I. MACFARLANE SMITH, Ph.D., M.A., B.Sc., Ed.B.

SPATIAL ABILITY

Its Educational and Social Significance

UNIVERSITY OF LONDON PRESS LTD
Warwick Square, London E.C.4

Dedicated to Joan

PRINTED AND BOUND IN ENGLAND BY
HAZELL WATSON AND VINEY LTD
AYLESBURY, BUCKS

FOREWORD

by Professor P. E. Vernon
PROFESSOR OF EDUCATIONAL PSYCHOLOGY, INSTITUTE OF EDUCATION OF THE UNIVERSITY OF LONDON

THIS is a book which needed writing; and Dr. Macfarlane Smith, who has been closely identified with its subject-matter for over one quarter of a century, is clearly the person to write it. At first sight it would appear to be a highly technical survey of the statistical findings of certain mental tests. But the conclusions which the author draws from his careful weighing of the evidence have very important implications for current educational policy. It is high time, therefore, that educationists should take the trouble to acquaint themselves with this technical evidence, to ponder and act on it. Briefly stated, Dr. Macfarlane Smith's thesis is that British education, particularly that given in grammar schools, while stressing the development of general or all-round intelligence, has over-valued the verbal type of ability at the expense of its psychological opposite—spatial ability. The Crowther Report, Sir Charles Snow and many other public figures have, of course, urged the claims of mathematical, technical and scientific education, together with Britain's need for technologists and scientists. But few of such advocates possess any scientific knowledge of the nature of these abilities they wish to encourage, what is their common essence, nor how this essence is related to other abilities or to temperamental traits and personality qualities. Nor are they, perhaps, sufficiently aware that our current system of selection for secondary and university education actively discriminates against the pupil or student who is most likely to be talented in these directions.

Dr. Macfarlane Smith outlines a large body of work on spatial, performance, mechanical and other non-verbal tests and shows that there is a major underlying factor or type of ability which is best defined as the capacity to perceive and hold in mind the structure and proportions of a form or figure, grasped as a whole. This view reconciles the somewhat divergent results of British and American workers, since the latter have often used less appropriate multiple-choice tests involving recognition of details rather than perception and reproduction of complex wholes. There is ample evidence of the usefulness of such tests in selection for technical courses and training, for geometry and art. But in addition a comprehensive survey of work on mathematical aptitude indicates that, apart from general (preferably non-verbal) intelligence tests, the most predictive tests are also those of the spatial factor. In contrast, mechanical arithmetic tests give very little indication of future mathematical or scientific ability (hence Crowther's advocacy of 'numeracy' is psychologically misleading). It would seem that the perception of form is a general characteristic of the abstract thinking involved in mathematics and science, as distinct from the verbal thinking involved in most school subjects.

A good deal of interesting work is surveyed, also, on defects in spatial ability associated with brain injury, cerebral palsy and leucotomy; and a discussion of the relations of this ability to types of attention (analytic vs synthetic) and to EEG brain waves throws further light on the neurological and mental processes involved. Finally the author makes a strong case for some relation between the ability and temperamental qualities akin to introversion, masculinity and initiative. The lack of understanding between the scientist and the humanist probably arises from the fact that their modes of thinking are intimately bound up with their whole personality organization.

The book covers much controversial ground and not all psychologists will endorse all of Dr. Macfarlane Smith's interpretations. I myself wonder, for example, whether some of the correlations and factor loadings he quotes are not too small to justify some of his more novel conclusions; and I would see more virtues in the verbal type of thinking than he seems to allow. But it is all to the good that his evidence and arguments should be presented and marshalled logically, since this will stimulate others to undertake further badly needed research. I would, then, particularly commend the book not only to educational policy makers, but also to research students in education and psychology who are searching for fresh ideas to explore.

January, 1963 PHILIP E. VERNON

ACKNOWLEDGMENTS

It is not possible to name individually all those who have helped in connection with the researches which form the basis of the present work.

My first introduction to a test of spatial ability occurred when I was a student at the University of Glasgow. Mr. C. A. Oakley, Lecturer in Industrial Psychology, requested volunteers to take a number of aptitude tests and I offered my services as a guinea pig. One of my fellow students who also took the tests and who has since taken a keen interest in my researches on spatial ability was Dr. A. K. Cairncross, the well-known economist. My long-standing interest in the subject seems to date from this incident and from the discussions to which it gave rise.

I received my initial training in psychology in the department directed by Dr. R. H. Thouless, whose classes were a source of lasting inspiration. My interest in the subject had been aroused by the experience of acting as a subject for Dr. Thouless in his first experiments (1931) on phenomenal regression, now called the constancy effect.

Dr. Thouless was then concerned with the problem of demonstrating the existence of a group factor in the abilities underlying the experiments on constancy. Thus, not only did he direct my interest to the field of shape-perception but he also familiarized me with one of the early techniques of factor-analysis. Some years afterwards, it became apparent that Dr. Thouless' researches were more closely related to mine than we realized at the time. We had in fact been studying different aspects of the same factor.

During my first investigation on spatial ability, I was greatly assisted by Dr. W. Stephenson of the University of Oxford, who responded to a postal request by supplying samples of test-materials, some of which were identical with those used by El Koussy in his well-known investigation (1935).

During the second world war, I served as a R.E.M.E. officer,

responsible for servicing radar equipment, and I had little contact with the work of psychologists, but after the war Professor R. A. C. Oliver encouraged me to continue my pre-war investigations. When I joined the staff of the National Foundation for Educational Research as their first officer in charge of Tests Services, Professor P. E. Vernon acted as a supervisor of this research.

My interest in spatial ability has been sustained over the years because of my role as supervisor of the researches of a number of M.Ed. students, many of whom worked in this field and contributed in various ways, notably R. Edwards (now Professor), J. S. Lawes, C. Stewart, C. C. Taylor, F. R. Witty and J. Wrigley (now Professor). I am also indebted for some help to a former colleague, J. J. C. McCabe, and to former students R. Anderson and J. C. Gardener.

In 1959 I was awarded the Research Fellowship of the Institute of Education of the University of Durham and I took the opportunity of carrying out a follow-up study of the validity of secondary school allocation procedures. For help in the collection of data I am indebted particularly to Mr. L. Charnley and Mr. J. Clitheroe, Officers of the City of Carlisle Education Committee, to Mr. T. H. C. Walker, Headmaster of Tynemouth Technical School, and to Mr. T. C. Wonnacott, Headmaster of Higginshaw Secondary Modern School, Oldham.

During my tenure of the Research Fellowship I received valuable assistance from Professor Brian Stanley, Director of the Institute of Education of the University of Durham, and also from the Institute Advisory Committee on Research. It was during this period that the present work took shape.

Subsequently, when I was appointed Director of the Research Unit at Garnett College, Roehampton, Professor P. E. Vernon read the draft of the book and made numerous helpful comments and suggestions. Professor Vernon's influence on the work will be apparent to all who are familiar with his writings.

While the present volume is the direct outcome of my tenure of the Research Fellowship of the Institute of Education of the University of Durham, the views expressed and conclusions reached in it are largely the result of reflections which have taken place from time to time over the last thirty years. For these views I must accept full responsibility, though, of course, they have been fashioned by innumerable circumstances and influences in my educational and family background. In so far as they have merit, much of the credit must go to the teachers who have stimulated and inspired me in the past. It would be ungrateful not to pay tribute to my old University, for it was while I was a student at Glasgow that my interest in the subject was first awakened. Undoubtedly another formative influence was a family tradition which valued technical skill and which fostered a youthful enthusiasm for biography and clan history.

It is scarcely surprising that a student with such a family background and attending a University associated with names such as Watt and Kelvin should become interested in the nature and inheritance of scientific and technical abilities. Thus, it may be deemed not inappropriate if I conclude by quoting an ancient motto of the Clan Macfarlane referring to a traditional interest in the science of astronomy:

Numen lumen, astra castra. (The Lord my light, the stars my camp.)

I. M. S.

GARNETT COLLEGE,
DOWNSHIRE HOUSE,
ROEHAMPTON LANE,
ROEHAMPTON, S.W.15

CONTENTS

ILLUSTRATIONS

LIST OF TABLES

PART ONE

Chapter one

Spatial Ability and the Selection Problem

The shortage of scientific manpower

IT has often been said that, in an age of scientific revolution, a nation's economic progress and survival depend largely upon the quality of its scientific, technological and technical education. Hence, it is widely recognized that the present acute shortage of all grades of scientific and technical personnel is one of the most critical problems affecting the future well-being of Great Britain. While the shortage appears to be world-wide,* in Britain it approaches the dimensions of a national crisis. Numerous authorities have estimated that in proportion to the population, both the United States of America and Soviet Russia are producing several times the number of scientists and engineers that are being produced in the United Kingdom. C. P. Snow (1959) has estimated that if we compare like with like, putting scientists and engineers together, Britain is training at a professional level one Briton to every one and a half Americans and to every two and a half Russians. He has given the following figures of graduates trained per year (scientists and engineers combined): United Kingdom 13,000; U.S.A. 65,000; U.S.S.R. 130,000.

Snow states that the Russian output of engineers is now much larger than that of the rest of the world put together, approximately, 50 per cent larger. While only slightly more pure

* C. F. McCrensky (1958) *Scientific Manpower in Europe*. London, Pergamon Press.

scientists of all kinds are being trained than in the United States, in physics and mathematics the balance is heavily in the Russian favour.

Reliable comparative figures are difficult to obtain, however, and the writer has found that estimates vary widely. A British Labour Party statement of policy, dated October 1961, asserts that per head of the population, the United States is educating each year two to three times as many, and the Soviet Union five times as many, highly trained technologists as Great Britain. J. Vaizey (1961), of London University Institute of Education, has estimated that the number of newly qualified scientists and technologists per million of the population is no larger in Britain than in Yugoslavia. It is half that of Canada, a third that of Russia and about a seventh that of the United States. In spite of these differing estimates, there is general agreement that Britain is lagging behind other countries in the production of scientists and engineers. In technological education, Britain lags behind not only great nations like Russia and the United States, but also behind smaller countries like Switzerland, France and Western Germany.

Statements regarding the long-term demand for scientific manpower and Britain's capacity to meet this demand have been even more contradictory. In a report issued by the Stationery Office in London in 1956, estimates were given both of the short-term and long-term demands for scientific manpower in Britain. It was suggested that to maintain an annual rate of growth of 4 per cent in industrial output, it would be necessary to increase the number of qualified scientists and engineers from the 1956 level of 135,000 to somewhere in the region of 220,000 in 1966—an increase of over 68 per cent. A 70 per cent increase in the number of engineers would be needed and a 50 per cent increase in the number of scientists. To produce this increase, the output of trained scientists and engineers would have to have been doubled by 1960. A British Labour Party pamphlet, published in 1961, has suggested, however, that the official target for the output of scientists and engineers in 1970 is quite inadequate. It proposed that Britain should aim to produce a 15 per cent increase in the yearly output, which would mean a doubling in seven years and a three-fold increase in ten. This aim was considered none too high to meet the challenge of the seventies.*

In January 1962, the American National Science Foundation published a report on education in the Soviet Union, which gives comparative figures for the U.S.A. and the U.S.S.R. This 900-page report was compiled by Mr. Nicholas de Witt, an associate of the Russian Research Centre at Harvard. It concludes that the Soviet Union is now turning out twice as many scientists and engineers as the United States.

The current American space programme, according to the report, will absorb the services of all available trained staff. The present (1962) annual total of American science and engineering graduates is 90,000 compared with 190,000 in the Soviet Union. Before 1970 the Soviet total will be expanded to 250,000—over twice the anticipated number in the United States. In 1959, 57 per cent of all B.A. graduates in the Soviet Union were in engineering, science and applied sciences. The comparable percentage in the United States was 24.

The report disposes of the widely held belief that American and West European educational standards are higher than those in the Soviet Union. It maintains that Soviet higher education in science and engineering transmits about as much knowledge as, and at times more than, American or West European higher institutions. The 'time inputs' required in Soviet education, it states, moreover, are invariably higher than in the United States.

A more optimistic view of the situation in Britain was expressed in a report published in October, 1961, by the British Advisory Council on Scientific Policy, which suggested that the country will have a surplus of scientists and technicians by 1970. According to the Council's calculations, by 1965 Britain should have achieved the necessary annual output of qualified men and

* For an exhaustive analysis of data on Britain's scientific and technological manpower, reference should be made to the comprehensive treatise by Payne (Stanford University Press, 1960).

women and from then onwards, production would exceed demand. This forecast arose from an earlier estimate made in 1956 by the Council's Manpower Committee that an annual output of 20,000 qualified people should be achieved between 1966 and 1971.

The 1961 report suggested that this figure would be reached by 1965 and that the output of 30,000 may be achieved by 1972. It recognized, however, that several contingencies might falsify the figures, and that in any case, supply and demand in individual disciplines were less easy to equate.* There was likely to be, for example, a continuing shortage of mathematicians.

In its annual report, published in January 1962, the Council commented that in some quarters the earlier report of its manpower committee has been "seriously misunderstood". Despite a six-fold increase of new qualifications likely in the early 1970's compared with 1938 there will still be a dearth of men and women with scientific education to fill posts in management, administration and other professions generally. A very similar view was expressed by the Institute of Physics and the Physical Society in a memorandum to the committee on higher education (1962). The memorandum stressed that America and Russia were investing heavily in physics to ensure a supply of well-trained scientists.

"We are certain," it stated, "that the supply of such people in Britain will not be remotely sufficient in 1965 or in any period for which it would be worth making a forecast."

The serious shortage of teachers of mathematics might well be one of the contingencies which may falsify the forecast of the

* Since this passage was written, the Advisory Council for Scientific Policy in Britain has reported (October, 1963) that the figures would in fact be falsified. Instead of a surplus of qualified scientific manpower in 1965, there would probably be a shortage of some 28,000. There would be major shortages of mathematicians, electrical engineers and possibly of physicists and mechanical engineers. The report emphasized that the shortage would be mainly one of technologists.

A statistical summary published in 1963 by the United States National Science Foundation has shown that in America specialized manpower in science was growing at the rate of about 4·3 per cent per year. Estimates for 1970 foresaw a total of 4,000,000 persons working as scientists, engineers, technicians or science and mathematics teachers. In these four manpower categories, there would be by 1970 an increase of 1,300,000 over the 2,700,000 estimated for 1963.

Advisory Council's manpower committee. The Labour Party pamphlet from which we have already quoted has described the situation as "desperate" and has called for "desperate measures to relieve it".

With almost monotonous regularity, leading authorities in Britain have been drawing attention to the seriousness of the teacher shortage and have emphasized that the situation is deteriorating. Sir John Cockcroft (1961) has expressed his views in the following passage:

"Part of the reason for our present mathematical deficiencies is the fact that the country is short of mathematicians. Their employment has increased by 50 per cent in the past three years, due partly to the development of computers in industry, and the supply is quite inadequate for industry, Government and the schools. As usual, the schools are taking the brunt of the deficiency. Last year (1960), the direct grant schools were able to fill only 61 per cent of their mathematics vacancies. The teaching of the subject is now in a serious state. The supply of mathematicians coming from the Universities is far too low . . ."

Professor Bryan Thwaites (1961), of Southampton University, made the following comments: "The truth is that the whole profession of mathematics is like a very sick man, a man in a high fever and still restlessly active, but suffering even so from a wasting disease, advancing so fast that one hesitates to speak too loudly of recovery. . . . If no recovery in fact comes about, mathematical education as it exists today is likely to die a natural death within twenty years. So desperate is the situation. . . ."

Summing up, Professor Thwaites said that "for the Universities, the present staffing deficiency was at least one year's total output of Ph.D.s and the maintenance of staffs at their present level required a doubling of the annual output. For the grammar-type schools, the present deficiency equalled at least three years' total output of graduate mathematicians, and the maintenance of staffs required at least a doubling of graduates in departments of education. These figures took no account of the difficulties of secondary modern schools. . . ."

No doubt many factors have contributed to the present difficult situation in Britain, but it seems probable that a vicious circle has already begun to operate and that this will create a more difficult situation in the future. The growing demands of industry for qualified technical personnel are denuding the schools of the normal supply of teachers of mathematics and of science and the shortage thus created will in turn have an adverse effect on the future supply of scientists and engineers to industry. While the shortage of teachers of mathematics and science is particularly serious because of its long-term effects, there are acute shortages in numerous highly essential technological occupations. Draughtsmen, for example, appear to be in short supply, as also are chemical engineers and technicians (Alexander, J., 1959). There is an unsatisfied demand for technicians of all kinds.

The pool of ability

McIntosh (1962) has stressed the fact that there are considerable untapped reserves of talent which ought to be developed and has suggested that there is an urgent need for scientific investigation to discover the factors which prevent this development. In considering the arguments for increasing facilities for higher education, it is necessary to take account of the size of the reserves of talent. But any estimate of the size of the pool of ability must be qualified by the reservation that this figure is valid only for conditions existing at any given time. Under different conditions, the figures might be very different.

Teachers are well aware of the fact that there are limits to the intellectual achievements of their pupils. Some pupils seem to be incapable of mastering Latin, at least when taught by conventional methods, and never reach a level which would enable them to read and appreciate Latin literature in the original. There is a growing awareness that mathematics also presents a stumbling block and many apparently intelligent adults seem to be unable to make progress in abstract mathematics. It is possible, however, that if methods of teaching mathematics could be greatly improved, even at the primary stage, the size of the pool of mathematical ability might be substantially increased.

In McIntosh's view, one of the most urgent educational problems is that of finding out how many people are capable of doing these things, how many are not doing them and what are the factors preventing them from realizing their full potential. He defines a pool of ability as that part of a population which is capable of success in a clearly defined type of higher education. There is not one pool of ability, but a number of pools and they differ according to the criteria by which they are defined. It is a matter of concern in Britain to discover how much greater is the pool of mathematical ability than the present output of mathematicians. McIntosh rightly emphasizes that in studying either a population sample or an age-group, it would be folly to use only one measure to estimate a pupil's potential ability. The size of the pool of ability cannot be estimated from anything as simple as the I.Q. distribution. The I.Q. is merely one of several measures which might be used to estimate potential ability, and it is necessary to make revised estimates based on a pupil's actual progress at school. There is clearly a very great need to devise methods for identifying the different types of ability which are necessary for success in subjects such as mathematics and science.

The criterion by which 'success' is judged must also be considered with care. A distinction must be made between 'being successful' in a course of education and 'benefiting' from it. Some students may benefit from a course in which they may have very little to show in the way of examination results.

The analysis of abilities

Any discussion of the 'pool of abilities' and of the problem of identifying talent necessarily requires some understanding of the structure of human abilities and of the techniques involved in their measurement. Only a very brief reference to the basic principles of mental measurement can be given here and the reader who is not familiar with the field should consult standard works, such as that by Vernon (1950).

The most far-reaching attempt to place the psychology of mental measurement on a sound, quantitative foundation was that of Spearman who claimed that mental abilities could be analysed into factors, a general factor *g* and numerous specific factors ($s_1, s_2, s_3 \ldots$). He believed that for any individual, *g* was a measure of the mind's ability to educe relevant relations between ideas or to educe correlates corresponding to a given relation.

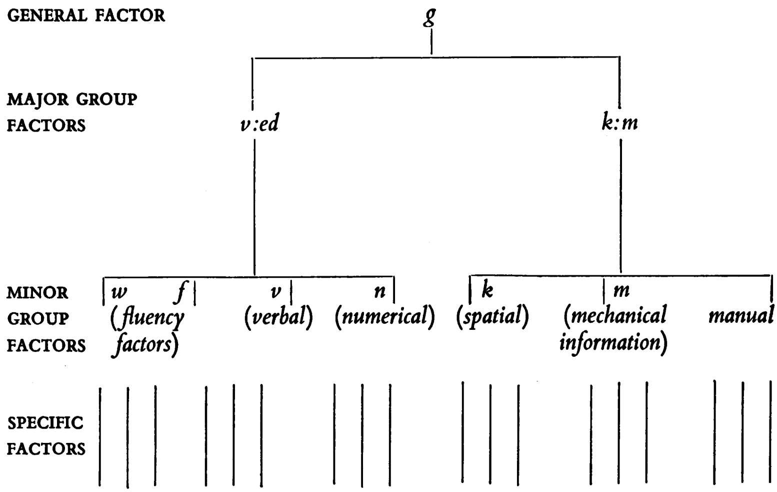
This two-factor theory of Spearman has provided a logical basis for constructing tests for measuring *g* and it has been found to give satisfactory results in practice. By analysing the correlations between tests it is possible to identify those tests which have high saturations in, or loadings of, the general factor *g*. Though each test has its own specific factor, when a number of such tests are combined together to form one test, the various specific factors tend to cancel out so that the total score provides a better measure of *g*.

More recent factorial studies have brought to light the existence of a number of group factors, which enter into some distinct abilities but not into all abilities, and which correspond fairly closely to the more important aptitudes. The evidence for the existence of these group factors is particularly strong in the case of the verbal, spatial and numerical aptitudes and the corresponding group factors are denoted by *v*, *k* and *n*. There is considerable doubt as to the existence of a unique group factor corresponding to mathematical aptitude.

It is possible to analyse mathematically the inter-correlations of a set of test-scores by a process known as factor-analysis. There are now many different methods of factor-analysis, but the results obtained are broadly equivalent. They consist essentially of systematic techniques for removing the effect of each factor in succession from the original table of correlations.

Usually, when the inter-correlations of the scores of a set of mental tests are analysed, the first factor to be extracted corresponds approximately to *g*. After the removal of *g*, the tests tend to fall into two main groups: the verbal-numerical on the one hand (sometimes referred to as the *v:ed*-factor); and the spatial-mechanical-practical on the other (sometimes referred to as the *k:m*-factor). If there are sufficient tests in the battery and the analysis is carried far enough, the two main groups sub-divide into minor groups factors: verbal *v* and numerical *n* on the one hand; and spatial *k*, mechanical information *m* and manual on the other.

Thus, we may think of human abilities as arranged hierarchically like a family tree, as shown in the following diagram.



Hierarchical structure of human abilities

The abilities needed by industry

The technical staff required in modern industry may be conveniently classified into four categories as follows:

1. Research scientists (including mathematicians and statisticians);

2. Technologists (both development and production engineers);

3. Technicians;
4. Skilled craftsmen.

There is a need to find ways of identifying potential abilities for any of these categories of occupations at a sufficiently early age to enable the talents to be developed to the full. Skilled craftsmen and technicians are likely to require among other qualities a high degree of spatial or mechanical ability, called by Vernon *k/m*. Research scientists, mathematicians and technologists will certainly require a high degree of general ability, usually denoted by *g*, but it is possible that they may also require some degree of spatial ability *k*. All four categories will require to have had at least a basic training in mathematics.

Thus, it is particularly necessary to find methods of identifying spatial and mathematical aptitudes at an early age. It has sometimes been said that the bottleneck in the present shortage of technical personnel in Britain is mathematics. There are not enough people studying mathematics. There are reasons for believing, however, that talent for mathematics may involve specialized aptitudes, which may be possessed by pupils who are not gifted in other scholastic subjects. This view has been expressed by McIntosh (1959), who made a number of case studies of such pupils in his follow-up study in Fifeshire. Clearly, the existence of such special aptitudes must be taken into account in the procedures by which pupils are selected both for grammar schools or technical schools or for grammar school courses or technical school courses. The type of secondary school organization, whether tripartite, multilateral, bilateral, or comprehensive, will have to be taken into account in considering selection procedures, but it will not affect the essential problem. Even in a comprehensive school, due account must be taken of a pupil's aptitudes, if he is to derive the maximum benefit from his education.

At the present time, the majority of scientists and technologists receive their secondary education in grammar schools. Many studies have been made of the efficiency of the selection procedures by which pupils are admitted to these schools. But, hitherto few of these studies have investigated the problem in the light of the shortage in the supply of scientific and technical manpower.

McIntosh (1959) has discussed the problems of guidance and selection in the light of the need to increase the production of pupils of university calibre, but he seems to have no radical suggestions to offer. The proportion of pupils who achieved a good Leaving Certificate in his age group of some 4,000 was 6·3 per cent and he estimated that with an ideal allocation procedure and full parental support, the figure might be increased to 11 per cent. These figures, however, relate to the situation in Fife in which 24 per cent of pupils are offered full secondary education. It cannot be concluded that a greater percentage might not reach the same standard with a different type of secondary school organization. Sir Geoffrey Crowther has repeatedly expressed the view that it is among those who fail at the early age of 11 plus that Britain loses most of her latent and badly needed talents.

Tests used in selecting for secondary courses

The tests which are most commonly used are known to be highly efficient for predicting all-round success in grammar-school courses. Thus McIntosh found that the combination of two verbal I.Q.s, semi-objective tests of attainment and scaled teachers' estimates in English and arithmetic yielded a correlation of ·872 with all-round marks in the fifth and sixth years of the senior secondary school. It is less certain, however, that these measures would be equally successful in predicting success in mathematics, physics and technical subjects. McIntosh found that teachers' estimates and the test of ability in English had greater validity for predicting marks in mathematics and science than had the test and estimates in arithmetic. This was true in the senior secondary schools, whereas in the junior secondary schools with their greater emphasis on practical subjects the relationship was reversed.

There is some evidence that some of the tests normally used for grammar school selection in Britain have *negative* value for predicting success in science courses at the university. Nisbet and Buchan (1959) followed-up a group of eleven-year-old pupils, who were eventually admitted to Aberdeen University. They compared the standardized scores obtained in the original selection tests with the results gained in the School Leaving Certificate examinations and in courses at the University (in first, third and final year examinations). The selection procedure employed in Aberdeen City involved the use of two verbal reasoning tests (I.Q.s), and attainment tests in English and arithmetic, as well as teachers' estimates.

Nisbet and Buchan calculated correlations between each of these scores, and the results obtained in the School Certificate examinations and in examinations in arts, science and medicine at Aberdeen University. They found that for the 1953 university entrants who took courses in science, their three criteria of success (first, third and final year) correlated negatively with the 11 plus attainment test-scores in English; two of the criteria (first and third year) correlated negatively with teachers' estimates and one of the criteria (for third year courses) correlated negatively with one of the verbal I.Q.s. Undue reliance cannot, however, be placed on these figures, since the number of students in the 1953 group was only 34, but a similar negative correlation between science and the 11 plus English score was found again in the 1954 group of 27 students, though it was not found in the 1952 group of 31 students. For the 1953 and 1954 entrants, the arithmetic score was a better predictor of success in university science than any of the other test-scores. All the test-scores were found to give positive, though moderate, correlations with criteria of success in university courses in arts and medicine.

It would appear from these findings that the 11 plus selection procedure in Aberdeen seems to operate moderately successfully from the point of view of its long-term validity for selecting potential arts and medical students, but it appears to give preference to pupils who are likely to be *less* successful in university science courses. Since the English tests tend to correlate negatively with criteria of success for science courses and positively for arts courses, there are grounds for suspecting that pupils who might succeed in university science are being rejected by the grammar schools because they lack the linguistic abilities required for success in the 11 plus examination.

University science courses appear to involve abilities which differ from those required for all-round success in the wider range of subjects taught in the grammar schools. Indeed, there is now some evidence that even for predicting success in the grammar school, the test batteries at present in use tend to over-emphasize linguistic abilities.

Nisbet and Buchan found that two of the 11 plus selection tests gave slightly negative correlations, both for science and medical students, with the number of passes in the Leaving Certificate Examination. Yates and Pidgeon (1957) have shown that the predictive value of the selection battery can be improved by including a spatial test in the battery. They studied numerous combinations of tests and estimates to discover the best combination of the predictor variables for predicting all-round success in the grammar school. Their criterion was the scaled estimates by the secondary school headmasters of the success of the pupils after two years in the secondary school. These estimates (based on the results of internal examinations) were scaled against scores in a verbal reasoning test given to all the pupils (876 in number).

The best prediction of all-round success was given by the combination 4 P.H.A + 2E + Sp.I, where P.H.A. is the Primary Head Teachers' Assessment, E is the score in a special English test by A. F. Watts, and Sp.I is the score in N.F.E.R. Spatial Test I (Macfarlane Smith). This weighted battery gave the highest correlation with the criterion, namely ·931.

Commenting on this finding, Yates and Pidgeon write (p. 72 of their Report): "The appearance of a test of spatial ability in this battery is deserving of comment. Such tests are not usually considered as useful predictors of success in grammar schools. It is possible, however, that the abilities that are measured by this kind of test are related to subsequent success in some branches of mathematics and science. It would seem to be desirable for further research to be undertaken to investigate this point."

If it should be confirmed that spatial ability is important for success in some branches of mathematics and science, it would follow that the 11 plus examination, as it is now conducted by most Local Education Authorities, is failing to identify an unknown percentage of pupils with an aptitude for mathematical studies. Pupils with high spatial ability may fall just below the border-line on the conventional combination of test-scores and may consequently be directed into secondary modern schools. There is no doubt that the tendency of current selection procedures (1963) is to reject rather than to select pupils of high spatial ability. Spatially gifted pupils are not represented in the grammar schools to the same extent as pupils of high verbal ability, as was shown by Dempster (1951) in an investigation carried out in Burton-on-Trent. The grammar-technical group of boys usually had higher verbal than spatial scores, whereas the modern school boys almost as frequently had higher spatial than verbal scores (roughly speaking in the ratio 2:1 in each case). Dempster made the following comment on his findings: "This evidence shows, of course, that the tests or examinations which had been used for selecting the grammar-technical group of pupils tended to give more weight to verbal ability than to spatial ability."

A similar result was obtained by J. C. Gardener, who administered verbal and spatial tests to 95 pupils in two Northumberland grammar schools. The only reason for testing this sample was that it could be done conveniently in the two chosen grammar schools.

Since the two tests (verbal and spatial) had been standardized with the same mean (100) and the same standard deviation (15), the standardized scores could be compared directly. In the sample of pupils tested, the numbers having standardized scores greater than or equal to 134, 131 and 128 on the two tests were as shown in Table I.

Table 1

Comparison of numbers of grammar school pupils obtaining high scores on verbal and spatial tests (Gardener)

NUMBERS OF PUPILS	M.H. VERBAL TEST (ADV. I)	N.F.E.R. SPATIAL TEST I
Scoring 134 *or above*	15	1
Scoring 131 *or above*	30	4
Scoring 128 *or above*	49	6

While it cannot be claimed that this sample of grammar school pupils is likely to be representative of all pupils in Northumberland grammar schools the discrepancies between the numbers of pupils obtaining high scores in the two tests is much too large to be ascribed to sampling error. There is no doubt that in these two schools at any rate, pupils with high spatial ability are not represented in numbers approaching those of pupils with high verbal ability. The reason, of course, is that the selection examinations tend to give an advantage to pupils with high verbal ability, because most of the tests are of a linguistic nature. A large proportion of spatially gifted pupils fail to gain admission to the grammar schools because they are handicapped in performing these tests. If it is true that such children are likely to include potential engineers, mathematicians and other scientific workers, then there should be a possibility that the supply of such workers may be increased by means of a suitable modification of the selection procedure. It might be supposed that pupils with high spatial ability who have failed to gain admission to a grammar school would probably be offered a place in a technical school. Such an assumption, however, would not be justified. In Britain as a whole, local authorities have been slow to make use of available tests of spatial ability.

While many authorities, such as Northumberland, have regularly administered both verbal and non-verbal tests of mental ability, true spatial tests have not been used extensively, except occasionally for the testing of relatively small groups of pupils recommended for transfer at 13 plus.

In 1951, the Chief Education Officer for Middlesex sent a letter to the L.E.A.s in 101 counties and county boroughs, enquiring about their methods of selection for secondary technical education. These L.E.A.s were chosen because it was known that they were responsible for at least one technical school in the area. Replies were received from 80 Chief Education Officers. These showed that in the great majority of cases, the usual battery of verbal reasoning, English and arithmetic tests was administered. A few authorities relied on tests of English and arithmetic alone, and a very few on a test of verbal reasoning alone. (In the latter cases, an interview and Primary Head Teacher's report was regarded as an essential part of the procedure.)

Only 18 of the 80 replies mentioned the use of a spatial test. M.H. Space Test I was mentioned 13 times and the recently published N.F.E.R. Spatial Test I was mentioned twice. Nine of the replies stated that some other spatial test had been tried at some time. Thus, Peel's Group Test of Practical Ability was mentioned six times; N.I.I.P. Form Relations twice; Peel's V.S. 17 once; four replies mentioned aptitude tests but gave no name; one mentioned a 'drawing interest and observation' test; and finally Birmingham referred to the very extensive battery of aptitude tests used in their vocational selection studies (Allen, E.P. and Smith P. 1931, 1934, 1939). Four replies stated that a spatial or other special aptitude test had been dropped, two of these mentioning that this was because of their high correlation with the more usual tests. Nine expressed doubt, advised caution or stated that they were using such tests only experimentally. Two stated that Peel's Group Test of Practical Ability appeared to be very satisfactory, and the Authority which used its own locally devised 'drawing interest and observation test' reported that it was very successful.

By 1952, some fifteen Local Authorities had used one or other of the recently published N.F.E.R. spatial tests. Since then, some nine Authorities have been regularly using N.F.E.R. spatial tests at the age of 11 plus as part of the normal selection procedure, though many more include a spatial test in the battery administered to the groups of pupils aged 13 who have been recommended for re-examination and transfer from modern schools to some form of selective secondary education.

In 1960, the National Foundation for Educational Research carried out a survey of the different procedures used by Local Authorities for selection at the age of 11 plus. Further enquiries were made of the nine Authorities who reported that they used spatial tests as part of their procedure. The replies showed that three of the nine no longer used spatial tests, two having based their decision on subjective impression only and the third on the result of statistical and other analyses though no details of this were given. Four Authorities used spatial tests with only part of their group. Two administered them to all children, though one of these made use of the results only in borderline cases.

Two follow-up studies had been carried out in Preston, one study showing that the spatial test was good for predicting success in technical subjects in the grammar school and the other study being less favourable. In a pilot study carried out by the National Foundation for Educational Research (Yates and Barr, 1960) in a county Borough (Wallasey) a variety of types of test were added to the standard battery of Moray House verbal reasoning, English and arithmetic tests. The criterion for the follow-up investigation was the head teachers' assessment based on cumulative records at the end of four years of the technical course. Of the tests used (Non-Verbal Test 3, Spatial Test 2, Watts' Group Performance Test, Clerical Test 1, and the Lambert-Peel and Devon Interest Tests), the spatial test gave the best prediction (·41), the degree of prediction not being significantly improved by any combination of tests.

Differential selection for grammar and technical courses

Reese Edwards (1960) has reported the results of an extensive survey, carried out between 1958 and 1960, of a large number of secondary technical schools. One chapter on the classification and selection of pupils is devoted to a discussion of the evidence for and against attempts at selection by means of psychological tests. He concludes from his investigation that opinion in England does not favour the use of spatial tests, though he quotes with approval statements to the contrary such as that made in a Report (1950) to Southampton Education Committee, to the effect that "where selection is for technical education alone, whether this is carried out at 11 or at a later age, spatial tests have been shown to have considerable value."

At the end of a detailed discussion of the selection problem, Edwards states his conclusions in the following words:

"It is here contended that the boy with a high level of *g* and relatively high verbal ability can successfully undertake either a course in a secondary grammar school or a course in a secondary technical school. . . .

"It can be said that:

1. A number of boys of good ability can take either a secondary grammar school course or a secondary technical course with a prospect of ultimate success.

2. Certain boys will be more successful in a purely secondary technical course than in a purely secondary grammar school course; and

3. Certain boys will be more successful in a purely secondary grammar school course than in a purely secondary technical school course, but that

4. It is exceedingly difficult, if not impossible, in the present state of our knowledge, to differentiate between boys in categories (2) and (3) at the age of 11 or 12.*

"When an age-group of 11 plus children is taken, it is possible, with a fair degree of accuracy, to separate on the basis of their past records in the primary school, the teachers' estimates of their ability and attainment, and of tests of *g* into those who are likely to succeed in a specialized course of secondary education and those who are not. After there has been a classification of pupils into those who can profit from a specialized secondary education course

* A somewhat similar view has been expressed by Gooch, cf. Gooch (1962) and the reply by the writer (1963).

and those who are not likely to succeed in such a course, then the final decision as to whether the grammar school course or the technical school course is to be taken should be made by the parents."

Thus, in Edwards' view, the decision as between grammar and technical school education should be made in *accordance with parents' wishes.* A consideration of aptitude would play no part in this most vital decision. He suggests, however, that it would be advantageous if the years from 11 to 13 should form a diagnostic period, during which all pupils should follow, as far as possible, the same type of course. A review could then be made at the age of 13 before a final decision is made as to which ultimate course, whether grammar or technical, the pupil should enter for the completion of his secondary education. This recommendation implies a postponement of the final decision until the age of 13. There is no mention, however, of the use of aptitude tests during the diagnostic period to facilitate the making of a decision. He does stress the fact that general ability is just as important for technical as for academic courses, and this is a point which has not always been appreciated in the past.

In a later chapter, he states that "Art and technical drawing are, in a sense, the key subjects in the boys' secondary technical school and craft subjects must be closely integrated with these, if the maximum value is to be derived from the teaching of crafts."

There is now abundant evidence that art and more especially technical drawing are subjects which require a high degree of spatial ability, as do most of the craft subjects, such as woodwork, metalwork and building crafts. On the other hand, it has been repeatedly shown that the majority of grammar school subjects, such as English, modern languages and social studies, depend in large measure on verbal ability. Thus, there are grounds for adopting the view contrary to that maintained by Edwards, that grammar school and technical school courses depend to a large extent on different, if not opposing, ability factors. It is true that both groups of subjects also involve the general factor *g* and the greater the contribution of this common factor the more difficult becomes the problem of differential selection. This problem will be discussed more fully in Chapter five.

Our results suggest that pupils who do well in tests of verbal intelligence (i.e. pupils whom Edwards describes as "having a high level of *g* and relatively high verbal ability") do not as a rule distinguish themselves in technical courses. Such pupils appear to advantage when success is judged by written answers to examination questions. When, however, the criterion depends on ability to do an actual job, whether involving an actual construction or its representation by means of a drawing, the possession of high verbal ability confers no advantage.

Edwards quotes statements made by some of the head teachers he interviewed to the effect "that a considerable number of pupils in their schools are found to be incompetent with their hands. Apparently, practical work is beyond them." The cure, Edwards thinks, is to be found by diverting "the energies of these pupils . . . into other channels, since it is inadvisable to let them continue the serious study of these subjects, if they prove to be unfitted to do so." It may be questioned whether the deficiency shown by these pupils is something lacking 'in their hands'. It is more likely to be something lacking 'in their heads' since spatial ability is quite as intellectual as verbal or numerical abilities are. It is very probable that those pupils 'whose energies should be diverted to other channels' have been committed to these courses because of the inadequacies of current selection or guidance procedures, in which undue reliance is placed on parents' wishes and little account is taken of the pupils' aptitudes.

The success of secondary technical schools

In spite of these serious defects in methods of recruitment, technical schools have fully justified their existence by providing a form of education which has satisfied a real need in the great majority of their pupils. It is interesting to find that Edwards reports on the work of the technical schools of 1960 in very much the same terms as the Spens Committee wrote in their report of

1938 of the education provided by junior technical schools. The Spens Committee stated that "we have found in the schools we visited an atmosphere of vitality, keenness and happiness that was not only refreshing, but afforded a sure index that the curriculum and its methods of treatment so appealed to the pupils that the process of education was developing smoothly and unrestrainedly."

Writing of the technical school of 1960, Edwards comments that, in spite of inadequate selection procedures and the general inferiority of accommodation and equipment, "it is everywhere noted for the success that attends its pupils and for the evident satisfaction that so many of its teachers find in their work. . . . Indeed, it has been found that with pupils of comparatively low intelligence, the constructive impulse, if given adequate expression and outlet, can have a remarkably stimulating effect upon other creative faculties, upon the expression of intelligence itself and upon progress in academic studies." (Edwards, 1960.)

Most heads of technical schools would probably interpret the success of their schools as being due to the obvious link between technical subjects and the needs of industry and commerce. They suppose that the technical bias gives a sense of purpose and reality to the work and so engenders a favourable attitude to work and a sense of responsibility.

It is possible, however, that there is an important contributory factor to the success of these schools which is not generally recognized. Many pupils, whose aptitudes are spatial rather than verbal and who have achieved only moderate success in ordinary scholastic subjects will discover on being transferred to a technical school that they can achieve much greater success in technical subjects. It is to be expected that such pupils with hitherto undistinguished records will be stimulated and encouraged and will communicate their enthusiasm to their teachers.

If this is indeed one of the secrets of the remarkable successes of these schools, it behoves administrators to ensure that their selection procedures are as valid as possible. Exclusive reliance on parents' wishes may well cause frustration among the unfortunate misfits as well as denying to many technically gifted pupils an opportunity to develop their talents to the full.

There are some 225 secondary technical schools in England and Wales and 63 bilateral or multilateral schools most of which provide technical courses. Nearly 100,000 pupils attend secondary technical schools and more than 45,000 attend bilateral or multilateral schools. It is likely that the number of pupils taking technical courses will increase very considerably as a result of the drive to increase the output of technically trained personnel.

The validity of the procedures by which pupils are selected for these courses is a matter of considerable national importance. There is a very great need to investigate thoroughly the validity of current selection procedures and in particular to assess what improvement, if any, can be achieved by making use of any of the existing tests of spatial ability or of practical interests. The work of the National Foundation for Educational Research in providing an information service and in providing facilities for the co-ordination of the efforts of research workers in this field is very much to be welcomed.

That there is dissatisfaction with existing methods of selection is illustrated in a letter with the signature H. E. Hopper, which appeared in *The Times Educational Supplement* for 3rd March, 1960. The letter requested advice and information regarding procedures for selecting pupils at the age of 11 plus for the new secondary technical schools and for later transfer of misfits. Only one reply appeared in the subsequent issue. It was a letter from the author in which he suggested that a properly designed spatial test should be included among the selection tests. There was no subsequent published comment on the subject of these two letters. May it be assumed from the absence of comment that no one had any more promising suggestion to offer?

Conclusions

It is difficult to avoid the conclusion from the foregoing discussion that most educational systems, including that in Britain, have been slow to adapt themselves to the rapidly changing needs of a society which is being transformed by the application of scientific discovery. That this conclusion is officially accepted in Britain can be gleaned from statements made by Government spokesmen, e.g., on 6th May, 1963, Lord Hailsham expressed the view that "we need to develop from being a nation of shop-keepers and colonial administrators to become a nation of technologists."

In the light of such pronouncements it is clear that Local Authorities have been slow to make use of spatial tests in their selection procedures. They have tended to rely almost exclusively on tests of linguistic and numerical abilities. While no one would deny that language plays an important part in intellectual development, by facilitating the growth of concepts of increasing complexity, we must beware of supposing that thought can take place only by means of words.

Certainly there are now grounds for believing that linguistic tests are not the most valid tests for selecting candidates for technical courses or for the more advanced courses in mathematics or science. It is, therefore, very probable that most of the existing technical schools are not securing the maximum number of pupils who can make adequate use of the facilities they provide.

It is true, of course, that there is now a tendency to break down the rigidity of the tripartite system and to provide courses with a technical bias in grammar, modern and comprehensive schools. This policy is to be welcomed but it does not absolve authorities from the responsibility for *identifying* technical talents and for ensuring that they are developed to the full. The view expressed as recently as 1962 by a prominent educationist that tests of spatial ability, however valid, are of little more than academic interest, scarcely accords with the needs of the times.

It is a matter of some urgency to ensure that the best facilities for technical education are readily available to *all* pupils who possess the requisite aptitudes. Only when action is taken to make accurate diagnoses of the abilities and aptitudes of pupils can it be claimed that the ideal of education according to ability and aptitude is any nearer to realization than it was in 1944.

Chapter two

The Spatial Factor and its Subdivisions

Early studies using simpler techniques (1917–1935)

THE reluctance of most Local Authorities in Britain to make use of tests of spatial or mechanical ability in connection with their selection procedures is a phenomenon which requires some explanation. To a foreign visitor, it must seem surprising that this attitude should prevail some twenty years after the passing of the Education Act of 1944, which laid it down that pupils should be educated in accordance with age, abilities and aptitudes.

To understand the widespread tendency to regard the verbal test as a measure of a pupil's 'intelligence' and to treat the standardized verbal score as an 'I.Q.', indicating potentiality for any type of secondary education, we must study the history of the evolution of the 'intelligence' test, particularly in relation to its use in educational selection. The original impetus to the development of mental tests arose from the need to improve methods of predicting scholastic success. Thus, the pioneering work of Binet was directed to the construction of tests which would separate successful scholars from those who would be unlikely to succeed in the ordinary school. Since scholastic success depends to a large extent on reading and writing, the tests which were found to be most successful for this purpose consisted, not unnaturally, of verbal material. These tests were called 'intelligence' tests at an early stage in their development and so it became customary to assume that intelligence was best measured by tests consisting of verbal questions. The great majority of the so-called intelligence tests constructed prior to the Second World War were of this type. It is true that many psychologists were fully aware of the fact that these tests tended to give a very great advantage to pupils who were gifted in the use of words. But it was not until 1931 that it was clearly demonstrated by Stephenson that there is a group factor of verbal ability distinct from general ability, and this demonstration gave an impetus to the development of non-verbal tests, such as 'Progressive Matrices', due to Penrose and Raven (1938).

Throughout this period, i.e. prior to the Second World War, technical abilities were conceived to be mainly practical or mechanical and the earliest procedures for selecting pupils for technical education were designed to measure aptitudes for manipulating mechanisms or other concrete materials. Among the many tests of manual dexterity which were devised were those consisting of peg-boards and eye-boards, tests of tapping and aiming, tests of tweezer-dexterity, of wire-bending, tests involving the manipulation of nuts and bolts, and many others.

One of the most successful of the early tests was the Mechanical Assembly Test devised by Stenquist, in which the task was to fit together pieces of familiar objects such as bicycle bells and locks. This test was found to give high validities (·8 to ·9) with boys of secondary school age when assessed against criteria of success in manual subjects. It was later modified and extended and called the Minnesota Assembly Test. It yielded correlations of about ·55 with marks for quality of practical work done in the high school.

During the First World War, it was administered to more than 14,000 recruits in the American Army, to provide a measure of non-verbal intelligence, additional to the verbal score of the Army Alpha test. It is interesting and most significant that the Army psychologists discontinued using the Assembly Test when they found that it yielded low correlations with the Army Alpha.

The latter was devised by Otis and was the first group verbal test to be used on a large scale. In this situation in which both a verbal and a non-verbal score were available, it was assumed that the verbal score provided the more valid criterion. It might have been argued that since the Assembly Test consisted of a miscellaneous collection of 'real-life' problems, it was the more valid test of general ability. The fact that it was the Assembly

Test which was dropped and not the verbal test may be regarded as evidence of 'bias' on the part of the psychologists. The Assembly Test had been shown to be valid when judged by the criterion of quality of work done, but the verbal test was preferred presumably because this type of test was more convenient to administer and was known to be the better predictor of scholastic success.

Clearly, the only valid criterion in this situation was that of efficiency from the military point of view; while the verbal test might be more successful in predicting success in a clerical job, the Assembly Test would almost certainly be a better predictor of efficiency in a purely technical arm.

Though verbal tests were generally described as 'intelligence' tests throughout the period between the wars, a different attitude prevailed from the beginning of the Second World War. It was not automatically assumed that the verbal type of test had the greater validity. Indeed in the British Army, the test used for general grading purposes was a non-verbal test—the 'Progressive Matrices' of Penrose and Raven. It was presumed, following Spearman's principles, that this type of non-verbal test provided a measure of *g*, the general intellectual factor uncontaminated by the verbal factor. But in both the British and American armed forces a very wide range of test-materials was used.

One of the earliest investigations of 'practical' ability was that by McFarlane in 1925. She devised a number of wooden constructional tests, like the wheel-barrow test, in which the parts of a wheel-barrow had to be fitted together. She also used the Cube Construction Test and Healy's Puzzle Box, which was rather like the boxes used by Thorndike in his experiments with cats. McFarlane found some evidence of the presence of a group factor additional to *g*, with her sample of boys, though not with the girls. She wrote that the performance tests "measured an ability whose uniqueness lies in the fact that those persons possessing it in high degree analyse and judge better about concrete spatial situations than do other individuals who perhaps excel in dealing with more highly abstract symbols." Her description suggests that she was referring to the same aptitude which is now more usually measured by paper-and-pencil spatial tests. In 1928, O'Connor devised the O'Connor Wiggly Block Test and claimed that it was highly successful in selecting prospective mechanical engineers. Since it consisted of a single test-item it could scarcely be expected to have high reliability and yet it gave validities of ·62 and ·42 for shopwork with two groups of boys, figures which are remarkably high for a test of low reliability.

In spite of these promising pioneering investigations many psychologists tended to think of performance tests as providing only rather unreliable measures of non-verbal intelligence. This was certainly the view adopted by Spearman. In *Abilities of Man* (1927), he quotes the results of Macrae as evidence that there is no common group factor underlying spatial or performance tests in general. He accounted for McFarlane's finding of a large overlap in her tests with boys, though not with girls, by supposing that this derived from acquired experience rather than from innate ability:

"Daily observation shows that many boys, unlike almost all girls, tend already in their second year of life to play with mechanical instruments in a very thorough way, which can scarcely fail to help them subsequently in all performances of a kindred nature."

Apart from this element of acquired experience in handling mechanical objects, Spearman tended to regard most performance tests as being somewhat unreliable measures of *g*. Kohs (1923) adopted a similar view in putting forward his Block Designs Test as a measure of general intelligence and, as late as 1936, Cattell wrote of performance tests as if they were unrelated to technical abilities.

Only very gradually was it realized that some of these tests measure a very important factor in addition to general intellectual ability. One of the early tests which showed special promise was the elaborate wooden form-board known as the Minnesota Spatial Relations Test, which was found to give a correlation of ·53 with marks in shopwork. Earle and Milner (1929) found some evidence suggesting the presence of a spatial or practical group factor in a number of performance and spatial tests (Cube Construction, Dearborn Form Board, Form Relations (N.I.I.P.), Memory for Designs (N.I.I.P.) and Stenquist Assembly).

Rodger (1937), working with Borstal boys, found that the Cube Construction Test gave the best results for predicting success in the mechanical trades. It was the investigation by Alexander (1935), however, which seemed to establish the fact that some performance tests do measure a factor over and above *g*. This study originated from his work in developing a performance test of nine items which he called the Passalong Test. He included this and other performance tests in a large battery which he administered to several groups of subjects. One group consisted of primary school pupils, another of secondary and technical school pupils and a third of adult women in a delinquent institution. It appears that Alexander was the first to apply Thurstone's recently devised Centroid Method to an investigation involving factor-analysis of abilities. Using this technique, he showed that some of the performance tests involved a factor additional to *g*, which he designated the *F*-factor. Later he developed a performance scale consisting of Passalong, Cube Construction, and Kohs' Blocks for measuring what he called 'concrete' or practical intelligence as distinct from abstract intelligence.

It can now be seen that the terms 'concrete' and 'abstract' to distinguish between spatial and verbal abilities were ill-chosen. This terminology, which had been used by McFarlane in her 1925 study, tended to suggest that pupils with exceptional spatial ability were less likely to be capable of abstract thought than verbally gifted children. Such an assumption, which appears to be quite unfounded, may have been responsible for much of the tardiness with which spatial tests have been considered for use in selection procedures by Local Education Authorities.

While on the subject of performance tests, mention must be made of the cognate tests, designed to measure mechanical ability. Early forms of these, devised by Cox (1928), were very elaborate and were both expensive and time-consuming. They consisted of mechanical models, the mechanism of which was concealed behind a screen, the subject being required to answer questions about the nature of the mechanism. Similar tests which were more easily administered were later devised by Vincent for the N.I.I.P. Brush found that Cox's Mechanical Models were the most useful of the mechanical tests he gave to students for predicting success on engineering courses. A simpler form of this test, originally devised by Stenquist (1922), consisted of pictures of systems of gears or pulleys, the task being to indicate what would happen when any of the wheels were made to revolve. Cox developed the idea in three tests, Mechanical Designs, Mechanical Explanations and Mechanical Completion. These were much more convenient to administer and Brush found that they were also successful in predicting success in engineering courses.

Paper-and-pencil tests of the type which we should now call spatial tests have been in use since about 1917. They seem to have originated as paper-and-pencil versions of wooden form-board tests rather like jig-saw puzzles. They were intended to be tests of general intelligence and were considered suitable for testing the intelligence of persons who had had very little education or who were thought to be poor in expressing themselves verbally. These spatial tests were used in investigations into methods of awarding scholarships to trade schools and technical institutes. A question which naturally arose from these investigations was whether the tests involved a special aptitude (i.e., a group factor over and above *g*), just as verbal tests have been found to involve a large verbal group factor additional to *g*. Much time and energy has been expended in an attempt to answer this question, and different investigators came to opposite conclusions at different times. The existence of such a factor was long denied by Spearman, who claimed (1950) that the spatial type of test was originally devised by him for the very purpose of measuring *g*. However, it seems to be generally conceded that the factor does exist and numerous American psychologists claim to have found several distinct spatial factors.

The history of research on the spatial factor (or factors) is closely bound up with work on the mechanical factor *m* and the practical factor *F*. In his study of mechanical aptitude, Cox (1928) found the mechanical factor present in mechanical models and explanations, in paper-folding and jig-saw tests, in practical trade tests and technical tasks. He identified the *m* factor with the mental activity involved in the comprehension of mechanical relations rather than in manipulative activities. The thinking has to be of the nature of correlate eduction.

At about the same time as Cox's study, Truman Kelley (1928) tested children in the age-range 10 to 16 and identified a spatial factor in the following tests:

1. speed in reading;
2. power in arithmetic;
3. memory for meaningful symbols;
4. memory for meaningless symbols;
5. manipulation of geometric forms.

In another experiment, he obtained evidence that the spatial factor could be separated into two parts which he designated:

ε—an ability involving the sensing and retention of geometric forms;

θ—a facility in the mental manipulation of spatial relationships.

He admitted, however, that the existence of the factor was doubtful in four of his five groups.

Earle, Milner *et al.* (1929), and Earle and Macrae (1929), published two relevant studies in the same year. The conclusion expressed by Earle and his colleagues was that "the special abilities entering into the performance tests under consideration appeared to be unrelated except in the case of those in which spatial elements enter. . . . These are related by a rather small group factor for spatial perception, as well as by the general factor *g*."

There followed a number of studies all tending to cast doubt on the existence of a spatial factor. Thus, Line (1931) concluded from his study of the growth of visual perception in children "that throughout the spatial tests there seemed to be no evidence of a group factor". Similarly, Fortes (1930) concluded from his analysis that "the tetrad difference criterion (two factors) proved that these tests measured only *g* and specific factors".

Stephenson (1931) concluded from a very extensive investigation involving 1,037 girls "that the evidence was against any group factor in the non-verbal sub-tests." In a later study, however, in collaboration with W. Brown (1933) he reported "some slight signs of a group factor on the border of significance". In the same year, Milton Smith (1933) reported having found a small group factor between a form-board test and Kelley's spatial test.

Later studies using more complex techniques (1935–1945)

The next important contribution was that of Alexander (1935) in the related field of performance tests, several of which were found by him to involve a factor *F* in addition to *g*. He called this a factor of practical ability.

In the same year an outstanding contribution was made by El Koussy (1935), in a research carried out under the guidance of Stephenson. In this very comprehensive investigation, 28 tests covering a wide range of abilities were administered to 162 boys aged 11 to 13, attending a central school in Sidcup.

The battery included the following spatial or mechanical tests:

Area Discrimination (El Koussy)
Memory for Designs (N.I.I.P.)
Form Relations (N.I.I.P.)
Fitting Shapes (Stephenson)
Form Equations A (El Koussy)
Form Equations B (El Koussy)
Form Equations C (El Koussy)
Overlapping Shapes (i) (Stephenson)
Overlapping Shapes (ii) (with directions, Abelson)
Pattern Perception (Stephenson)
Spatial Analogies (Stephenson's test modified)

Classification (Areas, directions and lines, Spearman)
Band Completion (El Koussy, after Spearman)
Correlate Eduction A (El Koussy)
Correlate Eduction B (El Koussy, after Spearman)
Mechanical Explanations (Cox)
Mechanical Completion (Cox)

The following reference tests for *g* were used:

Inferences (Verbal)
Alphabet Series (O's and A B C D E F G H)
Visual Perception (Spearman)
Comparison of Greys (Power, Spearman)
Comparison of Greys (Speed, Spearman)
Greys Analogy (El Koussy)
Letter Cancellation
Pitch Discrimination (Seashore)
Loudness Discrimination (Seashore)

To investigate a possible relationship between spatial tests and practical abilities, marks in school examinations in woodwork and drawing were also included.

El Koussy used a modification of Spearman's Tetrad-difference Technique, partialling out the influence of *g* from the table of correlations by means of the scores in the reference tests for *g*. He concluded that there was evidence for the existence of a factor in eight of the spatial tests but not in the others. He stated his main conclusion in the following words:

"There is no evidence for a group factor running through the whole field of spatial perception. . . . Spatial tests are primarily tests of *g*. But some spatial tests involve a group factor over and above their *g*-content. This group factor, called the *k*-factor, receives a ready psychological explanation in terms of visual imagery."

He mentions that the letter *k* was suggested by the first letter of the word 'kurtosis', though Burt (1949) has stated that it was originally applied to the spatial factor because kinaesthetic imagery was formerly believed to be essential for success in such tests.

The contrasting results for the two kinds of spatial test are shown in the following table.

Table 2

Factor loadings of spatial tests with and without significant k*-loadings (El Koussy)*

TESTS HAVING SIGNIFICANT *k*-LOADINGS	loadings *g*	loadings *k*	TESTS NOT HAVING SIGNIFICANT *k*-LOADINGS	loadings *g*	loadings *k*
Memory for Designs	·66	·58	*Classification (Spearman)*	·76	·22
Form Relations	·43	·49	*Area Discrimination*	·55	·00
Fitting Shapes	·62	·61	*Form Equations B*	·54	·17
Form Equations A	·67	·40	*Form Equations C*	·61	·25
Pattern Perception	·76	·61	*Overlapping Shapes (i)*	·67	·00
Spatial Analogies	·63	·50	*Overlapping Shapes (ii)*	·64	·00
Band Completion	·65	·46	*Classification (Spatial)*	·49	·16
Correlate Eduction A	·50	·62	*Correlate Eduction B*	·65	·01
			Mechanical Explanations	·73	·09
			Mechanical Completion	·55	·26
			Woodwork	·51	·20
			Drawing	·40	·19

El Koussy obtained reports of introspections from many of his subjects concerning the mental processes which occurred during the working of the tests. These subjects reported that in the tests with high *k*-loadings visual imagery was employed in reaching the solutions whereas in tests not involving the *k* factor, success depended on a process of generalization and abstraction without the use of imagery. Thus, when performing the Pattern Perception Test one subject reported:

"I simply look at this shape on the left and when I look on the right I see it straightaway."

El Koussy concluded from these reports that "the explanation of the *k*-factor consists in the ability to obtain and the facility to utilize visual, spatial imagery."

At a much later date Emmett (1949) refactorized El Koussy's table of correlations using Thurstone's Centroid Method. Three factors were found to be significant, by applying McNemar's rough test of significance to the second residual correlations. Axes were then rotated orthogonally to eliminate negative loadings and the second factor was identified as the spatial factor. This factor showed loadings of ·4 or over, not only in the eight tests listed by El Koussy, but also in the following tests:

Table 3

Tests having spatial loadings of ·4 or more (Emmett's analysis of El Koussy's data, after rotation)

TESTS	2ND FACTOR LOADINGS
Visual Perception (Spearman)	·663
Form Equations C (El Koussy)	·612
Greys Analogy (El Koussy)	·606
Alphabet Series	·605
Mechanical Explanations (Cox)	·576
Classification (Spatial)	·556
Mechanical Completion (Cox)	·484
Woodwork Marks	·410
Form Equations B (El Koussy)	·407

Emmett's orthogonal rotation to eliminate negative loadings has resulted in some tests having high spatial loadings though on psychological grounds they might not be expected to have any spatial content at all. Thus, after rotation, the second factor loading of greys analogy is ·606 and of alphabet series ·605.

We might attempt to account for the *k*-loadings of these tests by supposing that *k*-factor is involved not only in the perception of relations between shapes but also in the perception of configurations or patterns of a more general kind (analogies in shades of grey, patterns in sequences of letters). Alternatively, we might invoke the explanation of imagery proposed by El Koussy. An ability to retain an impression of a shade of grey might conceivably assist in performing the greys analogy tests, while a similar ability to retain an image of a group of letters might be of value in doing the alphabet series test (since some of the items require the subject to continue a series of recurring letters).

If we consider the second factor loadings obtained by Emmett before rotation, we find that this bipolar factor contrasts the typical spatial tests (which in this case have high positive loadings) with a somewhat miscellaneous group of tests having high negative loadings. Arranging these loadings in order of magnitude and taking account of sign, they are:

Table 4

Spatial loadings of spatial and other tests
(Emmett's analysis of El Koussy's data, before rotation)

TESTS	2ND FACTOR LOADINGS
Correlate Eduction A (Spearman)	·333
Spatial Analogies	·317
Memory for Designs (N.I.I.P.)	·287
Classification (Spatial)	·233
Form Relations (N.I.I.P.)	·224
Greys Analogy	·188
Alphabet Series	·129
Inferences (Verbal)	— ·240
Pitch Discrimination (Seashore)	— ·256
Area Discrimination	— ·338
Loudness Discrimination	— ·397
Letter Cancellation	— ·413
Comparison of Greys (Speed, Spearman)	— ·444

When the second factor loadings are thus arranged in order, we note that the typical spatial tests, such as Correlate Eduction A (Upside-down Drawing), Spatial Analogies, and Memory for Designs (N.I.I.P), are differentiated in the sign of their loadings from a verbal test such as Inferences and a group of tests involving discrimination between pitches, loudness of sounds, or shades of greys. The fact that the second group includes a number of non-verbal tests suggests that the bipolar factor involves a broader differentiation than that between spatial and verbal abilities. A possible alternative hypothesis is that it represents a differentiation between an ability to perceive and retain 'in mind' spatial patterns (as in Memory for Designs) and an ability to switch attention from one item to another when perceived in temporal succession. In Comparison of Greys, which has the highest negative loading, the subject is presented with a series of circular areas of differing shades of grey. He has to compare each shade in turn with the previous one, crossing it out if it is darker. It is not difficult to make the required discrimination in any particular case, but since the test is speeded, subjects who can make many judgments in quick succession are likely to be most successful. Letter cancellation, which has the second highest negative loading, is rather similar. The difficulty does not lie in deciding which letters to cancel but in making as many cancellations as possible in the time allowed. In these two tests, high negative loadings are associated with rapid switching of attention. The pitch and loudness discrimination tests require the subject to compare or discriminate between sensations presented successively in time. Thus, a tentative interpretation of the bipolar factor may be attempted on the grounds that it differentiates between a group of tests involving the perception and retention of spatial patterns and a somewhat miscellaneous group requiring attention to stimuli perceived in temporal succession.

In the year following the publication of El Koussy's monograph, Clarke (1936) submitted a Ph.D. thesis to London University, reporting a very similar investigation carried out entirely with girls. The main experiment involved the administration of 29 tests to some 200 girls in the age-range 12 to 15. Of these tests, 7 were verbal and 17 spatial, two of them being the same as El Koussy's. Five were devised by Clarke expressly to test El Koussy's hypothesis that visual imagery was the explanation of the *k*-factor.

Two analyses were carried out, one using the Thurstone Centroid method and the other using a Spearman analysis in the light of information obtained from the Thurstone treatment. Her conclusion was that there was *no* extensive group factor in spatial tests, such as the *k*-factor found by El Koussy. She did find a group factor of small extent among the visual imagery tests, but this did not overlap into the spatial tests. She made the interesting observation that according to the analysis by the Thurstone method and perhaps also by the Spearman method, the verbal factor and the imagery factor were inversely related to one another.

One of her imagery tests (Designs) seems to have been similar in principle to the N.I.I.P. Memory for Designs Test, while another (Reversals and Inversions) was very like El Koussy's Correlate Eduction A (or Inverse Drawing). Both Memory for Designs and Correlate Eduction A had been found to have high *k*-loadings (·62 and ·58) so that it appears probable that Clarke's imagery tests were actually *k*-tests.

The following tests had extreme (positive or negative) second-factor loadings:

Table 5

Spatial loadings of tests used by Clarke

TESTS	2ND FACTOR LOADINGS
Imagery Tests	
Noughts and Crosses	·444
Designs	·378
Clockface	·350
Reversals and Inversions	·205
Ball	·195
Spatial Tests	
Line Pattern	·377
Form Relations (N.I.I.P.)	·272
Dot Series	·205
Incomplete Drawings (Street Gestalt Completion)	·198
Fitting Shapes	·196
Non-verbal Selection	·170
Verbal Tests	
Verbal Analogies (Selective)	— ·138
Selection	— ·154
Verbal Analogies (Inventive)	— ·231
Verbal Fours (Inventive)	— ·248
Best Answers	— ·310
Inferences	— ·332
Disarranged Sentences	— ·517

If Clarke's imagery tests were actually *k*-tests, her results did in fact support El Koussy's conclusions, although she claimed that they did not. Since her sample of pupils consisted entirely of girls, the factor would not be expected to show itself so prominently as in a sample of boys. Her conclusion that the verbal factor and the imagery factor were inversely related implies that the *v*- and *k*-factors are inversely related. This observation has been repeatedly confirmed. It does not mean that verbal and spatial tests necessarily correlate negatively. But it does mean that the correlation will be lower than one might expect to result from the presence of the common general factor. The verbal and spatial factors will tend to oppose rather than reinforce one another but the correlation will still be positive because of the general factor.

Though Clarke's results offered some support to El Koussy's findings, the general effect of her work in 1936 was to shed doubt on the existence of the *k*-factor. In 1937, the present writer submitted a thesis on the same topic to the University of Glasgow in part-fulfilment of the requirements of the Ed.B. degree, the title being "The Form-Perception Factor". When the writer embarked on this investigation in 1933, he knew nothing of the work of El Koussy. He had become interested in the possibility that a group factor might exist in tests of spatial ability. At that time the only technique that was generally known was the somewhat laborious method of tetrad-difference analysis originated by Spearman, though the first paper by Hotelling on the method of Principal Components was published in 1933. The writer investigated the factor problem by carrying out a tetrad-difference analysis and by applying Hotelling's method. But the main purpose of the investigation was to construct a spatial test of some hundred items and to make a preliminary study of its validity.

Having read an article by Stephenson (1931) describing an investigation into the existence of a group factor in non-verbal tests, he wrote requesting the loan of samples of test-material. Stephenson very kindly replied, enclosing a number of tests, some of which had been used by El Koussy, who was then a Ph.D.

student working under Stephenson's supervision. The writer proceeded to construct a spatial test consisting largely of items based on materials which had been obtained from this and other sources. He was also indebted to C. A. Oakley, of the Scottish Branch of the N.I.I.P., who supplied a N.I.I.P. spatial test. One sub-test of the new test (Form Recognition) was based on designs from an article by Gottschaldt (1926). (Many subsequent test-constructors, including Thurstone and Witkin, have based tests on these designs.) Another sub-test was modelled on the non-verbal items of Cattell's Test IIIA.

At an early stage in the investigation, the writer had noted that subjects showed very great individual differences in the ability to make recognizable drawings of simple objects. He had been particularly impressed by the fact that some pupils made drawings of objects such as Bunsen burners which were grossly 'out of proportion'. Thus, some pupils drew a burner with a base which was twice as wide as the height. It seemed painfully obvious that an inverse ratio of the dimensions would have been more nearly correct. Yet pupils who produced drawings like this seemed to see nothing amiss with them.

Thus, from the outset of the investigation, the writer had the theory that the special aptitude which he sought to measure, if it existed at all, would be manifested in an ability to perceive and reproduce shapes correctly, i.e. with their dimensions and their relations in due proportion. To test this theory, he included in the battery of tests a drawing test which required the pupils to make drawings of eight familiar objects of standard shape, such as Bunsen burners or milk bottles. These drawings were marked for the correct representation of proportions.

The sub-tests of the spatial test proper were constructed largely on this principle, i.e. that the items should depend *critically* for success on the perception of the correct proportions of a figure or pattern. This has been the writer's guiding principle in his subsequent researches. He has been puzzled by the fact that other research workers in this field seem to be unaware of it, or at least do not refer to it. It was not mentioned in the symposium on hypotheses concerning the nature of the spatial factors, held at the A.P.A. Congress in Washington in 1952.

Altogether, nine spatial tests were constructed and these were duplicated and administered to first and second year pupils in a Scottish grammar school in June, 1934. The ages of the pupils ranged from 12½ to 14½, the average age being about 13. The experiment was carried out, the correlations calculated, a tetrad-analysis completed and the thesis written, before the publication of El Koussy's thesis. When El Koussy's monograph became available late in 1935, it was clear that the findings were substantially the same.

The following tests were used in the investigation:

1. *Area Discrimination.* Three series of forms, having the same shape, but differing in size (circles, squares, triangles), the pupil being required to number the items in order of magnitude.

   This test was a development of part of a similar N.I.I.P. test consisting of a series of circles. It is not the same as the test of that name used by El Koussy.
2. *Completion.* Two pages of the N.I.I.P. Form Relations Test, each page having a series of squares, containing gaps which differed in shape.

   The subject had to select from a number of given shapes, the correct ones to fill the gaps.
3. *Fitting Shapes (A and B).* Similar in principle to tests used by Stephenson and El Koussy. Groups of shapes are shown which, when properly orientated and fitted together, form a larger shape. Lines have to be drawn on the larger shape to indicate how the parts are fitted together.
4. *Form Equations.* Similar to El Koussy's Form Equations B. Each item consists of an 'equation' with spatial terms, the signs being omitted. The pupils were required to insert the signs (+) or (—) required to make the 'equation' true.
5. *Classification.* 'Odd man out'. The pupil is required to indicate which of the five does not belong to the same group as the others.