of the factor plots showed little evidence of obliqueness and orthagonal reference was maintained. Rotated Factor I was interpreted as "numerical facility" and all the "rights" scores of the carefulness tasks have numerical content. "Wrongs" scores do not have significant loadings on this numerical factor.

Rotated Factor III was interpreted as General Intellectual Factor - "g" + Verbal. Only one of the "rights" scores of the carefulness tests had an appreciable loading (Directional Plotting 0.44; "wrongs" incidentally 0.14) on this factor.

Rotated Factor VI from the "wrongs" battery was interpreted as a Carefulness Factor. The corresponding factor in the "rights" battery was not interpreted since only one test had an appreciable loading on it. (Complex Scale Accuracy 0.42). The loadings from the "wrongs" battery were plotting 0.64; complex scale accuracy 0.60; plotting accuracy 0.57 and Directional plotting 0.36. These were significant loadings and the postulated trait of "Carefulness" in the construction of the tests was so identified.

The question was whether the error score of a speeded test contained sufficiently unique variance to justify its treatment as a separate variable and the error scores of the 4 experimental tests used in this analysis did define a new factor. Thurstone (1938) isolated a similar factor. His tests were scored either % wrong or % omitted, and he said "The factor involved may be concerned with accuracy or caution".

Later, Fruchter noted that yet another factor could be obtained (which might be unique to error scores). This was labelled "sequential reasoning" and he said, "If this is the correct interpretation it may well be related to the reasoning by elimination which is sometimes used to select answers to multiple choice items". (Fruchter.B.1950).

Using measures of Accuracy for prediction came into the enquiries undertaken by Hilde T. Himmelweit and Arthur Summerfield, just after

the war. They undertook a comprehensive study of 232 students in the years 1947 - 50 testing cognitive ability, achievement, motivation and personality characteristics, to assess possible correlations with degree classes obtained. The index of accuracy (assessed indirectly from the cognitive tests as the ratio of WRONG responses to the total number of responses) correlated (-0.390) significantly at the 1% lewel. They made special note of the fact that the immediate post-war entry put pressure on the usual examinations; interviews were brief and there was little recent scholastic experience. Their "pen picture" of the successful student is of interest..."A person of high verbal ability, able to understand material presented in graphical or chart form, well informed on topics not related to his subjects, relatively free from tension, able to work at high pressure and with good powers of concentration - tends to avoid guesswork, preferring to leave unanswered questions about which he is not certain". (Himmelweit, Hilde T. and Summerfield, A.(1951)

The only direct attempt to associate "error" scores with subsequent performance in secondary schools was made during the investigation by S.C.Richardson (1956), to which reference has been made earlier. It will be recalled that in the investigation in Plymouth highly significant correlations had been found between the scores of 313 children who took the eleven-plus examination in 1949 and their assessment in grammar schools one and two years later. In addition to the general study, following a suggestion made by Professor P.E.Vernon, the error scores on "intelligence" tests for 155 boys were considered. The error scores were, of course, a negative index of accuracy and so the signs of the correlation coefficients found were reversed. The criterion of "success" in secondary schools was based upon examination marks and also marks for everyday work. For the boys, correlations of 0.774 and 0.761 were found between error scores and the first and second year criteria.

The value of accuracy was also confirmed during the enquiry relating to the performance of student teachers, also mentioned earlier. It was found that accuracy in a verbal intelligence test correlated 0.203 (significant at the 0.05 level of confidence) with the marks obtained in the final Theory Papers and that accuracy in a Vocabulary test correlated 0.258 (significant at the 0.01 level of confidence) with the Final Theory marks. (Warburton F.W.,1963)

In his book "Contrary Imaginations" and in his research thesis, Professor Liam Hudson used measures of Accuracy derived from a high level intelligence test. This was A.H.5. and he quoted A.W.Heim (1947). In a reference,(pages 77 and 78), to speed and accuracy, "The results suggest that the people who work naturally fast tend also to be mentally superior to the slower thinkers. However, the size of the correlation coefficients - no less than experience with people in everyday life - suggest that the exceptions to this are numerous. In fact, many subjects with a high score on the test were below average on speed and quite a few of the quick ones scored lower than average. The times on both tests (A.H.4 and A.H.5) correlated 0.588. This agreement was due more to the quick subjects being consistent in their speed rather than any marked consistency among the slow ones. On both tests, scores with unlimited time correlated negatively with the time taken - i.e. on both tests, the longer the time taken, the smaller the score..... A.H.4 discriminates speed rather than intelligence (correlation Limited: Unlimited time 0.411) but A.H.5 regulated speed to a comparatively minor role."

A.H.5 as a test is scored separately in two halves. Liam Hudson derived separate accuracy scores for each half and used "worse" accuracy for the lower score and "range" accuracy for the difference between them. Professor Hudson found potential "Arts" specialists among other things, "erratic" on the intelligence test, "accurate at times and slapdash at others" (p.22). However, "the young physical

scientist...was usually consistently accurate". In addition (footnote p.27) "I found poor accuracy scores (range accuracy) to correlate with interview ratings of emotional instability. Some boys seemed to express their instability by being consistently inaccurate; others by being inaccurate only in patches".

His main discovery lay in detecting individual differences from which he constructed two new classifications for ways of thinking - "convergers" and "divergers". The "convergers" were found to be much more accurate on the intelligence test. (p.69; in this case "worse" accuracy.) (Hudson L.1965 and 1961).

The implication of this work is closely tied up with the changing views on the nature of intelligence and changing attitudes to education. A good deal of academic education at school has been, until recently, based upon "conformity", learning from the authoritative statements of teachers. To a great extent experiment, "creative" thinking and "unusual" reactions have been discouraged, particularly for children of secondary school age. Standardised objective "intelligence" or verbal reasoning tests in which only the generally found answer is treated as "correct" probably tend to favour "convergent" thinkers. If this is so, as teaching methods change, it seems likely that lower levels of correlation will be found than have been customary in the past, between quotients obtained from these tests given at the age of eleven and later "success" in secondary schooling.

Before concluding this Section two comments on "Guessing" should be noted. L.J.Cronbach (1952) has pointed out that doubt may range from slight lack of confidence to total ignorance. He cites experiments to show that if people know they are to be penalised for guessing some are hesitant but others answer freely even when in doubt. He suggests that individual differences in this tendency can be eliminated by directing everyone to guess, but this introduces

large chance errors. Following studies made of the reliability of tests given with "guess" and "do not guess" directions he concluded that despite errors caused by over-timid people and those who will guess whatever the instructions, there is statistical superiority for "do not guess" rulings. This is yet another factor affecting scores on standardised objective tests.

Fairly recently A.W.Heim and K.P.Watts (1967) gained the impression that the naval ratings they were testing were loath to guess in open-ended tests, though they evidently had no objection to doing so in the less vulnerable situation of multiple choice of answers. They also commented on the connection between poor spelling ability and guessing "e.g. 'Posterity'. "The standard of living in Great Britain in the last ten years has gone up - this is POSTERITY."

To sum up this section, it is submitted that the literature reviewed shows:-

(a) Time limits for tests may well mask true ability levels indicated when children of approximately the same "intelligence" are being considered.

(b) With ability constant, accuracy seems related purely to speed and thus to the degree of error possibilities arising.

(c) The 'Carefulness' factor can be associated with the probable difference trait of tempo.

(d) That, especially in conditions where the 'normal' response is expected, accuracy seems to correlate positively with later success, and

(e) "Guessing" will introduce a sizeable proportion of chance errors in the scores obtained from multiple choice objective tests. This factor will be increased by failure to give specific instructions on the point. It is also complicated by poor standards of spelling - and changes in idiom. On this last point it may be noted that in the groups studied in the present enquiry a frequent "wrong" response was "Girls RAVE dresses" instead of "WEAR".

References in Section 3(b)

Bernstein E.	1924:	Quickness and Intelligence	- British Journal of Psychology.Monogram Supplement No.7.
Cronbach.L.J.	1952:	Essentials of Psychological Testing	- Harper and Bros. - New York (p.90)
Eysenck.H.J.	1947:	Dimensions of Personality	- Routledge and Kegan Paul. London.
Eysenck.H.J.	1966:	Intelligence Assessment- A Theoretical and Experimental Approach.	British Journal of Educational Psychology Vol.37.(p.86)
Fruchter B.	1950:	Error Scores as a Measure of Carefulness	- Journal of Educational Psychology. Vol.41 (pp.279-291);and later Psychometrika(1953) Vol.18(pp.257-265)
Heim. A.W.	1947:	An attempt to test High Grade Intelligence.	- British Journal of Psychology. Vol.37 (pp.70-80)
Heim.A.W.	1954:	Appraisal of Intelligence:	Jarrold. (p.120)
Heim.A.W. and Watts K.P.	1967:	An Experiment on Multiple Choice v. Open Ended Answering in a Vocabulary Test.	: British Journal of Educational Psychology. Vol.38 (p.346)
Himmelweit.Hilde T. and Summerfield A.	1951:	Student Selection; An Experimental Investigation II.	- British Journal of Sociology.Vol.2 (pp.59-75)
Howie D.	1956:	Speed and Accuracy from the Viewpoint of Factorial Analysis.	- The Australian Journal of Psychology No.8. (pp111 - 118)
Hudson L.	1961:	Arts/Science Specialisation:A Study of Differential Aptitude.	- Ph.D.Thesis, Cambridge 1961.
Hudson.L.	1965:	Contrary Imaginations	- Methuen, London.
Hudson L.	1968:	"Frames of Mind"	- Methuen,London(p.102)
Lele T.P.	1952:	An Investigation into the Relationahip between Intelligence, Nature of Test Items and the Speed of answering them.	- Ed.B.Thesis. Dundee.

/continued.

References Section 3(b) continued

Ley P.	1966:	"The Relationships between Intelligence, Anxiety, Neuroticism and Extraversion." P.Ley, M.Spelman, Ann D. M.Davies, S.Riley.	British Journal of Educational Psychology. Vol.36 Part 2.(p.191)
Richardson S.C.	1956:	Some evidence relating to the Validity of Selection for Grammar Schools.	British Journal of Educational Psychology. Vol.26. (pp.13-24)
Ruch G.M. and Koerth W.	1923:	Power vs.Speed in Army Alpha.	British Journal of Educational Psychology No.14(p.193)
Slater P.	1938:	British Journal of Psychology.No.29(pp.55-68)	
Sutherland J.D.	1934:	Speed Factor and Intelligent Reactions.	British Journal of Psychology No.24. (pp.276-294)
Thurstone L.	1938:	The Perceptual Factor	Psychometrika.No.3. (pp.1 - 17)
Warburton F.W.	1963:	Predicting Student Performance in a University Department of Education.	British Journal of Educational Psychology Vol.33 (pp.68-79)
Vernon P.E.	1950:	The Structure of Human Abilities.	Methuen, London, (p.84/85)
Yates.A.J.	1966:	The Relationship between Level and Speed on two Intelligence Tests, and Speed and Personality Factors in the Intellectual Performance of Young Children.	British Journal of Educational Psychology Vol.36(pp.166 and 312)

3(c). A Review of the Relationship between (i) Accuracy and (ii) Speed of response in the 1966 Tests, and placing in the Order of Merit Lists for English and Arithmetic in 1966.

This brief Section gives information relating to the same groups of 111 Boys and 112 Girls mentioned in Section 2(d) above. Product moment correlation coefficients were calculated. Accuracy is measured as percentage correct of those attempted. Speed is measured simply as the percentage of questions attempted out of the total possible. The criterion used was the Quotient obtained in 1966 for English (N.F.E.R.English Test 15) or Arithmetic (N.F.E.R.Arithmetic Test 15E).

Table 16

Sample	Subject	Correlation Coefficients (r)	
		'Accuracy'	'Speed'
112 Girls	English	0.898	0.800
	Arithmetic	0.935	0.733
111 Boys	English	0.803	0.807
	Arithmetic	0.927	0.453

All these coefficients are significant at .01 level of confidence when tested against the null hypothesis using (N - 2) degrees of freedom.

It is clear that in general it was "accuracy" rather than "speed" which determined the placing in the order of merit lists, particularly for Arithmetic, for the children considered here. If there is any element of spurious index correlation affecting these calculations, presumably they are all affected similarly. The almost equal correlation coefficients for "accuracy" and "speed" in attempting the English paper are worthy of special note, particularly as spelling had to be correct in 21% of the questions. A generally low level of spelling, which was not apparent in the samples, would, of course, have depressed all the "accuracy" scores.

3(d)A Review of the Relationship between (i)"Accuracy" (ii)"Speed" and in some cases (iii) measures of "confidence" in answering the 1966 Tests and placing in the Heads' Ranking Lists in 1969.

The method of obtaining the criterion measures in 1969 and scaling them using the N.F.E.R.Secondary Verbal Tests given to the whole age group in the schools concerned, has already been described in Section 2(d). So also has further scaling (T) based upon 'T' scores, to equalise means and standard deviations to "pool" data from the larger schools. As described in the previous section 3(c) and the Introduction 3(a) :-

(i)"Accuracy" is measured as the percentage of correct responses of those attempted,

(ii) "Speed" is calculated as the percentage of questions attempted of those possible, and

(iii)"Confidence" is assessed as the number of questions attempted compared with the actual number of the last question attempted - and the measure expressed as a percentage.

In appropriate cases "Accuracy", "Speed" and "Confidence" are all converted to T scores (Section 2(d)(i))

Dealing first then with Accuracy:-

Table 17 ("within" schools)

Product moment correlation coefficients between Heads' Ranking for OVERALL SUITABILITY for an Academic Course in 1969 and % correct on N.F.E.R. Verbal Test 15A in 1966.(Significance levels in parenthesis).

School(No.in sample, sex.)	r	r,based on T scales
B (29b) Modern	0.715 (.01)*	0.758* (.01)
K. (20g) Modern	0.597 (.01)	0.583 (.01)
L. (26g) Modern	0.400 (.05)*	0.484 (.05)*
M. (41g) Modern	0.698 (.01)*	0.556 (.01)
O. (34g) Grammar	0.285 (N.S.)	0.305 (N.S.)*
P**(21g) Grammar	-0.045 (N.S)*	-0.137 (N.S)*
Q**(71g) Grammar	-0.076 (N.S)*	0.000 (N.S).

Table 17 continued

*The calculations marked with an asterisk show correlation coefficients which are GREATER than those found between the actual "11+" Quotients obtained from this test and the Heads' 1969 Ranking (Table 3). In all the other cases above the correlations based on % correct are almost the same as those based upon "11+" Verbal Reasoning Quotients. Accuracy would seem to be very important.

** The homogeneous nature of the samples from these schools has been mentioned before (section 2(d)(ii)).

Table 18 ("within" schools)

Rank order correlation coefficients between % correct in N.F.E.R. Verbal Test 15A in 1966 and Heads' 1969 Assessments for OVERALL SUITABILITY.

School (no.in sample: sex)	r_s	significance level
E. (13b) Grammar	0.157	(N.S.)
G. (10b) Modern	0.909*	(.01)
G. (11g) Modern	0.843	(.01)
H. (25m) Modern	0.242	(N.S.)
I. (16m) Modern	0.412	(N.S.)
J. (12m) Modern	0.174	(N.S.)

*Again the coefficient marked is higher than that found for the relationship based directly on the Verbal Reasoning Quotients(Table 4). The other from school G and that from school I are close to the corresponding Table 4 figures, but the remainder are considerably less.

As in Section 2(d) "pooling" can be effected using the T scale conversion. For boys (School B is the only sample) this gives a figure of $r = 0.758$. For girls, in schools K,L,M,O,P and Q (213 cases) a figure of $r = 0.251$ is obtained.

Both these coefficients are significant at the .01 level of confidence. They are both greater than those found using the Verbal Reasoning Quotients themselves - as shown in Section 2(d)(iii).

Turning to ARITHMETIC/MATHEMATICS, the comparisons are with Section 2d(iv):-

Table 19 ("Within" schools)

Product moment correlations: 1969 Secondary Mathematics Assessments compared with % correct in 1966 N.F.E.R. Arithmetic Test 15E. (Significance levels shown in parenthesis.).

School (no. in sample: sex).		r	r, based on T scales
Boys	A.(38b) Modern	0.742 (.01)*	0.723 (.01)*
	B.(29b) Modern	0.784 (.01)	0.777 (.01)
	C.(70b) Grammar	0.302 (.01)	0.342 (.01)
	D.(59b) Grammar	0.096 (N.S)*	0.103 (N.S)
Girls	K.(20g) Modern	0.564 (.01)	0.589 (.01)
	L.(26g) Modern	0.500 (.01)*	0.531 (.01)*
	M.(41g) Modern	0.699 (.01)	0.588 (.01)
	N.(32g) Modern	0.205 (N.S.)*	0.693 (.01)
	O.(34g) Modern	0.375 (.05)*	0.364 (.05)
	P.(21g) Modern	0.275 (N.S)*	0.502 (.05)*
	Q.(71g) Modern	0.101 (N.S)	0.214 (N.S)*

Again the asterisks show correlations found to be greater than or equal to those found directly with the Arithmetic Quotients (Table 5), but this time the other figures are less than the respective ones in that Table.

Table 20 ("within" schools)

Rank order correlation coefficients between 1969 Mathematics and 1966 % correct in ARITHMETIC.

School (no. in sample: sex)	r_s	significance level
E.(13b) Grammar	0.225	N.S.
F.(16g) Modern	0.319*	N.S.
F.(9b) Modern	0.817	(.01)
G.(10b) Modern	0.406	N.S.
G.(11g) Modern	0.697	(.05)
H.(25m) Modern	0.155	N.S.

(continued)

Table 20 continued

School (no.in sample: sex)	r_s	significance level
I. (16m) Modern	0.802	(.01)
J. (12,) Modern	0.532	N.S.

With the exception of the one marked, the coefficients are generally notably less than their counterparts in Table 6.

Table 21

Product moment correlations using T scales to pool data from the schools listed in Table 19: 1969 Mathematics, 1966 Arithmetic % correct.

	r	significance level
Boys: 195	0.406	.01
Girls: 245	0.448	.01

Both these figures are slightly less than those found using the Arithmetic Quotients (which were 0.449 and 0.457 respectively).

Table 22

Product moment correlations calculated on samples drawn from the total population considered, ignoring differences between Secondary Schools. (ARITHMETIC % correct/MATHEMATICS 1969).

	r	significance level
Boys: 111	0.686	.01
Girls: 112	0.762	.01

Again, slightly less than the 0.718 and 0.773 found in Table 8 using Arithmetic Quotients.

Turning to ENGLISH, the comparisons are with Section 2(d)(v):-

Table 23 ("within" schools)

Product moment correlations: 1969 Secondary English Assessments compared with % correct in the 1966 N.F.E.R.English Test 15. (Significance levels shown in parenthesis.)

Table 23 continued

	School(no.in sample:sex)	r	r,based on T scales
Boys:	A. (38b)Modern	0.660 (.01)	0.771 (.01)
	B. (29b)Modern	0.652 (.01)*	0.609 (.01)
	C. (70b)Grammar	0.268 (.05)	0.285 (.05)
Girls:	K. (20g) Modern	0.012 (N.S)	0.154 (N.S)
	L. (26g) Modern	0.418 (.05)*	0.372 (N.S)*
	M. (41g) Modern	0.548 (.01)	0.375(.05)
	N (32g) Modern	0.465 (.01)	0.501 (.01)
	O. (34g) Grammar	0.614 (.01)*	0.579 (.01)*
	P. (21g) Grammar	0.339 (N.S)	0.434 (.05)
	Q. (71g) Grammar	0.251 (.05)*	0.212 (N.S)

Again, coefficients marked with an asterisk are greater than the corresponding figures in Table 9, when the English Quotients themselves were used.

Table 24 ("within" schools)

Rank-order correlation coefficients between Secondary English Assessments in 1969 and % correct in 1966 ENGLISH Test 15.

School(no,in sample:sex)	r_s	significance level
E.(13b) Grammar	0.500*	N.S.
F.(16g) Modern	0.768*	.01
F.(9b) Modern	0.400	N.S.
G.(10b) Modern	0.600	.05
G.(11g) Modern	0.821	.01
H.(25m) Modern	0.481	.05
I.(16m) Modern	0.636	.01
J.(12m) Modern	0.647	.05

Again, two of the coefficients are greater than the corresponding ones in Table 9 and the others are only considerably less in two cases (schools H and F, boys).

Table 25

Product moment correlations using T' scales to "pool" data from the schools listed in Table 23.

	r	significance level
137 Boys	0.507	.01
245 Girls	0.358	.01

These are less than the correlations found in Table 11.

Table 26

Product moment correlations calculated on samples drawn ignoring differences between Secondary schools. (ENGLISH % correct/ENGLISH 1969)

	r	significance level
111 Boys	0.612	.01
112 Girls	0.678	.01

Again, slightly less than those found in Table 12.

Comment on Section 3(d)(i):"ACCURACY"

It can readily be seen that accuracy in attempting the standardised tests is almost as good a measure as the direct assessment, whether it be of Verbal Reasoning Ability, or of Mathematics, or of English. It will be shown later (in Section 3(f)) that children are remarkably consistent in their levels of accuracy whether they are attempting tests of Verbal Reasoning Ability or of English and, to a lesser extent, Arithmetic.

It is also interesting to consider the possible reasons that accuracy scores from girls in two of the schools (one Modern and the other Grammar) are consistently almost as good as, or slightly better predictors of later assessments, overall, in English or Mathematics, than the quotients obtained from the standardised tests. It may be that special emphasis is laid upon "correctness" in those schools. It may, of course, be chance, since the samples were small, as are the differences between the correlation coefficients obtained.

It is submitted that the general pattern revealed by a study of the evidence reviewed in this Section stresses the importance of Accuracy when attempting standardised tests.

3(d)(ii) SPEED - As measured in terms of percentage of Questions attempted.

The first aspect to be considered is the relationship, if any, between placing in the Heads' 1969 lists for Overall Performance and Speed in tackling the N.F.E.R. Verbal Test in 1966.

Table 27 ("within"schools)

Product moment correlation coefficients between Heads' Ranking for Overall Suitability for an Academic Course in 1969 and Percentage attempted in the N.F.E.R.Verbal Test 15A in 1966. (Significance levels are shown in parenthesis; "N.S" means not significant).

School (no.in Sample; sex.)	r	r, based on T scales
B. (29b) Modern	0.390 (.05)	0.504 (.01)
K. (20g) Modern	0.665 (.01)	0.407 (N.S)
L. (26g) Modern	-0.292 (N.S)	not significant
M. (41g) Modern	0.531 (.01)	not significant
O. (34g) Grammar	not possible	not possible
P. (21g) Grammar	not possible	not possible
Q. (71g) Grammar	-0.005 (N.S)	not significant.

It is immediately obvious that there are wild fluctuations even in the "raw" correlations. It is also apparent at once that in some cases it is not possible to calculate correlation coefficients which have any meaning, since the vast majority of the children concerned in these schools attempted EVERY question on the paper. Where the entry "not significant" appears in full it means that the sum of the cross products of the T score deviations was negligible compared with the standard deviations of the scores.

This indeed was the pattern found throughout the consideration of the Speed measures. In general the candidates had found ample time at least to attempt the vast majority of the questions. Most had read

and attempted the last question and there was, therefore, no point in using "the number of the last question attempted" as a measure of "speed".

For completeness, however, the calculations were made where possible, and the tables are set out below for comparison with those in earlier Sections.

Where Rank Order methods were used corrections for tied scores were applied.

Table 28("within" schools)

Rank Order correlation coefficients between 1969 Heads' Assessments for Overall Suitability and Percentage attempted in the 1966 N.F.E.R.Verbal Test 15A.

School (no.in sample: sex)	r_s	significance level
E. (13b) Grammar	0.104	N.S.
G. (10b) Modern	0.467	N.S.
G. (10g) Modern	-0.020	N.S.
H. (25m) Modern	0.464	.05
J. (12m) Modern	0.291	N.S.

If one attempts to look at the overall picture so far, only one small sample from a Boys' school (B) gives a significant correlation and even this ($r = 0.504$) is considerably less than that found when accuracy was used as a measure ($r = 0.758$). It is perhaps worth repeating that the "accuracy" correlation was higher than the direct association ($r = 0.704$) for the boys in this school.

As noted, the position is little better when one considers the relationship between final ranking for Mathematics and Speed in tackling the 1966 Arithmetic papers.

Table 29 ("within" schools)

Product moment correlation coefficients between Secondary Assessments for Mathematics in 1969 and Percentage of Questions attempted in N.F.E.R. Arithmetic Test 15E taken in 1966. (Significance

Table 29 continued

levels are shown in parenthesis; "N.S." means not significant.)

	School (no.in sample;sex)	r	r,based on T scales
Boys	A. (38b) Modern	0.002 (N.S)	0.05 (N.S)
	B. (29b) Modern	0.530 (.01)	0.488 (.01)
	C. (70b) Grammar	0.011 (N.S)	not possible
	D. (59b) Grammar	0.084 (N.S)	not possible
Girls	K. (20g) Modern	0.087 (N.S)	0.275 (N.S)
	L. (26g) Modern	0.091 (N.S)	-0.172 (N.S)
	M. (41g) Modern	0.714 (.01)	0.343 (.05)
	O. (34g) Grammar	0.317 (N.S)	not possible
	P. (21g) Grammar	0.070 (N.S)	0.133 (N.S)
	Q. (71g) Grammar	0.104 (N.S)	not possible

Table 30 ("within" schools)

Rank Order correlation coefficients between Secondary Mathematics Assessments in 1969 and Percentage of questions attempted in N.F.E.R. Arithmetic Test 15E in 1966.

School (no.in sample: sex)	r_s	significance levels
E. (13b) Grammar	0.195	N.S.
F. (16g) Modern.	-0.046	N.S.
F. (9b) Modern.	0.375	N.S.
G. (10b) Modern	0.231	N.S.
G. (11g) Modern	0.283	N.S.
H. (25m) Modern	0.648	.01
I. (16m)Modern	0.619	.01
J. (12m) Modern	-0.129	N.S.

It was possible, though really of no value, to "pool" the correlation coefficients obtained using the T scores to see if there was any overall pattern emerging in respect of the relationship between Secondary Mathematics Assessments and percentage attempted in the 1966 Arithmetic Tests. For 108 girls in schools K,L,M AND P, the correlation coefficient found was $r = 0.157$. For 67 boys in

schools A and B the figure was $\underline{r = 0.245}$.

Neither are significant, in contrast to the figures of $r = 0.448$ and $r = 0.406$ for girls and boys respectively in connection with accuracy measures.

Finally, consideration was given to English. The relationship found here was slightly more promising. Details are given below:-

Table 31 ("within" schools)

Product moment correlation coefficients between Secondary Assessments for ENGLISH in 1969 and percentage of questions attempted in N.F.E.R.English Test 15 in 1966. (Significance levels shown in parenthesis: "N.S."means not significant))

	School(no.insample: sex)	r	r,based on T scales
Boys	A.(38b) Modern	0 .338 (.05)	0.459 (.01)
	B.(29b) Modern	0.532 (.01)	0.503 (.01)
	C.(70b) Grammar	0.317 (.01)	not possible
Girls	K.(20g) Modern	0.276 (N.S)	0.039 (N.S)
	L.(26g) Modern	-0.033 (N.S)	0.135 (N.S)
	M.(41g) Modern	0.581 (.01)	0.576 (.01)
	N.(32g) Modern	0.297 ((N.S)	0.299 (N.S)
	O.(34g) Grammar	0.022 (N.S)	not possible
	P.(21g) Grammar	0.259 (N.S)	0.474 (.05)
	Q.(71g) Grammar	0.390 (.01)	not possible

Two points are worthy of note. First, there are indications that the "raw" correlation coefficients are similar to those based on a "normal" distribution. Secondly, it would appear that at least as far as the children subsequently admitted to Modern schools are concerned, many did not complete the paper. Taken together these points possibly suggest that the paper was a little difficult for most of the children. In this connection it may be noted that in 7 out of the 10 schools considered, the correlations found between Arithmetic Quotients at "11+" and Secondary Assessments for Mathematics

were much higher than the corresponding figures for English.

In two cases (Schools M and P, both Girls' schools) the T scale correlation coefficients were higher than those found when accuracy was considered. In the case of school M the coefficient was slightly higher than that found for a direct association between English Quotients at "11+" and Secondary English Assessments.

Table 32 ("within" schools)

Rank Order correlation coefficients between Secondary English Assessments in 1969 and percentage of questions attempted in the 1966 N.F.E.R.English Test 15.

School (no.in sample: sex)	r_s	significance levels
E. (13b) Grammar	0.210	N.S
F. (16g) Modern	0.221	N.S.
F. (9b) Modern	0.529	N.S
G. (10b) Modern	0.576	.05
G. (11g) Modern	0.030	N.S
H. (25m) Modern	0.490	.05
I. (16b) Modern	0.704	.01
J. (12m) Modern	0.362	N.S

In the cases of schools H and I these figures are slightly higher than those found when accuracy was considered.

In this case there is sufficient information to consider "pooling" scores.

Table 33

Product moment correlations calculated using T scales to pool data from various schools. 1969 English Assessments compared with percentage of questions answered in 1966 ENGLISH Test 15.

Boys	r	significance level
Schools A and B (67 boys)	0.478	.01
Girls		
Schools K,L,M,N,P. (140 girls	0.283	.01

Table 33 continued

Both of these figures, though significant, are smaller than those found for "accuracy", which were 0.507 and 0.358 for boys and girls respectively. In turn, these were smaller than the direct associations found using 1966 English Quotients, for which the correlations were 0.547 (boys) and 0.411 (girls).

Table 34

Product moment correlations calculated using samples drawn from the total population and using the scaling on Secondary Verbal Test 2 only, ignoring differences in variances and means in different Secondary Schools. 1969 ENGLISH assessments compared with 1966 percentage of questions attempted in ENGLISH.

	r	significance level
Boys. 111 cases	0.539	.01
Girls 112 cases	0.644	.01

Again, these figures are both noticeably smaller than those found earlier for accuracy (boys r = 0.612, girls r = 0.678). These in turn were smaller than the direct associations using English Quotients (boys r = 0.658, girls r = 0.786).

CONCLUSIONS Except for the conclusion that in the samples considered, speed was much less related to final order than accuracy, no clear pattern emerges.

In many cases it was not possible to show any relationship, and in most cases no significant relationship was found. Except in English, and even there for the more able children, the inference could be that the tests were practically the equivalent of "untimed" ones in that most children finished them. However, this is a speculative statement. The children may have only "almost" or "just" finished the tests, when the time limit expired. The element of stress, of "working against the clock", may well have existed and exerted an influence which it would not have done had the children been reassured in the original instructions

that"there was no time limit".

It seems possible that speed had a greater part to play in determining the raw score obtained in the English tests than it did in the other tests. It will be recalled that there was a very much higher correlation between the percentage of questions attempted in English and the English Quotient obtained, than between similar measures for Arithmatic (Table 16). This, of course, would not affect the standardisation of the tests, though it could mean that the factors influencing assessment of English were more complicated than those for Arithmetic. The "speed" element involved may simply be a reflection of rate of comprehension of written material - the English papers contained much more script to be read and understood than the Arithmetic or Verbal Reasoning papers. This would be quite justifiable, since reading ability is part of "English".

3(d)(iii). "CONFIDENCE" as measured by the percentage of questions attempted compared with the actual number of the last question attempted.

This measure has been mentioned earlier, in the introduction to Section 3. With the single reservation that one is assuming that every question "skipped over" in the middle of the paper has been read, "confidence" as defined above, would provide a measure of "preparedness to risk an answer".

One can dispose of this measure in relation to the Arithmetic Test with little delay, since in almost every case the last question on the paper had been attempted and therefore the "confidence" score equalled exactly that for "speed" mentioned in the previous section. In one girls' school (M) there was a noticeable difference - one can speculate that children in the largest contributory Primary School had been advised to work steadily through the paper, attempting every question in turn. Such children would have a "confidence score" much closer to 100 than their "speed scores". In the event, they turned out to be youngsters whose subsequent performance in Mathematics was not as good as that of their faster classmates. The correlation between final order in Mathematics and "confidence" at "11+" in Arithmetic was therefore lower than that for "speed" i.e. $r = 0.381$ compared with $r = 0.714$.

This kind of result was found also in connection with the Verbal Reasoning Tests where again most children had sufficient time to reach the end of the paper. For school M (41g) the correlation between Overall Ranking at 13 and "confidence" in Verbal Test 15A at 11 was $r = 0.494$ compared with $r = 0.531$ for "speed" in the same test. For school K (20g) "confidence" was a slightly better predictor than "speed", $r = 0.685$ compared with $r = 0.665$. All four of the above coefficients are significant at the .01 level of confidence.

Though four were calculated, no significant rank order coefficients were found for the smaller schools involved.

It will be recalled that speed might have played a greater part in determining scores in the English tests. It is not surprising therefore to find that some relationship with the measure of "confidence" could be detected in this subject. The following table, giving the Rank Order Coefficients found, can be compared with Table 32.

Table 35

Rank order correlation coefficients between Secondary English Assessments in 1969 and "confidence" defined as the ratio of the number of questions attempted to the actual number of the last question attempted.

School (no.in sample:sex)	r_s	significance level
E. (13b) Grammar	0.210	N.S
F. (16g) Modern	0.210	N.S
F. (9b) Modern	0.450	N.S.
G. (10b) Modern	0.606	.05
G. (11g) Modern	-0.001	N.S
H. (25m) Modern	0.429	.05
I. (16b) Modern	0.691	.01
J. (12m) Modern	0.399	N.S

Reference to Table 32 above will show that these figures are almost the same as the earlier ones. For schools G(b) and J they are slightly larger than those found for Speed. For the remainder they are either the same, or somewhat smaller. In passing one may note that for school G(b) the figure is almost exactly the same as that found for accuracy. For J the figure is much less than that found for accuracy, and for both G(b) and J the accuracy figures were much less than those based directly on the English Quotients.

But the relationship was sufficiently interesting to suggest that the product moment figures be examined, and for 3 girls' schools this

was done.

Table 36

Product moment correlation coefficients between Secondary English Assessments in 1969 and "confidence" defined as the ratio of the number of questions attempted to the actual number of the last question attempted:

School (no.in sample:sex)	r, based on T scales	significance level
L.(26g) Modern	0.052	N.S
M.(41g) Modern	0.457	.01
P.(21g) Grammar	0.490	.05
These can be "pooled" and for 88 girls give		
	r = 0.309	.01

For schools L and M these figures are somewhat lower than those found for "Speed". For school P the correlation is marginally higher, but this was really an artificial result as 6 of the 21 girls involved had "confidence" scores of 100. "Pooling" gave a correlation about the same as that found for "speed", especially when it is borne in mind that many more children were included in the "speed" assessment.

CONCLUSIONS In the samples considered the measure of "confidence" was of little or no value. Clearly this will always be the case where children have ample time to finish the paper, as the score would then be no different from that for "speed".

At the other extreme, in cases where hardly anyone completes the paper there might be some value in such a measure as a different one from "speed".

In intermediate cases, however, where many children complete the paper, the measure is not likely to be more useful than "speed" for the following reason.

To obtain a higher correlation than that found for "speed" alone

the slower candidates must in the end be judged more able than the quicker ones. While those who finish the test "stand still" in the order of merit relating to speed, they must, for a higher correlation with ability to result, be overtaken in ability assessment by children who did not reach the last question. As more able children can in general "process information faster" such a turn of events is unlikely.

There follow Summary Tables Indicating Relative Sizes of Correlation Coefficients.

The numbers in the following Tables relate to T scale correlations plus numbers for Rank Order Correlations for Schools and Groups.

Table 37

Significance levels of correlation coefficients between measure shown and final order, overall or in stated subject.

Significance level of Confidence	VERBAL REASONING using V.R.Q.	% correct	% attempt	ENGLISH using E.Q.	% correct	% attempt	ARITHMETIC using A.Q.	% correct	% attempt
.01	5	5	1	10	8	5	10	9	2
.05	2	1	1	6	5	2	4	4	1
not significant	6	7	10*	2	5	11	5	6	15
	13	13	12*	18	18	18	19	19	19

* 1 calculation not carried out.

Table 38

Relative sizes of correlation coefficients - giving details of number of cases in which the coefficient found using the Quotient directly was greater than that using "percentage correct" and so on.

		Direct Quotient greater than % correct correlation.	% Correct greater than direct quotient.	% attempt greater than direct quotient.	% correct greater than % attempt.	% attempt greater than % correct.
VERBAL REASONING	All cases	8	5	0	9	2
	Significant cases	5	3	0	6	1
ENGLISH	All cases	13	5	2	13	5
	Significant cases	13	3	2	9	4
ARITHMETIC	All cases	14	6	1	18	1
	Significant cases	12	3	1	13	1

OVERALL CONCLUSIONS for Section 3(d)

1. Of the measures examined, for the children under review "ACCURACY" is more significantly related to final assessments of merit than either "SPEED" or "CONFIDENCE".

2. In 33 sets of correlation carried out using the T scale to normalise distributions, "Accuracy" was found to correlate more highly than the direct association between the respective Quotients and the Final Order after 3 years in the Secondary School, on 12 occasions. It was within 0.1 of that found for the direct association on a further 16 occasions. Accuracy would appear to be a major constituent of success in standardised tests.

3. From the table which immediately precedes this set of conclusions it is apparent that the direct calculations using the "proper" quotients are much more closely related to final assessments of merit than either "accuracy" or "speed". However, it is also clear that "accuracy" is much more closely related to final order than "speed".

4. There is a general point. The smaller correlations for "speed" may merely indicate that more people finished the papers - the correlation would vanish if every child attempted every question and at that point "accuracy" would coincide with raw score obtained. Where the "speed" correlation is not equal to zero, some children did not attempt every question. When this factor comes into play, as it seems to have done in English, which is it better to do - to be fast or to be accurate ? Considering the T score scales (Tables 23 and 31) for two schools, M and P, it appeared to be better to be fast. Note also the significant correlations for "confidence" for these schools (Table 36). Perhaps they encourage "risk taking" for the sake of increased fluency in writing. However, for the two boys' schools and for the three other girls' Modern schools "accuracy" was more closely related to the final order of merit and this is reflected in the figures obtained from "pooling" data.

As far as the direct correlations were concerned in the "11+" examination itself (reported in Section 3(c)) the evidence again favours accuracy.

In the Rank Order correlations two groups were found (School I and the boys in School F) where it was "better" to be fast than accurate. But for the majority "accuracy" was more closely related to final order in the Secondary School than "speed".

More evidence on this point is given in the Section dealing with the Consideration of Individual Cases, which follows.

3(e) Comparison of Individual Cases

When one considers the information given in the previous Section, a question which comes to mind is the possible effect of "intelligence". Presumably the more intellectually able would be more accurate, and therefore should fare well in the "follow up" review. There are two ways of examining this aspect. One is to use partial correlation techniques, the other "matched pairs". It will be recalled that quite marked adjustments had to be made to the distributions used earlier to transform them to more "normal" shapes. Therefore, a review of "matched pairs" seemed to offer a more reliable basis for judgement. The data relating to all the 183 children in 4 schools (code letters A,C,M and O: 2 Grammar and 2 Modern) were reviewed and "pairs" from within the same school were matched for age and "intelligence". For the latter the measure used was the Verbal Reasoning Quotient obtained from the N.F.E.R. test in the "11+" examination. In all 28 boys' and 42 girls' pairs were chosen. The original intention was to accept only the exact matches in terms of V.R.Q., and age in years and months. This produced very small samples, so a latitude of $\pm$ 1 month and $\pm$ 1 point of I.Q. was considered. Eventually, in the cases of girls, this was relaxed even further to permit the inclusion of some falling within a range of $\pm$ 2 months and $\pm$ 2 points of V.R.Q. However, these tolerances were balanced as far as possible-the "older" children having the lower Verbal Reasoning Quotients.

The table below gives details of the variations from exact matching:

Table 39

	"Pairs" having a difference in V.R.Q. of			"Pairs" having a difference in age of		
	0 pts.VRQ	1 pt.VRQ	2 ptsVRQ.	0 mnth.	1 mnth.	2 mnths.
Boys (N=28)	22	5	1	14	14	0
Girls (N=42)	20	14	8	21	12	9

Each pair was then considered individually and notes made of relative accuracy in the standardised tests of Verbal Reasoning, English and Arithmetic, and subsequent ranking for Overall Suitability, English and Arithmetic respectively. The question was "Does the more accurate, or the lesser accurate member of the pair achieve a higher subsequent position?". There is, of course, another possibility - that the members were equally accurate, or achieved an equal position in the"follow up' review.

The tables below set out the general position and are followed by an evaluation of the data.

Table 40 OVERALL SUITABILITY

"Matched"Pairs - Accuracy in Verbal Reasoning Tests Compared with ranking 3 years later for Overall Suitability for an Academic Course.

	More accurate achieves higher final ranking	Same accuracy Same final rank.	Lesser accurate achieves higher final ranking.
Boys (N=28)	14	7	7
Girls (N=42)	25	4	13
Total (N=70)	39	11	20

Table 41 ENGLISH

"V.R.Q.Matched" Pairs - Accuracy in 11+ English Test Compared with ranking 3 years later for performance in English.

	More accurate achieves higher English ranking	Same accuracy or final ranking	Lesser accurate achieves higher final ranking
Boys(N=28)	19	4	5
Girls(N=42)	32	nil	10
Total(N=70)	51	4	15

Table 42 ARITHMETIC

"V.R.Q Matched" Pairs - Accuracy in "11+" Arithmetic Test Compared with ranking 3 years later for performance in Arithmetic.

	More accurate achieves higher Arithmetic ranking	Same accuracy or final ranking	Lesser accurate achieves higher final ranking
Boys (N=28)	22	2	4
Girls(N=42)	30	nil	12
Total (N=70)	52	2	16

The tables give the impression that when "matched" pairs are considered the more accurate member generally achieves a higher position 3 years later. This can be further examined using the CHI - SQUARE Test and the null hypothesis. (Garrett,1958. Lindley and Miller,1953). Tables can be drawn up showing the observed and expected frequencies. These are 3 x 2 tables and since any two entries made in a row determines the size of the third, and the sums of the columns are fixed so that only one entry is free, the number of degrees of freedom are 2 x 1 = 2. Tables of the CHI - SQUARED* distribution will then give the probability of finding values by chance alone, as high as those observed.

It is suggested that there are two plausible hypotheses. First, that the chances of the more accurate partner, or the lesser accurate partner achieving a higher final ranking - of of there being no difference - are all equally likely. On this basis the "expected" frequencies would be N ÷ 3 in each column. The other possibility, in view particularly of the discussion a year or two ago about "over and under-achievers" is that for "matched" pairs, the achievement of final ranking orders will be "normally" distributed. If, therefore, a "normal" curve is considered, divided over a base line of 6σ into

* Yates's correction not applied: expected frequencies 5 or more.

3 equal segments of 2 σ each, the expected proportion of cases falling into each category is given below:-

					Proportion of cases
Between	+3	and	+1	:	0.16
"	+1	"	-1	:	0.68
"	-1	"	-3	:	0.16

The actual number of cases can be translated into expected frequencies using these proportions.

The value of CHI - Squared for each of these possibilities has been calculated and is given below:-

Table 43

Values of "CHI -Squared" calculated on the assumption that there are equal chances with pairs matched on the basis of V.R.Q. and age, of the more accurate partner, or the lesser accurate partner achieving a higher final rating or of there being no difference. (2 degrees of freedom).

	Overall Assessment.	English Assessment.	Arithmetic Assessment
Boys(N=28)	3.500*	14.391	26.001
Girls(N=42)	15.858	38.286	35.536
Total(N=70)	17.516	51.800	57.029

All of the above are significant at the 0.01 level except the one noted with an asterisk where $0.20 > p > 0.10$

Table 44

Values of "CHI-Squared" calculated on the assumption that there is a normal distribution of cases in which, with pairs matched on the basis of V.R.Q. and age, the more accurate partner, or the lesser accurate partner achieves a higher final rating, or of there being no difference. (2 degrees of freedom).

Table continued

	Overall Assessment	English Assessment	Arithmetic Assessment
Boys(N=28)	29.260	59.456	83.767
Girls(N=42)	76.713	125.262	113.358
Total(N=70)	104.056	182.757	194.371

All of these are significant at the 0.01 level of confidence.

The inference is clearly that with the same age and level of ability, the more accurate children achieve higher final placings.

However, a few words of caution will not be out of place. As far as overall assessment is concerned the "pairs" were matched with respect to ability and age and consideration was of accuracy on Verbal Reasoning Tests. But in the other cases, while "pairs" were matched for age and Verbal Reasoning ability it was accuracy in English or in Arithmetic Tests which was assessed. The "pairs" were NOT matched for English or Arithmetic Quotients. In the cases of 40 "pairs" for English and 45 "pairs" for Arithmetic the more accurate partner had a higher English or Arithmetic Quotient.

Despite this, the general conclusion seems a sound one for in the admittedly small number of cases where English and Arithmetic Quotients were equal as well as age and Verbal Reasoning ability, there were three times as many cases for English and five times as many for Arithmetic, in which the more accurate partner subsequently achieved a higher ranking. Details are given below:-

Table 45

Cases of pairs 'matched' for age and Verbal Reasoning ability and for ability in English or Arithmetic.

	More accurate partner achieves higher final ranking	No difference	Lesser accurate partner achieves higher final ranking
English(N=9)	6	1	2
Arithmetic (N=7)	5	1	1

It would seem that accuracy is a significantly worthwile attribute.

References in Section 3(e)

Garrett.H.E.	1958 :	Statistics in Psychology and Education (pp.253 - 258 and 450)	: Longmans Green & Co. New York.
Lindley.D.V. and Miller.J.C.P.	1953 :	Cambridge Elementary Statistical Tables (page 7)	: Cambridge University Press.

3(f) Stability of Accuracy

A point of possible interest was whether the children in the samples displayed consistent lack of accuracy when attempting more than one test.

Therefore all the "percentage correct" scores of 192 boys and girls from 4 schools (2 Grammar and 2 Modern) were listed and considered. The tests (4 in all) were the two Verbal Reasoning tests, the English Tests and the Arithmetic Tests used in the "11+" battery. In all, therefore, 768 scores for accuracy were reviewed, only one of which was 100%.

It was considered that reasonable consistency in degree of accuracy would be shown if the range of the scores fell within a 10% group. (i,e,a 5% variation from the middle score). Considering all 4 tests the figures given below show the % of children with "consistent" levels of accuracy:

Boys (N = 107)	75.7%
Girls (N = 72)	71.5%

When one considers the diverse nature of the content of the tests (English, Arithmetic and Verbal Reasoning) these seem highly significant. It was really the Arithmetic Test which "sorted out" the inaccurate, for when only the 2 Verbal Reasoning Tests and the English Test were considered the figures rose sharply to:

Boys	84.11%
Girls	81.02%

It might be thought that a trait of carefulness was emerging. If so, would such a measure prove stable as time passed ? It is worth noting that the tests just mentioned were administered on separate occasions over a month apart. However, it was possible to make a direct comparison between the degree of accuracy shown by these children in 1966 when attempting a National Foundation Verbal

Reasoning Test and that shown by them 3 years later on a similar N.F.E.R. Test (Secondary Verbal 2) in 1969. The variation was marked. Only 35.52% of the boys and 25% of the girls had "% correct" scores within $\pm$ 5 points of their earlier ones. 22.43% of the boys and 19.4% of the girls had improved their level of accuracy by 5% or more. 42.05% of the boys and 55.6% of the girls were found to have "% correct" scores more than 5% below their previous levels.

Possible explanations for this decline in accuracy include the probability that the Secondary Verbal Test was much more difficult - the youngsters were near the lower limit of the age range for which the test was designed and quotients were markedly lower on average, as Section 2(e) shows. Again, increasing age could have led to increasing confidence or tendency to guess an answer when in doubt.

It seems reasonable to conclude that in the sample reviewed there was a marked consistency in accuracy in attempting standardised tests of varying content at the age of eleven. The indications are that standards of accuracy vary when intervals as long as three years are allowed to elapse.

4. Notes on Accuracy in the 1966 Tests and (i)1969 Assessment of Aggressive Tendencies (ii)Choice of Composition in 1966.

(i) Professor Liam Hudson has reported finding a noticeable link between inaccuracy and aggressive tendencies. There was"a strong connection between inaccuracy on the I.Q.Test and violence in the drawing. Out of 27 boys with high accuracy scores only one did a drawing involving violence; $p < .005$". (Hudson.1965).

This suggested that it might be worth seeking further information about such a personality trait for the children reviewed in this enquiry.

Accordingly, several Heads were asked if they would, in consultation with class teachers, categorise children in the survey on what was basically a 3 point scale, but to which '+' and '-' signs could be added, as "aggressive", "normal" or "timid". No explanation was given of the purpose of the enquiry - so that judgements would not be biased - but it was explained that "aggressive" could be interpreted as willingness or eagerness to attack class work as well as classmates. Information was obtained from 5 Secondary Modern schools relating to 93 girls and 34 boys.

Inaccuracy was considered in two ways:-

(a) Where the general level of accuracy was 60% correct or less

(b) Where, by considering a 'profile' of accuracy scores, marked variations in levels of accuracy were apparent.

It is sufficient to report that no apparent link or pattern was discernible. Those children judged "aggressive" or "timid" were distributed equally among those of high consistently accurate scorers as well as the inaccurate or variable ones.

It may be that the samples were too small, or that teachers' observation of overt behaviour of 13 year old children is not sufficiently

discriminating. For example, teachers might consider mildly aggressive tendencies desirable in girls of this age, who would probably appear lively, with a touch of independence. These very qualities might be called "cheekiness" when associated with boys. The present sample was largely composed of girls. Possibly a personality test would reveal hidden tendencies.

No real evidence either way was obtained from this review.

(ii) Another possibility was that aggression might be linked with flights of fancy, and as one of the possible essay choices had a subjectively adventurous flavour, this was also scrutinised.

The question was "You are exploring a deserted aerodrome when you hear a mysterious buzzing noise from a hangar. Suddenly the doors are flung open. Describe what happened next".

Again, however, no evidence was found to suggest that low levels of accuracy, or variable levels of accuracy, were linked with any particular essay choice.

Reference in Section 4

Hudson L. 1965 : "Contrary Imaginations"(p.57). Methuen, London.

5. Brief Notes on Other Enquiries Involving the Assessment of Performance in Secondary Schools.

The purpose of this Section is to stress that there are many other factors operating which affect performance and assessment in Secondary Schools. These vary in strength and importance from the vague suggestion that examiners give higher marks to the writers of the first scripts read and therefore, as alphabetical order is customary, favour those with surnames beginning with A,B or C, (Page, 1969), to the carefully documented major effects of environment (Wiseman, 1964).

The assessable performance of a 13 or 14 year old is influenced by the interaction of the personal characteristics of the child, circumstances at home and out of school hours, and factors which affect life within the school.

Considering first the personal characteristics of the child, clearly there are individual differences in reasoning ability and measures of these correlate very highly with assessments of success in the Secondary schools,as the studies quoted earlier show. The nature/nurture argument about the degree to which innate capabilities can be identified, modified or developed, continues and is likely to do so. The effects of environment and heredity are difficult to separate (Vernon, 1960), but that there are important inherited differences, including neurological ones, must be beyond dispute. (Burt,1969).

Similarly, there is evidence that personality differences are correlated with examination scores. The personal need for achievement has been studied closely by McClelland (1953) and more recently by Atkinson and Feather (1966). In this connection a point with important implications for homo- or heterogeneous groupings within schools, is that for maximum incentive a child shall not perceive tasks as being too easy or too difficult (Nisbet J. 1968). Clear

preferences for intermediate levels of risk where personal skills are involved has been reported in work at Princeton University. (McClelland. 1967).

Academic motivation has been studied as part of a wider survey by Dr.Entwistle (1967). He noted that J.A.Finger (1966) showed it to be independent of intelligence, yet closely related to school attainment. However, as the main purpose of the present enquiry is a study of speed and accuracy, it is perhaps of interest to note that in an enquiry mentioned earlier, it was found that levels of aspiration among University Students correlate considerably less than measures of accuracy with final degree class obtained. (Himmelweit and Summerfield,1951).

That more intelligent children have difficulties in adjusting to transfer to Secondary school was shown by Chazan (1962). However, it has been noted that such children are also much more likely to be capable of reporting such difficulties.(Philip,1968a).

As children in Secondary schools are generally grouped 'a year at a time', the effects of differences of age in months has been studied. In problems of adjustment, the older child has an advantage. The younger, less mature children are more likely to suffer a setback in their educational progress immediately after entering the Secondary school. (Nisbet and Entwistle 1969a).

In children about to enter Secondary schools it has been shown that there is a link between ability to attain concepts and the ability to use economical systematic strategies in tackling problems. In addition,at this age, tendencies to extraversion are an asset.(Banks,1964). This received further confirmation in a wider study by Rushton (1966).

The complex relationships between success in school work and measures of extraversion and neurotic tendencies have been the subject

of enquiries by Child (1964) and Hallworth (1964), who showed that ~~xx~~ teachers' ratings were based almost equally upon measures of extraversion, intelligence, and emotional stability. Shirley Cunningham (1967), in Aberdeen, found school attainment related to neuroticism and also that there is a sex difference in the relationship of the measures of extraversion and attainment. It would appear that extraverted girls and introverted boys tend to be more successful in school work than the opposite personality types. (Entwistle and Cunningham, 1968).

A brief reference has been made earlier to the Comprehensive survey recently completed in Aberdeen (Nisbet and Entwistle, 1969b). It is worth indicating the sizes of the correlation coefficients found between measures of academic success at 13 and personal characteristics besides intelligence.

	Boys (N = 1293)	Girls (N = 1245)
(i) Attitudes to work:	0.696	0.700
(ii) Academic motivation:	0.503	0.418
(iii) Ambition:	0.460	0.402
(iv) Social Maturity:	0.417	0.498
(v) Neuroticism:	-0.148	-0.196

Measures relating to home environment gave the following correlation coefficients:-

	Boys	Girls
(vi) Parental encouragement	0.603	0.586
(vii) Socio Economic Rating	0.308	0.310

These two measures are also reflected in the relationships found with such factors as the kind of employment Mothers took up (semi-skilled or unskilled being related to detioration in school work). There was also a link with the number of rooms occupied by the family and, hence, the space available for homework or private

study. Parental encouragement and educational level of the home seemed to be of special importance in relation to improved performance as the Secondary School course progressed. Very strong links between parental encouragement and measures of I.Q. (r=0.604) and more still with school attainment (r = 0.660) had been found earlier.(Fraser 1959). It was found that the ten individual aspects considered correlated more highly with attainment than they did with I.Q. Parental attitudes to the education and future occupation of the child had the greatest effect. Thus the home can override potential talent. Professor Fraser noted, "there would appear to be a common thread.... a normal home background, emotional stability, freedom from tension and economic insecurity, and consistent encouragement from parents, are necessary for a child if his school work is to reach the level allowed by his intelligence".

The major surveys carried out in the Manchester area in 1951 and 1957 also stressed the vital importance of these factors. After the 1957 survey a computer was used to begin a factorial analysis of 29 main variables and the findings were further clarified by graphical rotation to give six meaningful factors. Two-thirds of the variance on the educational measures was accounted for by Factor I, which accounted for 26% of the total variance. Poor home conditions such as cruelty and neglect, cleansing notices, probation, illegitimate children,and negative correlation with the number of Jurors per thousand voters accounted for over 77% of the social variance. This major factor was labelled "general educational-social factor". Factor II, more closely associated with attainment than intelligence, suggested a lack of maternal care, particularly from the psychological aspects, when the social variables were considered. This factor contributed some 21% of the Educational variance. Factor III, which was significantly negatively associated with measures of brightness, yet not positively with backwardness, seemed

to emphasise lack of physical aspects of care. This factor accounted for 8% of the total educational variance. The remaining 3% was virtually all attributable to Factor VI, basically Juror index and number of persons per acre - a purely economic factor. (Wiseman,1964a).

Considerations of extremes of social class led to the conclusion in North Ireland that home environment had more effect than the "qualifying examination" for transfer to selective Secondary education, in determining the subsequent careers of young people.(Cave 1967). This was also reflected in the findings of Jahoda (1953). He reported that Lancashire working-class boys looked for jobs which would keep them "with the lads". Of course, the need for working-class children to carve out their own careers and to make their own educational decisions received considerable emphasis by Jackson and Marsden (1962).

The significant effects of poor home background on children about to enter Secondary schools were shown by Douglas (1964a). In his study of 5000 children born in 1946 it was shown that home conditions correlate with increases or decreases in apparent ability between the ages of 8 and 11. Such children begin their Secondary school careers at a disadvantage. It has been shown that almost all children suffer from problems of adjustment when they transfer from Primary to Secondary education, but those from the higher social classes are least affected (Murdoch,1966).

Intelligence tests usually have a high verbal content and the language difficulties experienced by children from working-class homes have been explored, notably by Bernstein (1960) and Jahoda (1964). The latter found striking discrepancies in the vocabulary levels of middle class boys aged 10 - 16 and their working-class fellows, although they had been "matched" using non-verbal tests. However, one study of discrepancies between scores on verbal and non-verbal tests of intelligence failed to show that socio economic status significantly affected the direction of the discrepancy (Watt,1965). Another study

in which large numbers of children had also been involved and which failed to show any link between home environment and improvement or deterioration in intelligence test scores in early adolescence, was that by Ann P.Lamont (1961). She noted the importance of the early years in influencing mental development and concluded "it may well be that the main effect of environment has already been felt before the age of eleven".

The many problems and considerations involved in the move from Primary to Secondary education were studied in the Aberdeen survey mentioned earlier. (Nisbet and Entwistle 1969). An enquiry included in Professor Nisbet's general programme showed that co-operation between schools and liason about transition arrangements had a significant effect upon improved adjustment to Secondary schooling. (Philip,1968b). "Streaming",of course, is another school factor. Albeit referring to children between the ages of 8 and 11, it has been shown that "at each level of ability the children in the upper stream improve their scores ... and those in the lower streams deteriorate".(Douglas,1964b). In Secondary schools, there has been strong criticism of the use of the "11+" examination data to "stream" children, it being suggested that over two-thirds would be wrongly allocated.(Thompson,1966). "Streaming" has also been criticised by Jackson (1964).

Another factor affecting performance is co-education. In a recent report on a study extending over 20 years, it is claimed that G.C.E. 'O' level examination results are better for boys from co-educational schools than from single sex schools. While the figures relating to girls are less clear cut there are indications of similar advantages in co-edcuational schools.(Dale 1969).

The influence of peer groups in the home environment has already been mentioned. In school,too, "Fear of peer exclusion is a powerful determinant of academic performance.... many...students associate

solitude with failure and to them popularity is the key to success.... students were inclined to earn lower marks than they could, since acceptance within the group was more likely for those doing merely "satisfactory work" than for those doing "extremely well" ". (Strom. 1969).

That the school environment had a major role in affecting scholastic attainment was shown by Warburton (1964). He found the main factors were

(i) Socio-economic level, showing chiefly as a good school neighbourhood, high juror index and good school buildings,

(ii) progressive methods of education, and

(iii)good teaching conditions, based mainly on small size of class and good attendance.

Ann Sutherland (1963) indicated that differences in class teaching appeared to be more important than other variables such as sex, social status of the school,and I.Q. measures,in influencing discrepancies in school attainments at the transfer examination stage.

The problems of evaluating "progressive" education are many and will increase as the practices become widespread. A pertinent comment is that unexpected side effects can occur. "What is particularly worrying, especially to the young teacher...is that deviant behaviour can exist side by side with considerable enthusiasm. The line between mad keenness and over-excitement is not easily drawn, and the total experience for the inexperienced teacher might well be somewhat traumatic." (Salt,1969).

New techniques are still really only being studied at University level. Programmed learning and active involvment of students has been shown to have advantages in certain situations over traditional lectures with "passive" students.(Pikas,1969). A less complicated change with advantages at this level is that of posing questions, inviting answers and then supplying the correct information. This

technique has been shown to produce a higher level of performance in a later test than just delivering a lecture in the form of direct statements of fact. (Sime and Boyce,1969).

As far as attainment in school is concerned, if it was necessary to single out one influence, "the most important factor of all appears to be the cultural level of the home and the parents' interest in, and aspirations for, their children's education. Hence the further improvements that we would desire in material conditions, social welfare and schooling, will not of themselves eliminate the handicaps of the lower working and less educated classes.... It is the people and not only their material conditions that we must change". (Vernon,1969).

Further evidence, if any was needed, of the significance of such parental support was supplied a week or two ago, when it was shown that there was a high correlation between attendance at meetings of parent teacher associations and socio-economic status, and also that it is the parents of the abler children who support such associations. (Kent.1969).

However, as Professor Wiseman,in this country,and David Lavin, in the United States, have pointed out, this is essentially a multivariate and dynamic situation. As Wiseman says "educational attainment is not a single entity, but a shorthand description of the reactions of individual pupils to various forms of educational measurement. ...the picture is not one of the pupil being surrounded by a multitude of forces...but rather that the pupil himself produces some of the forces and interacts with others". (Wiseman,1964b). Lavin (1965) comments, "neither psychological nor sociological factors alone are capable of substantially enhancing our understanding of academic achievement... it is at the level of the interaction between these two types of factors that any major breakthrough is likely to come". Wiseman notes, "the immediate research task is to attempt to structure the field, in the hope that we may achieve further insights

into the complex of inter-relationships of variables and factors". Lavin says, "If we can conceptualize, and can measure new dimensions of student performance, and if performance can be predicted accurately, such knowledge could be used in admissions - (and) - for modifying educational organisations so that schools can achieve their purposes more readily".

Here then, as in the field to which the reference was originally made, "The further outlook is thus a picture of a growing body of probable associations between observations of behaviour and concepts which are supported by demonstrable operations...The general features of scientific activity of co-ordination, prediction and control, are thus possible and can be improved with increasing knowledge of relevant conditions". (Smith F.V.,1960).

References in Section 5

Atkinson and Feather	- 1966 -	"A Theory of Achievement	- John Wiley & Son, New York.
Banks. J.	- 1964 -	"The Relationship between problem solving in Arithmetic and concept attainment, intelligence and personality characteristics."	- M.Ed.Thesis. University of Manchester.
Bernstein. B.	- 1960 -	"Language and Social Class."	- British Journal of Sociology. Vol. 10.1960(p.60-61)
Burt, Sir Cyril	- 1969 -	"The Mental Differences Between Children" article in "Black Paper Two"(pp16-25)	- The Critical Quarterly Society. London.
Cave, E.D.	- 1967 -	"The Effects of the '11+' result on the subsequent careers of pupils"	- British Journal of Educational Psychology.Vol.37 Pt.1.(pp.41-46)
Chazan, M.	- 1962 -	"School Phobia"	- British Journal of Educational Psychology. Vol. 32.(pp.209-217)
Child,D.	- 1964 -	"The Relationships between Introversion, Extraversion,Neuroticism and Performance in School Examinations."	- British Journal of Educational Psychology. Vol. 34.(pp.187-195)
Cunningham.S.	- 1967 -	"The Relationship between Neuroticism and Extraversion and the School Attainment of Aberdeen School-children."	- M.Ed.Thesis. Aberdeen University. 1967.
Dale.R.R.	- 1969 -	"Changeover to Co-Education" - article in "Verdict on the Facts". (p.54)	- Advisory Centre for Education, Cambridge.
Douglas. J.W.D.	- 1964 -	"The Home and The School" (a)pp.116-117.(b)p.114.	- MacGibbon and Kee, London.

/continued

References in Section 5 continued

Entwistle.N.J.	- 1967 - "The Transition to Secondary Education: The age of Transfer and the correlates of Academic Success in the first year at Secondary School".	- Ph.D.Thesis. Aberdeen University (pp.37 and 133)
Entwistle N.J. and Cunningham, Shirley.	- 1968 - "Neuroticism and School attainment - a linear relationship ?" (quoted by Nisbet.J. & Entwistle N.J. 1969)	- British Journal of Educational Psychology. Vol.38.
Finger J.A.	- 1966 - "Academic Motivation and Youth Culture Involvement: Their relationships to school performance and career success"	- School Review,No. 74.(p.177-195)
Fraser.E.D.	- 1959 - "Home Environment and The School" (p.59 & 72)	- University of London Press,London.
Hallworth.H.J.	- 1964 - "Personality Ratings of Adolescents: A Study in a Comprehensive School".	- British Journal of Educational Psychology. Vol.34.(pp.171-177)
Himmelweit H.T. and Summerfield. A.	- 1951 - "Student Selection: An Experimental Investigation II"	- British Journal of Sociology.Vol.2 (pp.59-75)
Jackson.B.	- 1964 - "Streaming: An Education System in Miniature."	- Routledge and Kegan Paul.London.
Jackson. B. and Marsden. D.	- 1962 - "Education and the Working Class" (p.134)	- Routledge and Kegan Paul,London.
Jahoda.G.	- 1953 - "Social Class: Attitudes and Levels of Occupational Aspirations in Secondary Modern Leavers"	- British Journal of Psychology.Vol.44. (pp.95-107 especially p.104)

/continued

References in Section 5 continued

Jahoda. G.	- 1964 -	"Social Class Differentials in Vocabulary Expansion.	- British Journal of Educational Psychology.Vol.34. (pp.321-323)
Kent. C.C.	- 1969 -	Survey on Parents and Schools. Kent County Council report in "Education" 5th.Sept. 1969.	- "Education" Journal of Association of Education Committees London.Vol.134.No.10. (p.1082)
Lamont.Ann.P.	- 1961 -	"A Study of the Relationships of Social Class to Changes in Mean Intelligence Test Score between the ages of 11 and 13."	Ed.B. Thesis. University of Aberdeen.
Lavin.D.E.	- 1965 -	"The Prediction of Academic Performance" (pp.162 and 168)	- Russell Sage Foundation: John Wiley & Sons,New York .
McClelland. D.C.	- 1953-	"The Achievement Motive.	Appleton-Century-Crofts: New York.
McClelland.D.C.	- 1967-	"The Anatomy of Achievement Motivation" Translation of original work by Heinz Heckhausen. (p.95)	- Academic Press, New York.
Murdoch.W.F.	- 1966-	"The Effect of Transfer on the Level of Children's adjustment to School"	- M.Ed. Thesis. University of Aberdeen.
Nisbet.Jennifer.	-1968-	"The Relationship between Academic Motivation and the Social Class,Sex,Intelligence and School Attainment of Aberdeen Schoolchildren"	- M.Ed.Thesis. University of Aberdeen.
Nisbet. J,and Entwistle N.J.	- 1969-	"The Transition to Secondary Education" (a)chapter 8 (b) chapters 6 & 7	- Scottish Council for Research in Education No.59. University of London Press.

/continued

References in Section 5 continued

Page. D.	- 1969 -	"Lecturer urges children to disrupt G.C.E. System" Article in "The Daily Telegraph" for July 30th.1969.(p.21)	- The Daily Telegraph. London.
Pikas. A.	- 1969 -	"Comparisons Between Traditional and Programmed Learning as a Function of Passive Performance and Active Application and Time Till Application."	- Journal of the Assoc.for Programmed Learning and Educational Technology. Vol.6 no.1. (pp.20-25)
Philip.I.	- 1968 -	"The Effect of Changes in the Traditional Transfer Procedure on Children's Adjustment to Secondary School" (a)p.15 (b)General conclusion.	- M.Ed.Thesis University of Aberdeen.
Rushton.J.	- 1966 -	"The Relationship between Personality Characteristics and Scholastic Success in 11 year olds"	- British Journal of Educational Psychology.Vol. 36.Pt.2(pp.178-184)
Salt.J.	- 1969 -	"Problems of Integrated Education" Article in "Trends" in Education" No.16.p.24.	- H.M.S.O. London.
Sime.M. and Boyce G.	- 1969 -	"Overt Responses, Knowledge of Results and Learning"	- Journal of the Association for Programmed Learning and Educational Technology.Vol.6 No.1.(pp.12-19)
Smith F.V.	- 1960 -	"Explanation of Human Behaviour" (p.450)	- Constable & Co. London.
Strom. R.D.	- 1969 -	"Psychology for the Classroom"	- Prentice Hall, New Jersey.

/continued

References in Section 5 continued

Sutherland. Anne E.	- 1963 -	"Children with High Discrepancy in Standardised Scores of English and Arithmetic at the Transfer Examination Stage".(p.24)	- B.Ed.Thesis University of Aberdeen.
Thompson. D.	- 1966 -	"Secondary School Streaming"	- Educational Review, Birmingham.Vol. 18.No.3. (pp.199 and 200)
Watt.Joyce,S.	- 1965 -	"An Investigation into the Degree of High Discrepancy found between a Verbal and a Non Verbal Test of Intelligence"	- Ed.B.Thesis. University of Aberdeen.
Warburton.F.W.	- 1964 -	"Attainment and the School Environment" Chapter 6 of"Education and Environment" (pp.115 and 116)	- Manchester University Press.
Wiseman S.	- 1964 -	"Education and Environment".(a) pp. 138-146. (b) pp.173, 174 and 178.	- Manchester University Press.
Vernon.P.E.	- 1960 -	"Intelligence and Attainment Testing" (p.144)	- University of London Press.
Vernon.P.E.	- 1969 -	"Intelligence and Cultural Environment" (pp218 and 68)	- Methuen and Co. London.

6. GENERAL SUMMARY AND CONCLUSIONS.

(The numbers given on the right hand side throughout this section denote references to earlier pages of this report). Page No.

1. As far as I am aware, there has been no previously reported study concentrating on the relationship between accuracy or speed in attempting standardised tests taken at the age of 11 and later performance in Secondary School. This enquiry therefore provides evidence on an aspect of individual differences not previously studied.

2. It is submitted that "accuracy under pressure" and the relationship with schoolwork is relevant for work in the classroom at the present time. It seems reasonable 2
to suppose that the emphasis which teachers place upon accuracy will influence pupils when they have to make a choice between accuracy or speed in their work.

3. One of the major difficulties encountered when using tests to assess individual differences (and hence to predict probable future performance) is to try to ensure that the person taking the test is strongly motivated to suceed. It is submitted that this 2
assumption can be made in the present study as the taking of the "11+" examination is optional in the County concerned.

4. With the exception of the Composition paper, the tests used in the "11+" examination reviewed were all prepared and standardised by the National Foundation For Educational Research or Moray House. Bearing in

Page No.

mind previous research findings, the Composition Paper was marked by 4 independent judges, as suggested by Professor Wiseman. The scores were then standardised using the N.F.E.R. English Test as a basis and an age allowance was given. 57

5. The scripts of 1812 candidates were examined and the Heads of 27 secondary schools were asked to provide "order of merit" lists for "overall suitability for academic education", "English" and"Mathematics" judged at the end of the third year of the secondary course. In order to 40 standardise these assessments 1600 N.F.E.R. Secondary Verbal 2. Tests were given to all members of the third year age groups in the schools concerned. Because 41 some of the assessments were available only as "A", "B", or "C" gradings, in the end a detailed study was made of 553 children in 11 Secondary Modern Schools and 6 Secondary Grammar Schools. 43

6. Before considering accuracy and speed in detail, it was necessary to study the"11+" examination itself. First a review was made of the history and development of the examination in general.

(a) Principal factors which led to an examination were noted:-

(i) Historically Grammar Schools have provided virtually the only route to the Universities and Professions. 5

(ii) While Grammar Courses were appropriate for a minority of pupils only, such children were nevertheless to be found among all classes of the population. 6

(b) Main factors which fixed the age at 11 were listed:-

(i) There was a need to standardise the age of entry to Grammar Courses. 6

Page No.

(ii) In the 1920's psychologists paid special attention to mental changes associated with "adolescence". 6
It should be noted that Entwistle has recently shown that there is no evidence to support the notion of a spurt in mental ability associated with the onset of puberty. 122

(c) Factors influencing the content of the examination and selection procedures were mentioned:-

(i) The need for the Authorities to be fair to all candidates and to be seen to be fair to all. 9

(ii) Administrative convenience, in that standardised tests can be marked easily and give reliable results. 10

(iii) The demand for more social and educational equality. 4

(iv) Economic difficulties in the immediate post-war period; the shortage of buildings and of teachers. 5

(v) Changing views of the nature of "intelligence". 4

7. It is considered that the "11+" examination reached its stage of finest refinement in the late 1950 s and early 1960 s. A review of the development of the examination was carried out.

(a) Early investigators showed that the correlations between the Entrance Examinations for Secondary Schools and later School Certificate results were of the order of r = 0.40 10

(b) McClelland, reporting in 1942 on a review in Scotland where transfer was at the age of 12, found correlations with performance 3 years later to be r = 0.738 for a battery consisting of standardised tests 14
of "intelligence", English and Arithmetic. This was just slightly lower than if "traditional" type tests of

Page No.

attainment were used. The correlation with the "intelligence" test alone was found to be r = 0.690. 14

(c) Rutter, in England, found similar results, when comparing the Entrance Examination with School Certificate results. Again the importance of the "intelligence" test was noted. This had been the sole measure used for prediction in the years 1940 - 1943 and had given correlations of the order of r = 0.6 15

(d) The part to be played by Teachers' Estimates was stressed by Bosomworth in 1953. 18

(e) In 1955, Wrigley gave a very valuable summary of researches in England and Northern Ireland, showing the prime position of "intelligence" tests as predictors of probable later performance in Secondary Schools. 22

(f) The findings of a major project were reported by Yates and Pidgeon in 1957. They found raw correlations of the order of r = 0.75 between Secondary School rankings at the end of the 3rd. year and Primary Heads' scaled assessments on entry. Third year positions correlated 24 approximatelty r = 0.7 with Verbal Reasoning Test Quotients on entry. The correlations between standardised tests of English and Arithmetic in the Entrance Examinations and the third year rankings were of the order of r = 0.65. "Boosting" the coefficients to take account of the effects of selection raised these figures to r = 0.91, r = 0.89 and r = 0.86 respectively. 24

They went on to show that even with these high levels of correlation, misallocations of some 10% of the candidates would still occur. 25

(g) In 1957 also, a group of eminent psychologists published a book dealing with Secondary School Selection.

Page No.

Among other things they noted that "intelligence" tests were not measuring innate ability. 28

The move away from the "11+" examination gathered momentum after 1957.

8. Purely as a matter of comparative interest similar studies in Tertiary Education were noted, where the value 29
of "intelligence" tests as predictors of later success has been found to be less even than that reported above in surveys dealing with Secondary Education. 30

9. Again, as an aside from the main study, figures were given relating to the stability of measures of Verbal Reasoning between the ages of 11 and 14 years. Correlations of the order of r = 0.8 reported by N. France 31
in 1964 were matched by the findings of the present survey, where the correlations were of the order r = 0.76 60
A note has been made of evidence from the United States 59
and Scotland and also that all such test measures should 117
be thought of as indicators of a "kind of region or general level". 59

10. Details were given of the procedure adopted in the 35
County to "select" children for admission to Grammar 36
Schools in 1966. It was shown that approximately 30% of the age group were deemed to have qualified for admission to such courses. 37

11 (a). The "follow up" review of the main "11+" procedure which has been described in this study was not the main purpose of the survey. It was considered sufficient to show only that significant correlation existed and it is submitted that this has been done. A summary is given below. Three different methods of assessment of relationship

Page No.

were used: rank-order correlation for small schools: product-moment correlations "within" schools and using scaling based upon "T" scores to pool correlations "between" schools : and product-moment correlations using samples drawn from both small and larger schools.

The first figures noted relate to samples drawn from the different schools. The third year Secondary Rankings for "overall" suitability for an academic course were scaled using the Secondary Verbal Test given to all the age group. These rankings were then compared with the "overall" order of merit lists obtained from the"11+" examination in 1966. Product-moment correlation coefficients were calculated.
For 74 Boys the figure found was $r = 0.678$
For 245 Girls the figure was $r = 0.774$ 47

Both of these were found to be significant at the 0.01 level of confidence using the null hypothesis and (N - 2) degrees of freedom. This simple test was used because of pressure from limited time available. The results from the Secondary Schools were not available until the end of July, 1969. Conversion to Fisher's "z" would have been preferable.

It was found that the distributions of scores "within" the Secondary Schools were skewed, due, probably, to the effects of selection and the fact that the "11+" examination was "optional". Distributions were therefore "normalised" by converting all scores to "T" scores; these were used 42
to "pool" correlations to obtain figures for larger samples. When the correlations based upon these conversions were calculated, the boys' sample related to one school only.
For 29 boys a figure of $r = 0.778$ was found. 45

Page No.

When a similar calculation was carried out for girls between the total "11+" marks and the 3rd. year Secondary overall rankings, scaled on the Secondary Verbal Test, using the "T" score conversions to "pool" data obtained within 3 Modern and 1 Grammar Schools the correlation found was $r = 0.574$ (relating to 121 girls). 47

Both of these correlations are significant at the 0.01 level of confidence, using the null hypothesis and (N - 2) degrees of freedom.

For small samples, ranging in size from 10 to 25 children within 5 schools rank-order correlations ranging between $r_s = 0.4$ and $r_s = 0.9$ were found. All but one were significant. 46

It is submitted that significant correlation has been shown between the order of merit determined by the "11+" examination in the County in 1966 and the "overall" assessments for both boys and girls in grammar and modern schools three years later.

11 (b). It will be noted that these figures were lower than many found in previous studies. There are many probable reasons and it is suggested that account should be taken of the following:-

(i) Information was available about relatively small numbers of children in several different schools. 39

(ii) The samples from two of the girls' grammar schools were notably homogeneous. Approximately 70% of these girls were found to have Secondary Verbal Reasoning Quotients within a narrow 12 point band and original "11+" order of merit total scores within a group of 40 points. 46

Page No.

(iii) Fewer schools than expected actually assess their third year pupils by means of examinations and even then, rarely by one common to all members of the age group. 39

(iv) Ideally, a more comprehensive battery of tests should have been used to standardise the secondary assessments. As it was, there was only time and a willingness on the part of the schools involved to administer one Secondary Verbal Test. 39

(v) As the information from the Secondary Schools was not available until the end of July and early in August 1969, there was not time to extract more detailed information from the data and only very simple correlation calculations could be carried out. 39

(vi) Changing teaching methods, based less upon conformity and learning from the authoritative statements of teachers may well influence correlations with standardised objective tests in which only the generally found answer is treated as "correct". The latter may well favour "convergent" thinkers. 75

(vii) Experiments in changed organisation, such as "team teaching" and "mixed ability" groupings are in progress in several of the schools studied. These affect the work of 1st., 2nd. and 3rd year forms. They probably also affect correlations with entrance examinations taken when the primary schools were more "traditional" in their approaches to education.

12. In the survey, correlations were found between ranking based upon one Verbal Reasoning Test taken at the age of 11, in 1966 and the "overall" Secondary Assessments made in 1969. Though significant at the 0.01 level of

Page No.

confidence, these were found to be lower than those based upon the examination at"11+" as a whole.
For 29 Boys the figures were $\underline{r = 0.704}$ compared with $\underline{r = 0.778}$ 49
For 213 Girls the figures were $\underline{r = 0.187}$ compared with $\underline{r = 0.341}$ 50
In both cases conversions to "T" scores were used to pool data.

13. In the sample studied, correlations were found between Secondary Assessments for Mathematics in 1969 and Quotients obtained from the 1966 Arithmetic Tests. Using "T" scale conversions to pool data from different schools these were found to be:-

For 196 Boys, $\underline{r = 0.449}$
For 245 Girls, $\underline{r = 0.457}$ 52

(Both were significant at the 0.01 level of confidence, using the null hypothesis and (N - 2) degrees of freedom.

14. Within the sample studied, correlations were found between Secondary Assessments for English in 1969 and Quotients obtained from the 1966 English Tests. Using "T" scale conversions to pool data from different schools, these were found to be:-

For 137 Boys, $\underline{r = 0.547}$
For 245 Girls, $\underline{r = 0.411}$ 55

(Both were significant at the 0.01 level of confidence, using the null hypthesis and (N -2) degrees of freedom.

15. Rank-order and product-moment correlation coefficients were calculated for 8 small samples of children to find the relationship between scores obtained in the 1966 Composition Paper and the 1969 Assessments for English. These were found to vary

Page No.

from $r = 0.250$ (not significant) to $r_s = 0.827$. The latter was significant at the 0.01 level of confidence. The tendency was for significance to be established within the Secondary Modern Schools. The Grammar School samples were too small for any valid conclusion to be drawn. 57

16. The probable reasons for lower levels of correlation than those generally found in previous studies and which are listed under paragraph 11 above apply equally to the calculations mentioned in paragraphs 12 - 15 inclusive. 132 133

17. For the children in the samples studied, consideration was given to the ACCURACY, SPEED and a new measure for which the name CONFIDENCE was proposed, which they displayed in the way they tackled their scripts in the "11+" examination.

18. ACCURACY in attempting the"11+" standardised tests was defined as % correct of those questions attempted. It is submitted that for the first time it has been demonstrated that children at the age of 11 display consistent standards of Accuracy in answering mental tests of widely differing content (Verbal Reasoning, English and Arithmetic) taken over an interval of more than one month. 62 108

19. SPEED in attempting the "11+" standardised tests was defined as % attempted of the total number of questions possible. 62

20. CONFIDENCE, or "willingness to risk an answer", was defined as the ratio of the number of questions attempted compared with the ordinal number of the last question attempted. This ratio was expressed as a percentage. It is suggested that there may well be some value in such a measure in tests in which time-limits play a significant role, as they were thought to have done in the "11+" English Tests in the present study. For 88 girls in 1 Grammar and 2 Modern Schools a significant product moment correlation, based upon "T" scale conversions, of $r = 0.309$ (0.01 level) between "Confidence" in tackling the "11+" English Tests and later Secondary Assessments for English was found. 62 96 96

Though other significant correlations were found, they were not as closely associated with later success as either measures of Speed or Accuracy.

Other comments on the influence of "guessing" on the validity of test scores were noted. A.W. Heim and K.P. Watts found naval ratings loath to guess when faced with "open ended" tests, though prepared to do so if given a multiple choice of answers. In studying Secondary Schoolchildren in Scotland, Jennifer Nisbet concluded that for maximum incentive, tasks should not be perceived as either too easy or too difficult. Clear preferences for intermediate levels of risk have also been found in work at Princeton University. 76 112 113

21. Speed and "confidence" measures are both affected by the gradient of difficulty of the test. This aspect was studied for the "11+" tests in the present enquiry by examining indices of facility and discrimination for the items and also the patterns of

Page No.

answers and omissions. It was concluded that generally the later questions were more difficult and that the vast majority of the children worked steadily through the tests. 64

22. From a study of the pattern of answers and omissions, it was concluded that the English Test used in the "11+" examination was found more difficult, at least by children subsequently admitted to Secondary Modern Schools, than the tests of Verbal Reasoning or Arithmetic. 90

23. A review of the published accounts of studies of Speed and Accuracy in tackling "intelligence" and other tests was given. Principal points noted were:-

(a) The findings of Sutherland, that the time taken to solve a problem seemed to be governed by the number of steps required. He suggested a natural preference in 66
everyday life for certain speeds of working and considered that these would be overruled in the test situation by the condition of working at maximum pressure. Similar evidence was found by Slater, reviewing 450 children of 67
secondary school age. Duncan Howie, working in Australia with children of the same age as those in the present enquiry, found a speed factor, indicating a difference trait of tempo in functioning. With ability 70
held constant, accuracy appeared as the converse of speed and to be linked with the degree of error possibilities. 71
It is submitted that this reflects other findings relating to "the number of steps involved". Fruchter, albeit working with adults, found a factor unique to error scores which

Page No.

he labelled "sequential reasoning". Banks, with children about to enter secondary schools, found a link between the ability to attain concepts and the ability to use economical systematic strategies in tackling problems. 113

(b) The effects of personality differences, in that extroverts may work quickly but inaccurately and also build up inhibition, so reducing speed as the test progresses, has been studied by Eysenck, Ley and others. 69 Among older students the existence of fast and slow "starters" has been shown by Liam Hudson.

(c) The contrary aspect, that of using measures of inaccuracy to detect emotional instability has also been studied by Hudson. With boys slightly older than those in the present enquiry, he found inaccuracy linked with aggressive tendencies. 110

A small survey was therefore carried out as part of the present study. This related to 34 boys and 93 girls in 5 Secondary Modern Schools. No apparent link was found between inaccuracy in attempting the "11+" tests and the Heads' assessments of aggressive tendencies. 111 Nor was any link found between inaccuracy and the choice of subject for the "11+" Composition Paper, although one of the possible subjects seemed to offer scope for fantasy. 111

(d) The probability that time limits may unfairly handicap slower working students has been shown by Yates, working with young adults in Australia. It has 67 also been pointed out by Lele, studying secondary schoolchildren in Dundee. 71

Page No.

(e) The fact that error scores obtained from the "11+" examination tests could be used to predict later performance in secondary schools was noted by S.C. Richardson, who studied the cases of 155 boys. 73
Correlations between error scores and later performance have also been found in Tertiary level studies where coefficients ranging from 0.2 to 0.4 have been obtained. 74

24. In the present study, it has been shown that Accuracy in tackling the "11+" tests was much more closely related than Speed to final placing in the 1966 order of merit for English and Arithmetic. In the cases of 111 Boys and 112 Girls, product moment correlation coefficients of the order of r = 0.90 were found between Accuracy and final placing in the same test, whereas they varied from r = 0.453 to r = 0.807 for the relationship with Speed. 79

25. In the present study, when the measures obtained from the "11+" Verbal Reasoning Test 15A were compared with the "overall" rankings in the Secondary Schools after 3 years and "T" score conversions used to pool data, the findings reported were:-

		Significance level	
Accuracy:	Product moment		
29 Boys	r = 0.758	0.01	81
213 Girls	r = 0.251	0.01	
Direct Quotient:			
29 Boys	r = 0.704	0.01	50
213 Girls	r = 0.187	0.01	
Speed:			
29 Boys	r = 0.504	0.01	87
213 Girls	r, not significant.*		

*So many children attempted all the questions that it was not possible to calculate a "speed" score for girls.

Page No.

26. In the present study, when the measures obtained from the "11+" Arithmetic Tests were compared with the rankings for Mathematics in the Secondary Schools 3 years later and "T" score conversions were used to pool data, the findings reported were:-

Accuracy:	Product moment.	Significance level.	
195 Boys	r = 0.406	0.01	
245 Girls	r = 0.448	0.01	83
Direct Quotient:			
195 Boys	r = 0.449	0.01	
245 Girls	r = 0.457	0.01	52
Speed:			
67 Boys	r = 0.245	n.s.	90
108 Girls	r = 0.157	n.s.	89

27. In the present study, when measures obtained from the "11+" English Tests were compared with the rankings for English in the Secondary Schools 3 years later and "T" score conversions were used to pool data, the findings reported were:-

Accuracy:	Product moment.	Significance level.	
137 Boys	r = 0.507	0.01	
245 Girls	r = 0.358	0.01	85
Direct Quotient:			
137 Boys	r = 0.547	0.01	55
245 Girls	r = 0.411	0.01	
Speed:			
67 Boys	r = 0.478	0.01	
140 Girls	r = 0.283	0.01	91

Page No.

28. In the present study, comparisons of "matched pairs" were made. These pairs were matched for sex, age in years and months and verbal reasoning ability at the time of the "11+" examination. In all, 28 pairs of boys and 42 pairs of girls from 2 Grammar and 2 Modern Schools were reviewed. On the basis of examination on the "CHI - Squared" test it is submitted that the more accurate partner was significantly more successful in obtaining a higher ranking 3 years later 105
whether Secondary "overall" ratings, English or Mathematics Assessments were considered.

In the very small number of cases in which the partners forming the pair were not only matched for sex, and verbal reasoning ability at "11+" but were also matched for ability in English or Arithmetic the odds were at least 3 to 1 in favour of the more accurate partner achieving a higher Secondary Ranking 3 years later. 106
To obtain the same raw score, it follows that the more accurate partner must have been working more slowly.

29. Conclusions relating to Accuracy and Speed.

One must note the reservations that the samples mentioned in paragraphs 25 to 28 are limited in size and that adjustments were made to distributions of scores to normalise the data. In addition, the significance of the correlation coefficients found was tested only by using the null hypothesis and (N -2) degrees of freedom.

Nevertheless, it is submitted that the evidence reviewed leads to the conclusion that Accuracy is more important than Speed in the sample studied, whether one considers placing in the same examination or rankings obtained 3 years later after work in the Secondary Schools.

30. The question may then well be asked as to the extent to which these findings relating to Accuracy in tackling the "11+" tests compare with other factors known to influence performance in Secondary Schools. In Section 5, therefore, a brief account was given of other recent research involving the assessment of Secondary School work. The Factors found to affect performance include parental interest and support, the cultural level of the home, the influence of peer groups both in and out of school, socio-economic conditions of both the home and the school environment, progressive methods of education, arrangements to ease the shock of transfer from primary to secondary schools, streaming based upon ability, academic motivation, social maturity, tendencies on the part of the child to neuroticism or extroversion and individual differences in verbal reasoning ability.

It is submitted that the relationship found in the present enquiry between Accuracy in completing the "11+" tests and Secondary School Assessments made three years later, falls somewhere in size in the middle of the range of associations found due to these other factors. It would appear to be less important than parental attitudes and encouragement, but more significant than socio-economic rating or tendencies to neuroticism.

31. Because of the work involved, measures of accuracy seem unlikely to be added as additional predictors to the already highly refined batteries of standardised tests which go to make up the usual "11+" examination. It is submitted that there may well be some value in considering assessments of accuracy in Borderzone procedures, where more or less equal candidates are under review.

32. As has been noted on page 119, the situation involved in the prediction of academic attainment in secondary schools is a multivariate and dynamic one. The pupil produces some of the forces and interacts with others. Future progress is likely to involve teams of researchers and the analysis of data using a computer.

Despite the reservations about small samples and simple correlation techniques it is submitted that the present enquiry provides useful evidence relating to accuracy in completing tests as a significant feature. It is also suggested that there are important implications for practising teachers relating to the choice which may have to be made between accuracy or speed in the completion of classwork.

BIBLIOGRAPHY

For greater convenience, details of articles to which reference is made are given at the end of the relevant section.