The construction and validation of a spatail test, using diagrammatic material based on projections and sections of solid objects

J. S. Lawes

Abstract of:

The Construction and Validation of a Spatial Test, using Diagrammatic Material Based on Projections and Sections of Solid Objects.

M.Ed. Thesis submitted October 1960

by J.S. Lawes

An important development in mental testing has been the construction of tests for measuring the ability to obtain, and utilise, visual spatial imagery. These spatial tests have proved valuable in predicting subsequent success in such spheres as engineering apprenticeships, technical drawing and woodwork. They have also been employed successfully in selection for secondary schools.

Although a limited number of such tests already existed, it was felt that a new test based on three-dimensional material, and particularly projections and sections of this material, could prove of value.

Two hundred possible test items, of thirteen types, were therefore prepared and "tried-out" on a representative sample of school children. An item analysis of the resulting data provided indices of Facility and Discrimination which were used to select the one hundred most suitable items.

The selected items were then organised into a revised

draft which was inserted into the selection examination for the entire ten-year-old population of a city. This large-scale trial, as well as showing that the test and instructions are suitable for the age group, provided information from which tables for converting raw scores to standardised scores were constructed; enabled a second, confirmatory item analysis to be made; showing a significant difference in mean scores for boys and girls, a recognised property of spatial tests; and provided the following figures:

Range:	1 - 99
Mean:	42.1
Standard Deviation:	20.745
Reliability:	0.9642

In an investigation with 85 boys the test correlated more highly with a recognised spatial test than with a verbal test, and was a better predictor of success in metalwork than was the verbal test.

Extended validation, notably a four year follow-up of the 85 boys, and a factor-analysis are required before it can be certain that the test is truly "spatial" with the uses associated with such tests.

THE CONSTRUCTION AND VALIDATION OF A SPATIAL TEST, USING DIAGRAMMATIC MATERIAL BASED ON PROJECTIONS AND SECTIONS OF SOLID OBJECTS.

Thesis submitted for the degree of
Master of Education

by

James Sidney Lawes.

October 1960.

(University College)
Durham.

CONTENTS

CHAPTER 1.

BACKGROUND TO THE TEST.

Historical.

Mental testing originated in the latter half of the nineteenth century, when Sir Francis Galton devoted some time to devising tests aimed at classifying men according to their natural gifts, and to developing methods of dealing with the results. Galton's[1] interests ranged widely over many fields of psychology, anthropology, sociology, eugenics and much else, so that although the beginnings of mental measurement date from him and from the contemporaneous work of the American McKeen Cattell, it is the name of Binet which stands high in the history of testing.

Binet, with Simon, devised a series of tests aimed at measuring intelligence. These were first published, in Paris, in 1905, revised in 1908 and 1911 and subsequently were adapted for use in a number of countries. This work stimulated the construction of tests of many types by a whole series of workers. The last half century has seen the development of mental testing into one of the major

---

1. See, for example. Hereditary Genius. 1869.
   Inquiries into Human Faculty. 1883.

branches of psychology. Tests have been developed for individual testing and for the testing of groups; for testing intelligence, achievement, special abilities and many other capacities and characteristics.

With such a wealth of tests came the need for better methods of dealing with the results. As the art and science of testing grew so did the statistical techniques. In the later form of his tests Binet himself resorted to statistical methods, recording performance in them in terms of mental age. He established "norms" for his tests, i.e. the point on the test scale which the average child of any given chronological age would be expected to reach.

All reputable tests are now provided with information of this sort, that is they are standardised.

Development of Factor Analysis.

Apart from techniques designed to improve the usefulness of any particular test, there has been the development of a whole field of mathematics to aid the comparison of measurements obtained from a number of tests. This analysis of data has for its basic tool the correlation coefficient, a number expressing the degree of relationship between sets of scores.

If the scores obtained in one test are in the same relative positions and of the same relative sizes as the scores obtained by the same individuals in a second test then the correlation between these two sets of data is perfect and designated numerically 1. If the second results are exactly the reverse of the first set then the correlation coefficient (r) becomes -1. A correlation which is neither perfect nor inversely perfect is represented by a decimal fraction between 1 and -1. A variety of methods is available for the calculation of 'r'.

C. E. Spearman was one of the first to see the value of applying correlation methods to mental test results and the development of analytic techniques owes much to him. He devised many of the mathematical methods and himself used them in attempts to explain the structure of the mind. Using techniques such as his "tetrad difference" method on correlation coefficients, Spearman obtained results from which he formulated the theory that the correlations can be accounted for by a postulated general factor of ability 'g', which operates in the performance of all the tests, and specific factors which are brought into use only in particular single tests. This Two-Factor theory, first formulated by Spearman in 1904 and published in detail in 1927[1], was in opposition to the theories of Thorndike and other

1. Spearman, C. (1927). The Abilities of Man.

American psychologists who, also at the beginning of the century, advocated not a common factor plus specifics but an unlimited series of specific factors. These two antagonistic views have in time become modified in the light of further work.

One of the main lines of investigation in Britain has been in the establishment of further factors which operate over a series of tests but are not common to all tests. That is, for a number of varying tests, there is correlation of all in as far as they measure 'g', and for some of them there will be further overlap over and above that accounted for in terms of 'g'. While Spearman preferred to ascribe this to "specific overlap" and insisted that its occurrence was rare, it is now usual to take such correlation as evidence of a "group factor."

Emergence of Group Factors.

In one of the early analyses Burt (1917)[1] published evidence of verbal, numerical and practical group factors in school subjects, in addition to a general factor. Cox J. W.[2] in 1928 showed the presence of a factor common to various mechanical ability tests. Cox labelled this factor

1. Burt, C. (1917). The Distribution and Relations of Educational Abilities.

2. Cox, J.W. (1928). Mechanical Aptitude.

'm' and states, "The outstanding feature in which the 'm' tests differ from the customary 'intelligence' tests, ..., is the spacial (sic) character of the material employed." The presence of 'm' was shown by Spearman's 'tetrad difference' technique which was also used by Stephenson, (1931)[1], in an investigation of verbal and non-verbal tests which showed for the verbal tests a factor 'v' in addition to 'g', but for the non-verbal tests nothing but 'g'.

The existence of 'v' was corroborated by Alexander (1935)[2] with, in addition, a group factor present in performance tests but not in verbal tests. This factor was called 'F' and described as a practical factor. A further group factor 'X', running through measures of school attainment and identified as "persistence or determination" is also worthy of note. These factors were obtained by the use of a "multiple factor" technique and confirmed by "tetrad" analysis. Koussy (1935)[3] in work published

---

1. Stephenson, W. (1931). Tetrad Differences for Non-Verbal Sub-tests. Tetrad Differences for Verbal Sub-tests. Tetrad Differences for Verbal Sub-tests relative to Non-Verbal Sub-tests. J. of Educ. Psych. XXII.

2. Alexander, W.P. (1935). Intelligence, Concrete and Abstract. Brit. J. of Psych. Monog. Supp. XIX.

3. El Koussy, A.A.H. (1935). The Visual Perception of Space. Brit. J. of Psych. Monog. Supp. XX.

during the same year desdribes a tetrad analysis of 28 tests given to 162 boys between the ages of 11 and 13. Of the 15 'spatial' tests included 8 were shown to have loadings on a factor over and above their 'g' content. This 'k' factor is explained by Koussy as "the ability to obtain and the facility to utilise visual spatial imagery."

This brief summary serves to show how the Two-factor theory has come to be modified by the discovery of group factors. Spearman[1] himself in time came to recognise the usefulness of the group factor idea particularly in so far as it concerns 'v' and 'k' and particularly favoured the 'Bi-factor' extension of the theory by Holzinger and Swineford.

Meanwhile there had been parallel developments in America, affecting the Thorndike theories. In 1928 a pattern of verbal, number, spatial and speed factors, similar to that of Burt, was established by Kelley[2] for batteries of tests which he gave to three groups of children of different age levels. Since 1930 L. L. Thurstone has been the leader of American factor analysis, both in the development of methods and in the interpretation of results.

1. Spearman, C., & Wynn Jones, Ll. (1950). Human Ability.
2. Kelley, T. L. (1928). Crossroads in the Mind of Man.

Using his 'Centroid' technique he has carried out a long series of investigations and extracted what are known as Multiple Factors, far less in number than the specific factors of Thorndike and of more equal variance than the general plus specific factors of Spearman. They do not in fact include a factor general to all tests. Thurstone's first account in 1938[1] describes nine factors to which he attaches names, among them being

S - Visual or Spatial. P - Perceptual.

N - Numerical. V - Verbal

and M - Memory.

Subsequent work has confirmed these factors at different age levels and, particularly among children, the factors can be shown to be not entirely independent but to reveal what Thurstone has called a second-order general factor.

We can thus see that there has been some 'rapprochement' between the two opposing theories. These theories depend to a considerable extent on the methods used to obtain the factors and the fact that different methods show similar factors, albeit under different names, is significant. It matters little whether the analysis be of Group

---

1. Thurstone, L.L., (1938). Primary Mental Abilities.

Factor or of Multiple Factor type if the same factors still emerge. Particularly noticeable is that both Group Factor methods and Multiple Factor methods show the existence of verbal and spatial factors in particular types of tests.

Spatial Factors.

Attention will now be turned to the spatial factor. The relationship between it and various practical factors will be examined and evidence concerning the appearance of the spatial factor and its practical uses will be summarised.

That El Koussy's 'k' and Thurstone's 'S' are very similar, if not identical, is generally recognised; to quote but one authority Vernon (1950)[1] states "The Thurstones ...... obtained a factor which they call S, obviously the same as El Koussy's 'k'." The evidence about the relationship between 'k' and the earlier 'M' and 'F' factors is not clear cut but the suggestion of Slater (1940)[2] and Kerr (1942)[3] is that Spatial and Mechanical Tests measure similar abilities, while Price (1940)[4] and Dempster (1948)[5]

1. Vernon, P.E. (1950). The Structure of Human Abilities.
2. Slater, P. (1940). Some Group Tests of Spatial Judgment or Practical Ability. Occup. Psych. XIV.
3. Kerr, G. (1942). Aptitude Testing for Secondary Courses. Occup. Psych. XVI.
4. Price, E.J.J. (1940). The Nature of the Practical Factor (F). Brit. J. of Psych. XXX.
5. Dempster, J.B.B. (1948). The Selectors Point of View. Brit. J. Ed. Psych. XVIII.

conclude that spatial and performance tests measure the same ability, at least in some cases. Vernon in his hierarchical theory[1], in its simplest form, proposes a major group factor k:m covering spatial, performance and mechanical tests for most practical purposes. He suggests that "F is identical with the spatial factor 'k', and that Cox's 'm' is largely composed of 'k' also".[2] Watts and Slater (1950)[3] suggest that there is a range of 'non-verbal' abilities from those that are expressions of general mental ability, through problems involving spatial relationships without actual manipulation, up to those that require manipulation of concrete materials.

There is ample evidence to confirm El Koussy's conclusion that a spatial factor operates in certain types of test situation. Although Slater in 1941[4], and again with Bennett in 1943[5], could find no evidence of a spatial factor when working with children aged from 10 - 13, re-analyses

---

1. Vernon P.E. and Parry, J.B. (1949). Personnel Selection in the British Forces.

2. Vernon, P.E. (1950). Op. cit.

3. Watts, A.F. and Slater, P. (1950). The Allocation of Primary School Leavers to Courses of Secondary Education. 1st Interim Report, N.F.E.R.

4. Slater, P. (1941). Tests for Selecting Secondary and Technical Schoolchildren. Occup.Psych.XV.

5. Slater, P. and Bennett, E. (1943). The Development of Spatial Judgment and its Relation to Some Educational Problems. Occup.Psych. XVII.

of the data by both Adcock (1948)[1] and Emmett (1949)[2] does reveal a definite 'k' factor. This re-analysis by Emmett was part of a work in which he examined a number of previous investigations. As well as showing a 'k' factor in Slater's data he confirmed the results of Drew (1947)[3] and showed that they could be interpreted as revealing a 'k' factor at 11. Emmett also reports high 'k' loadings in seventeen of El Koussy's tests as compared with eight in the original work. El Koussy himself in the report of a lecture to the XII th International Congress of Psychology in 1948[4] states "The existence of the k-factor from 12 upwards with certain paper-and-pencil tests is repeatedly corroborated."

Macfarlane Smith (1948)[5] in work carried out in 1934 with Scottish pupils aged 12½ to 14½ obtained results which confirm those of El Koussy and provides the further information that there is a significant difference between the

---

1. Adcock, C. (1948). A Re-Analysis of Slater's Spatial Judgment Research. Occup. Psych. XXII.
2. Emmett, W.G. (1949). Evidence of a Space Factor at 11+ and Earlier.
3. Drew, L.J. (1947). An Investigation into the Measurement of Technical Ability. Occup. Psych. XXI.
4. El Koussy, A.A.H. (1948). A Further Examination of the k-factor. XII th. International Congress of Psychology.
5. Smith, I. Macfarlane. (1948). Measuring Spatial Ability in School Pupils. Occup. Psych. XXII.

mean scores obtained by boys and girls. This sex difference is confirmed by the same writer in 1954[1] in an account of a Test which measures the spatial factor at age 11, and also by Emmett in the work cited above. A significant sex difference is now taken to be an indication that a particular test is measuring spatial ability.

Evidence of the age at which spatial ability can be measured is equally abundant. El Koussy's[2] original research was with boys aged 11 - 13, Macfarlane Smith's[3] with pupils of 12½ to 14½, Adcock's[4] re-analysis of Slater's data refers to an age range of 11 - 13. Emmett's (1949)[5] work was specifically designed to show a Space Factor at 11+ and reverses Mills (1947)[6] conclusion that Moray House Space Test 1 fails to measure the space factor at 11+. Dempster (1948)[7] and Peel (1949)[8] both state that if the ability can

1. Smith I. Macfarlane. (1954). The Development of a Spatial Test. Durham Research Review 5.

2. El Koussy, A.A.H. (1935). op. cit.

3. Smith, I. Macfarlane. (1948). op.cit.

4. Adcock, C. (1948). op.cit.

5. Emmett, W. G. (1949). op.cit.

6. Mills, L. F. (1947). An Investigation into the Properties of a Space Test. B.Ed. Thesis. Edinburgh.

7. Dempster, J. J. B. (1948). op.cit.

8. Peel, E. A. (1949). Evidence of a Practical Factor at the Age of Eleven. Brit. J. Ed. Psych. XIX.

be measured at 13 then it is equally evident at eleven. Watts and Slater (1950)[1] state cautiously that there is some hope that variations in F and k can be measured at 11 plus. Renshaw (1950)[2] concludes that his space tests showed clear evidence of a space factor at 11 and Macfarlane Smith (1954)[3] has shown that his Spatial Test 1 definitely measures the spatial factor in association with g at age 11.

Practical Applications of Spatial Tests.

It would therefore seem that in the age range from 11 to 14 years it is possible, using appropriate tests and measures, to obtain some measure of the 'k' factor. This is important as one purpose behind the construction of a spatial test is that it should prove of use in the allocation of children to different broad types of education and in addition, or alternatively, enable some forecast to be made of a child's potentialities in a particular subject or group of subjects.

Allocation to types of education has at present its greatest meaning with regard to the transfer from primary to secondary schools. Figures quoted by Yates and Pidgeon (1957)[4] show that the majority of Education Authorities in

1. Watts, A.F. and Slater, P. (1950). op.cit.
2. Renshaw, T. (1950). Factorial Study of Two and Three Dimensional Space Tests. Ph.D.Thesis.Edinburgh.
3. Smith, I. Macfarlane. (1954). op.cit.
4. Yates, A. and Pidgeon, D. A. (1957). Admission to Grammar Schools. (N.F.E.R. Publication. No. 10).

England and Wales administer selection tests for this transfer each year to those children who will "have reached the age of eleven but not the age of twelve" at the end of August or the beginning of September (actual dates range from 31st August to 2nd September). Since these tests, for administrative and other reasons, are given to the children concerned in February or March, the age at which the test will actually be taken varies from approximately 10½ to 11½ years, with possibly a lower limit in those areas where 'under-age' candidates are permitted. In the investigations reported by Yates and Pidgeon the subjects were within this approximate range, being within the limits 10.7 to 11.5 on 1st March.

These investigations were planned to correlate scores obtained by pupils in various tests and batteries of tests used in allocation to secondary schools with their subsequent performance in the second and third years of the secondary school courses. One conclusion of this research was that a battery consisting of: Primary Heads Assessment weighted 4, plus standardised attainment test in English weighted 2, plus a Spatial Test, yielded a higher prediction than any other battery. The spatial test included is Spatial Test 1 of the National Foundation series, designed by Macfarlane Smith and described by him in the work already mentioned.[1]

The comment of Yates and Pidgeon on the appearance of a spatial test in the battery is of particular note: "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 desirable for further research to be undertaken to investigate this point."

The relation between various school subjects and spatial tests has been the subject of a number of investigations including the following, referred to in order of publication. Alexander (1935)[2] in his monograph concludes that for mechanical drawing the 'X' factor is most important, followed by 'g' and then 'F' and 'v' in about equal amounts, but that for "shopwork" the order of importance of the factors is 'X', 'F' and 'g'. Blackwell, A. M. (1940)[3] in an investigation into mathematical ability obtained a factor 'O' which was of next greatest importance after 'g'. This is an "operation in imagery factor, involving the manipulation of spatial and verbal data." This factor was found

---

1. Smith, I. Macfarlane. (1954). The Development of a Spatial Test. Durham Research Review. 5.

2. Alexander, W. P. (1935). Op. cit.

3. Blackwell, A. M. (1940). A Comprehensive Investigation into the Factors Involved in Mathematical Ability of Boys and Girls. Brit. J. Ed. Psych. X.

to play a relatively larger part in the mathematical ability of boys than girls, a pointer that 'k' may, at least in part, be involved.

Correlations of Verbal Intelligence scores and scores on a battery of Spatial and Mechanical tests with Engineering Drawing and Engineering Mathematics marks obtained by boy apprentices are reported by Holliday, F. (1940)[1]. For Engineering Drawing the correlation with Spatial/Mechanical tests is almost ten times greater than the Verbal Intelligence correlation; for mathematics the correlation with the verbal test is about twice that with the spatial/mechanical tests. Extended "follow-up" investigations by the same writer[2,3] show significant correlations between the Engineering Drawing and the spatial and mechanical tests, up to 2 years 7 months after the tests were administered.

In an inquiry into tests of technical aptitude, Shuttleworth (1942)[4] found that the three tests which correlated

---

1. Holliday, F. (1940). An Investigation into the Selection of Apprentices for the Engineering Industry. Occup. Psych. XIV.

2. Holliday, F. (1941). A Further Investigation into the Selection of Apprentices for the Engineering Industry. Occup. Psych. XV.

3. Holliday, F. (1943). The Relation between Psychological Test Scores and Subsequent Proficiency of Apprentices in the Engineering Industry. Occup. Psych. XVII.

4. Shuttleworth, C. W. (1942). Tests of Technical Aptitude. Occup. Psych. XVI.

most highly with performance ratings on the complete course of subjects in a Junior Technical School were Space Perception, Designs (reproduction) and Form Relations.

Holzinger and Swineford (1946)[1] concluded that "correlations ...... indicate that the general factor is a much better predictor of geometry than is the orthogonal space factor. For shop work and drawing, the predictive value of these two factors appears to be reversed." On the other hand both Barakat (1951)[2] and Wrigley (1958)[3], in factorial studies of mathematical abilities, provide evidence of a definite connection between geometry and the spatial factor.

Finally three investigations of Macfarlane Smith will be quoted. In 1948[4] he reports significant correlations between spatial scores and ~~the~~ Art, Practical Geometry, Engineering Drawing and a Drawing Test; correlations with Handwork and Algebra were not significant. The 1954[5] report

1. Holzinger, K. J. and Swineford, F. (1946). The Relation of Two Bi-factors to Achievement in Geometry and Other Subjects. J. of Educ. Psych. XXXVII.
2. Barakat, M. K. (1951). A Factorial Study of Mathematical Abilities. Brit. J. of Psych. Stat. Sect. IV.
3. Wrigley, J. (1958). The Factorial Nature of Ability in Elementary Mathematics. Brit. J. Ed. Psych. XXVIII.
4. Smith, I. Macfarlane. (1948). op. cit.
5. Smith, I. Macfarlane. (1954). op. cit.

shows that for most technical subjects Spatial Test 1 (N.F.E.R.) is a more valid measure of ability than is the Kent Senior Intelligence Test; while in 1959[1] marks obtained in school examinations in Engineering Drawing, metalwork and woodwork are stated to correlate highly with Spatial Test 1 with an interval of three years between test and examinations.

Such is the background to the present work which is a report on the construction of a spatial test suitable for pupils in the age range 10½ to 11½ years, the standardisation of the test on an appropriate population and the correlation of this test with marks in certain school subjects.

---

1. Smith, I. Macfarlane. (1959). Selection for Technical Education. Journal of the Institute of Education of Durham University. Vol. 10. No. 51.

CHAPTER II.

CONSTRUCTION OF THE TEST.- PRINCIPLES AND PRACTICE.

In the construction of a test of this nature certain principles of test construction should be observed. These principles are designed to improve the test in three important attributes, Objectivity, Reliability and Validity.

Objectivity. A test is objective (a) if the items are set so that each person answering a particular item is faced with the same situation and understands what is expected of him; (b) if the possible answers are so designed that there is only one correct answer; and (c) if the checking and scoring of the answers is definite and rigid so that each and every marker will give the same score; the personal, subjective opinion of the marker having no influence.

Reliability is the degree of accuracy with which a test measures. A test has high reliability if on re-test a group of subjects obtains practically the same scores as in the original testing.

Validity. A test is valid only in so far as it measures that which it purports to measure.

These three characteristics of any test or examination are not separate, independent entities; reliability is greatly influenced by objectivity and validity is limited by reliability. All three are greatly improved by careful planning, careful design of items, the try out of these items on a population representative of that for which the test is intended, and statistical analysis of the items to select those most suitable for the final form of the test.

In the preliminary planning the purpose of the test and the nature of the material should be carefully delimited. In the present work the purpose is implicit in the title words "and Validation of a Spatial Test"; that is the test is to be spatial, an instrument for measuring the 'K'-factor, and is to be valid for the purpose proposed in Chapter I, the allocation of pupils to school courses. This means that throughout the design of test items there is present the idea that those items must have the seemingly twofold purpose of testing "the ability to obtain and utilise visual spatial imagery" and also be of predictive value towards certain school subjects, in particular mathematical, scientific and technical subjects.

As for the nature of the material this is stated in definite terms as "diagrammatic material based on projections and sections of solid objects." Three dimensional material has been chosen for the test since Emmett[1] (1949) has shown that higher k-loadings are obtained in items with three-

1. Emmett, W. G. (1949). Op.cit.

dimensional objects than with two dimensional objects, although the more recent work of Yates and Pidgeon[1] shows that a Spatial Test based mainly on two-dimensional material correlated more highly with their criteria of success than did a Spatial Test based on three dimensional material. The work of Renshaw (1950)[2] also showed that the space factor is involved in both two- and three-dimensional tests, with a slight indication that it is the two-dimensional tests which are more closely linked with the factor.

In theory the number of items prepared should be at least twice the number which will be included in the final form of the test and these items should be varied in type, cover a range of difficulty, and have simple systems of response and of marking. A total of 200 items were in fact prepared with these points in mind and are described more fully in Chapter III.

Far more items are prepared than will ultimately be required so that a statistical analysis may be carried out to select items which are of optimum value for the intended purpose of the test. This analysis, which greatly improves both Reliability and Validity, is carried out on the results obtained from a try-out of the test on a population similar

---

1. Yates, A. and Pidgeon, D. A. (1957). Op.cit.

2. Renshaw, T. (1950). Op.cit.

to that for which the test is designed. This sample should number between 100 and 200 and be fully representative. The sample chosen in this instance was from Grammar and Secondary modern pupils in their first year at those schools and contained both boys and girls. Details of the try-out are found in Chapter IV.

Item Analysis is the procedure by which those items are selected which contribute with greatest efficiency to the success of the test as a whole. Items which are answered correctly by all candidates or those answered correctly by no candidate contribute nothing to the predictive value of a test. Therefore for all items a Facility Index, that is the percentage of candidates answering the item correctly, is calculated and those with very high or very low facility are rejected.

It is not enough that items should be within a given range of Facility, it is also necessary that each item chosen should be capable of discriminating between those who possess the measured ability to a high degree and those who possess it to a lower level. Facility and discrimination are different concepts; the fact that 50% of testees answer an item correctly does not necessarily mean that they all possess the ability to be measured to the same degree. Discriminating power of an item can be defined as the degree to which success or failure in that item by itself indicates success or failure in the test as a whole. Defined in this way the Discrimination Index of an item can be calculated by using total score in the test as the criterion. This is the method which has been used on the items described, details of the analysis being presented in Chapter V.

The items selected in this way form the revised draft of the test for standardisation and validation. This revised draft is discussed in Chapter VI, while the standardisation on a complete year group of pupils in a city, is described in Chapter VII together with a second item analysis and the calculation of reliability coefficient.

Validation, within the terms discussed at the beginning of the chapter, involves showing that the test (a) measures spatial ability (b) has predictive value. These two criteria of validity are not of necessity separate. Earlier the phase "seemingly twofold purpose" has been used about the test because while it is possible that success in the test could correlate highly with marks in school subjects for the sole reason that some test items were designed with the purpose in mind, it seems most likely from previous researches[1] that success in these subjects does in fact depend, at least in part, on spatial ability. The criteria could be therefore, not entirely separate but overlap to some extent.

1. See Chapter I Section on "Practical Applications."

To obtain a measure of the extent to which the test does in fact measure spatial ability a factorial analysis of the test to provide K-loadings of the various sub-tests would be required. This is a time-consuming process and is not included in the present work, but in Chapter VIII correlations between the test and recognised Spatial and Verbal tests are reported together with correlations with the second criterion, success in school subjects.

CHAPTER III.

THE FIRST TRY OUT DRAFT.

For the first draft of the test two hundred items were devised. These items are of thirteen types, all of them original designs although, as will be shown later, some were inspired by items in previously published tests, but with considerable modification to three-dimensional material. In all items an attempt has been made to present a problem in three dimensions; in the majority this is achieved by using oblique views of solid blocks. In some sub-tests these blocks have been shaded to give the impression of solidity, in others simple line drawings are used, there being no evidence available as to which type of diagram achieves the best results.

If the possession of the k-factor indicates "ability to obtain and facility to utilise visual spatial imagery", then test items should require for their solution both the "ability" and the "facility to utilise". The spatial image to be obtained in the items described here, is that of the solid object and the utilisation or manipulation of this image is included in the solving of the problem.

In many cases the types of problem posed deal with projections and sections of the solids, and means have had to be devised for presenting such problems without introducing terms such as "projection","section" and "profile" which would place a premium on the testee's knowledge. It is axiomatic in test construction that instructions should be understandable by all, that the weakest pupil should understand what he is expected to do even if he has not the ability to make the correct response. Ten of the thirteen sub-tests in the first draft are preceded by detailed instructions and trial items to ensure fuller understanding. These instructions were based on informal trials with children of the appropriate age range.

The sub-tests were designed in a number of forms including Multiple Choice, Matching, Equations and True/False, with a variety also of methods of response. In all methods of response the overriding consideration has been that of simplicity, in Multiple Choice the placing of a cross (X) on the chosen answer, in Matching the writing of the appropriate letter under a diagram, for example. In two sub-tests, F and M, the response is of a more complicated nature, consisting, in the former, of putting circles round E's in a pattern, and in the latter of drawing lines on a given shape.

In all items, except those of the last sub-test M, the marking is entirely objective, there being but one correct answer, all others being wrong. In sub-test M this is not so and during the marking of it a list has been composed of admissible responses.

Some thought was given to the mechanical aspect of marking. At first it was intended that all answers should appear at the right hand edge of a page to facilitate the use of a marking list, but on second thoughts this was abandoned in favour of having the response placed directly on, or under, the appropriate diagrams. This was done in the interests of the testees since it simplified the instructions and also eliminated any error which might occur in transferring mentally the response decided on from its position in the problem to the end of the line.

The 200 items of the first draft were divided into two tests of 100 items each for the First Try-out, and presented in duplicated form. The composition of the two tests was as follows:-

<u>Test I</u> consisting of 6 sub-tests.

A.	Counting Faces.	20 items.
B.	Plan Views.	10 items.
C.	Profiles.	30 items.
D.	Block Building.	10 items.
E.	Overlying Patterns.	10 items.
F.	3-D Networks.	20 items.
	Total	100 items.

Test II consisting of 7 sub-tests.

G.	Front and Side Elevations	10 items.
H.	Form Boards.	10 items.
I.	Three Sections.	10 items.
J.	Nails in Blocks.	20 items.
K.	Corner Recognition.	10 items.
L.	Vertical Sections.	20 items.
M.	Paper Folding.	20 items.
	Total	100 items.

Each of the sub-tests is described and discussed in the following pages.

A. Counting Faces. Fig. 1.

In this test an oblique view of a solid is shown together with plan views of each type of face found on the solid. The pupil is required to mark on each plan a number indicating how many times that particular face appears on the solid, not forgetting the base and those faces which are not seen. Scoring is straightforward, one point for each face correctly numbered. In all, there are 8 solids with 20 faces referred to them.

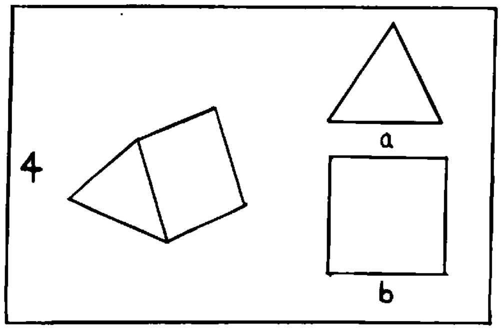


Figure 1.

B. Plan Views. Fig. 2.

Each question shows a solid model formed by placing various shaped blocks together. A cross (X) has to be placed on the one of four diagrams which shows the view looking down on the model. There are ten items in all.

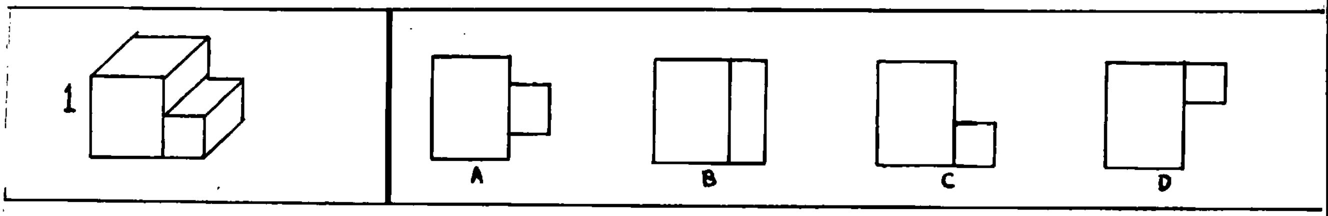


Figure 2.

C. Profiles. Fig. 3.

The items in this sub-test refer to four blocks A, B, C and D shown at the top of the page. The blocks are A a cube, B a flat square equal to half the cube, C a double cube and D a double flat square. Each question shows the front elevation of two or more of these blocks placed together. An X has to be placed on the correct one of three diagrams showing the profile from the right hand side. In all there are 30 items, the first ten showing the blocks in the same positions as illustrated at the start of the test, the second ten with some of the blocks turned round and the third ten with some of the blocks used more than once.

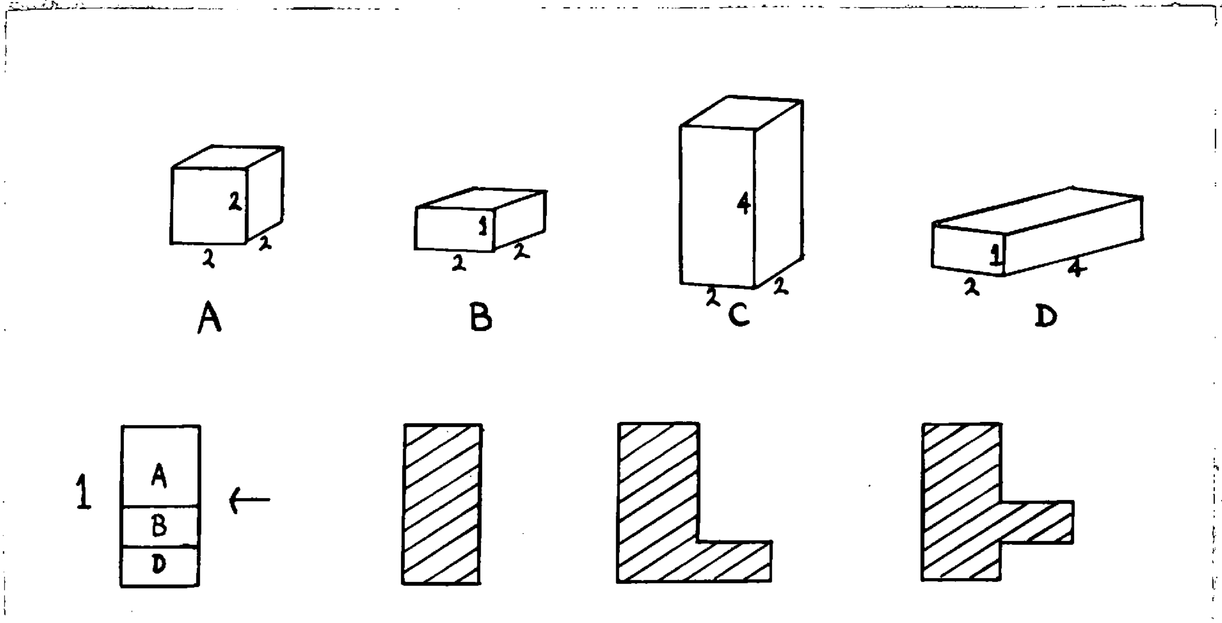


Figure 3.

D. Block Building. Figure 4.

Inspired jointly by the Fitting Shapes of Macfarlane Smith and the Square Completion of Watts, both of which use two-dimensional material, the items of this sub-test consist of drawings of a larger cube or rectangular block plus four smaller blocks. Three of these small blocks placed together form the large block. A cross (X) is placed on the one small block not required.

in the study of Chemistry is in the understanding of structural formulae. There is also about it an element of Pattern Recognition translated into three dimensions.

A regular pattern of 25 crosses is presented together with a diagram of a solid block the visible corners of which correspond to a number of the crosses. Circles have to be placed round these crosses. Although in effect calling for the representation of a three-dimensional drawing this type of task does not suffer from the usual difficulties encountered in marking drawings since a correct solution is only obtained by circling the appropriate crosses. The marking is entirely objective.

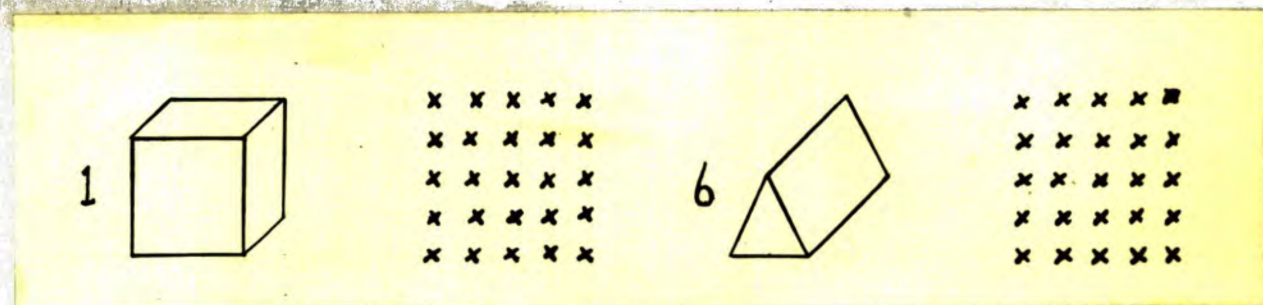


Figure 6.

G. Front and Side Elevation.

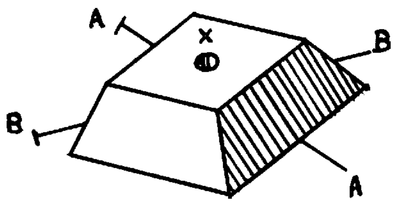
H. Form Board.

I. Three Sections.

These three tests all employ Matching methods. In G ten solid blocks have to be matched with pairs of drawings showing front elevation and side view. H is an attempt at a Form Board on paper with shapes representing holes to be filled by given objects. I is another attempt at presenting the problem described in E, above, with objects being matched with sets of three sections.

J. Nails in Blocks. Figure 7.

The example given below (Fig. 7) is more or less self explanatory. Instructions include the information that hole X is vertical and Yes or No is to be underlined.



7 Does nail A—A pass across hole X ? YES NO

8 Does nail B—B pass across hole X ? YES NO

Figure 7.

K. Corner Recognition. Figure 8.

A three-dimensional variant of Form Recognition, this test provides three-line drawings of corners copied from blocks illustrated at the head of the page. The letter of the appropriate block has to be placed under each diagram.

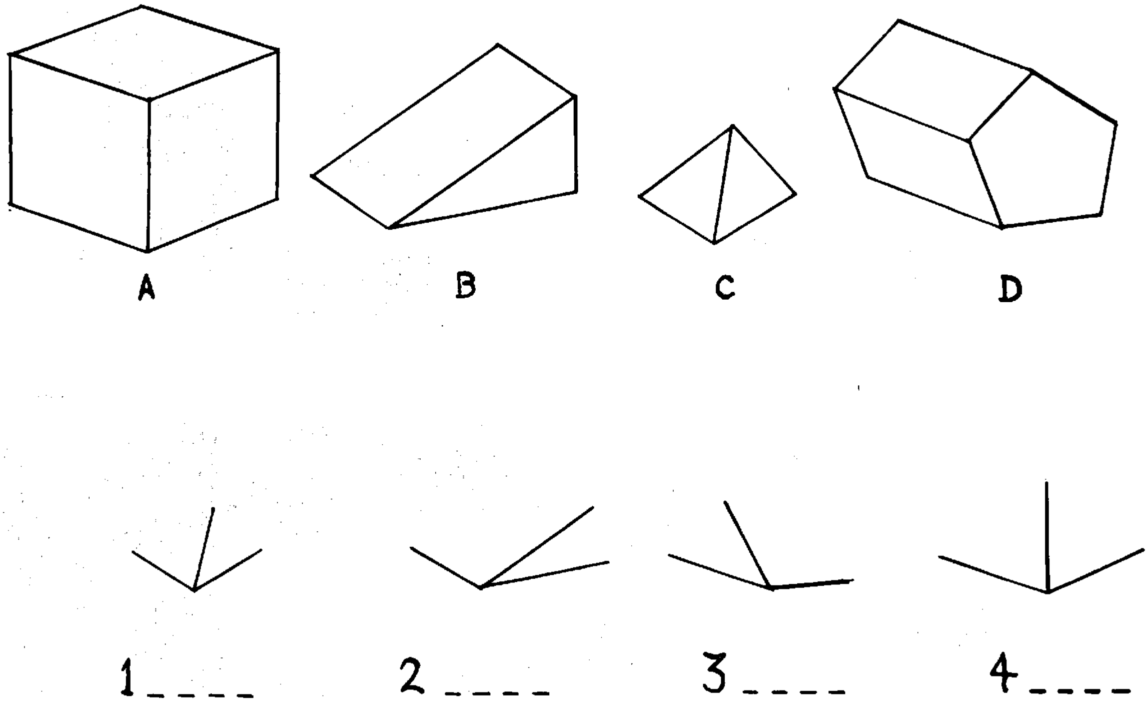


Figure 8.

L. Vertical Section. Figure 9.

This is an attempt at posing straightforward problems involving sections. The position of a vertical section is indicated on a block and there is a choice of three (in later items, four) alternatives, on one of which a cross is to be placed.

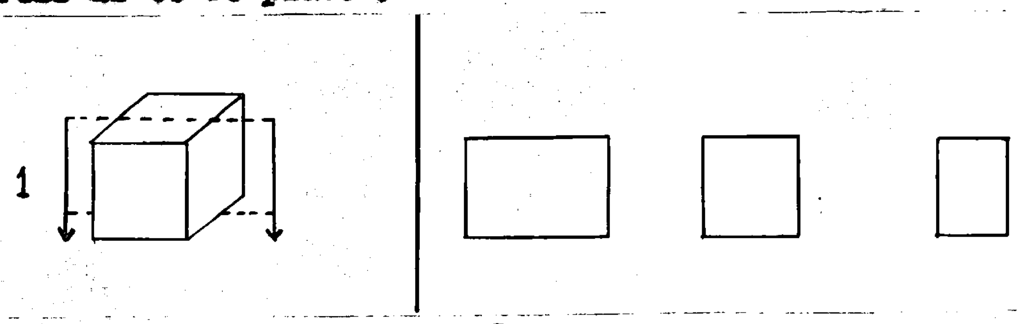


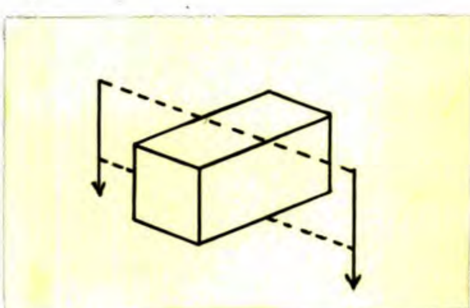
Figure 9.

viewing the material have any significant effect on the scores.

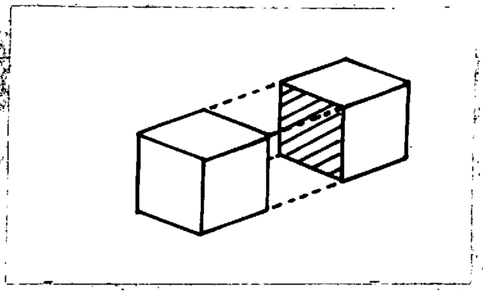
These then are the thirteen types of items which comprise the try out draft. Sub-tests H, J and K were not preceded by practice tests, it being felt that the instructions given at the head of the test paper were in each case sufficient for all pupils to understand. This belief was borne out by the facility figures obtained from the try-out. In all other sub-tests a practice test was worked immediately prior to each test. These practice pages included more detailed instructions than those given on the test proper, sample items, and one or more trial items. The answers to these trial items were given during the practice period and time was allowed for relevant questions.

It is probable that the instructions given in some of the practice tests suggest to a certain extent the method of approach required to reach the solution. Examples of this are Test L (Vertical Sections) in which the instructions are as follows:-

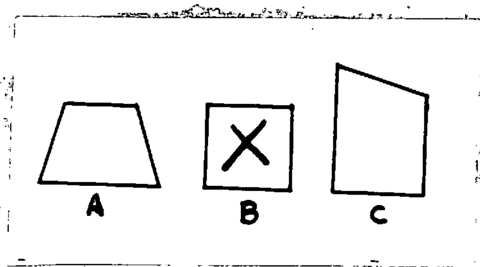
This diagram represents a wooden block which is to be sawn vertically into two parts in the place shown by the dotted lines.



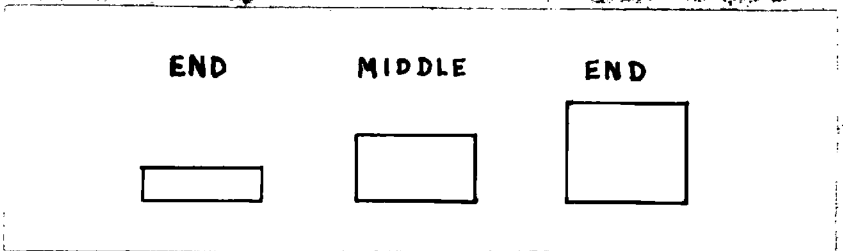
In this next diagram the two halves are moved apart to show the cut face. This cut face is shaded.



The cut face is the same shape as one of these three drawings.

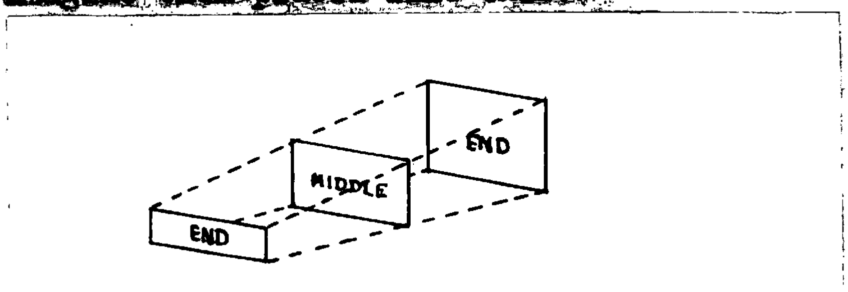


B is the correct shape and so is marked with a X; and Test I (Three Sections) where the instructions are:-

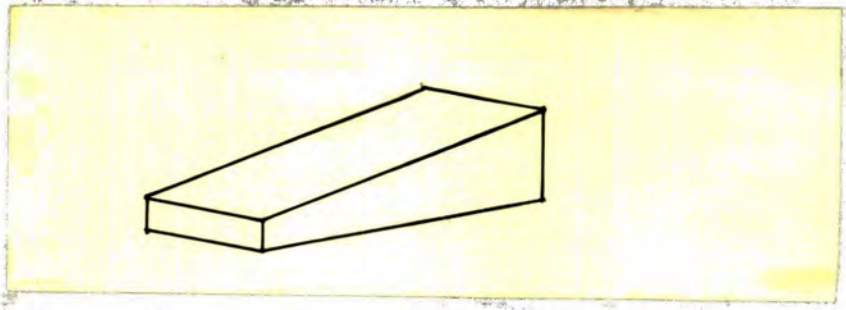


These three drawings represent the end views and middle sections of a solid object, in the order End, Middle, End as marked.

Imagine them placed like this:-



And you see that they form this object:-



That the instructions suggest the method of approach probably in no way affects the value of the test as a measurement of 'k'. It is more probable that, since the instructions tend to involve the use of spatial imagery, the testees will tend to use their spatial ability, rather than other abilities, to answer the problems.

Some 150 copies of each of the two halves of this try out draft were duplicated by the author from stencils prepared either directly by himself or by Gestetner Limited using a photographic process.

CHAPTER IV.

THE PRELIMINARY TRY-OUT.

As stated in an earlier chapter a preliminary try-out of the test items on a sample population provides the figures from which indices of difficulty and of discrimination can be calculated. Other advantages also accrue from such a try-out; the clarity or otherwise of the instructions becomes apparent, some indication of the time required to answer the items is obtained and any difficulties arising in the marking system can be checked.

The try-out of the first draft was carried out, in July 1958, on a sample drawn from the school population in the County Borough of Gateshead. In this area selection at 11+ results in allocation of the pupils either to one of the selective schools, comprising two Grammar Schools, Boys and Girls, and a mixed Secondary Technical School, or to one of a number of non-selective secondary schools. The try-out sample consisted of a representative group of those in their first year at these various schools, that is to say it was a sample of pupils who had sat an 11+ selection examination during the previous year.

The test was administered to a total of 109 pupils in four different schools, Boys' Grammar, Girls' Grammar, Low Fell Mixed Senior and Shipcote Senior Boys. The actual numbers from each school are shown in Table I.

SCHOOL	BOYS	GIRLS	TOTALS.	
Boy's Grammar	15	-	15	30
Girl's Grammar	-	15	15	
Low Fell Mixed	33	15	48	79
Shipcote Boys	31	-	31	
Totals.	79	30	109	

Table I. SAMPLE POPULATION - UNSELECTED.

The testing was carried out over a period of 10 days in the last few weeks of the summer term and efforts were made to keep the test situation as uniform as possible in the four schools. The usual arrangements as to spacing of desks, prevention of copying, answering of questions etc. were adhered to as far as practicable and all instructions given by the test supervisor, not in every case the author, were read from a prepared script. No special attempt was made at motivation except to state that this was a new test and the object was to investigate the test. This resulted in excellent co-operation and all pupils exhibited a high degree of what might be called "test sophistication", i.e. they were obviously familiar with paper and pencil tests and the general rules governing their administration.

Pleasure was expressed at doing tests rather than lessons and requests for more tests were common.

The two tests of 100 items each were administered separately, if possible on consecutive days, and always with Test 1 (sub-tests A - F) first. The teacher in charge, following exactly the procedure in the instruction script, instructed the pupils when to turn over, when to follow in their own booklets the instructions for the practice test and when to start each sub-test. He also read out the answers to the practice test and answered questions relevant to the practice tests. During the working of each sub-test no questions were allowed. The rules read to the pupils at the beginning included the instruction that they should omit any item they were unable to answer and go straight on to the next and that they should lay down their pencils when they reached the end of the sub-test. In this way the invigilator was able to see when all pupils had completed any particular sub-test.

The question of the time to be allowed for each sub-test was considered very carefully. In a test of this type, which is essentially one of Power rather than of Speed, it is usually stated that ample time should be allowed. To quote Thorndike[1] (1949) "the test should be given with quite

1. Thorndike, R. L. (1949). Personnel Selection.

ample time limits, so that most individuals will have a chance to try all items."

In theory all pupils should have time to complete all items, but the situation occurs where all except one, or possibly two, pupils have finished the sub-test and these one or two are unlikely to reach the end within reasonable time, or due to inability will completely fail to finish the test. Obviously some limit must be set to cope with this situation if only from the point of view that the other children will begin to lose interest from prolonged waiting. But there is a more important reason. To measure spatial ability, items are designed which depend on visual imagery for their solution. Those endowed with the ability are able to provide the answers with little difficulty or hesitation. Given very generous time limits it is possible that those not so endowed will be able to arrive at the correct solution by some other means such as counting sides or corners. In such a case the test ceases to be an accurate instrument for the measurement of 'k'.

For these reasons the supervisor of each group was asked to stop pupils working when 75% had completed the sub-test. A form was provided on which the time of commencement and the time of stopping of each sub-test was noted. It was found in practice that on those tests which could be completed rapidly all pupils had finished before a count of pupils laying down their pencils could be made. Examples of this were sub-tests K and H.

The scripts from all the schools concerned were marked and checked by the author and his wife. There being one answer, and one only, to each item, each correct response counts one mark giving a maximum possible total of 200 on the thirteen sub-tests. During the marking note was made of the solutions which were to be accepted and those which were to be rejected in Test M, the only sub-test which is not entirely objective. Marking also gave information as to the suitability of the instructions printed in the test. In one or two cases responses were made in which, while the intention of the pupil was clear and therefore the paper was valid, the nature of the response was not that required. For example, the placing of a tick by a diagram instead of a cross on it, or crossing out the wrong alternative instead of underlining the correct one. Such cases resulted in more emphatic instructions being included in the revised draft.

Taking the pupils* in the two grammar schools as being representative of the total population in the first forms of the selective schools and the pupils from the other schools as being the unselected portion, the figures in

Table I give the following percentages:-

Selective Schools:- 27.52%

Non-Selective Schools:- 72.48%

But figures supplied by the Director of Education show that the percentages in the total school population of the age group are:-

Selective Schools:- 23.55%

Non-Selective Schools:- 76.45%

Since it is essential that a sample should approximate as closely as possible to the true population, six scripts from the grammar schools were rejected by random selection. To keep the proportion of boys to girls in the grammar school sample as close as possible to that actually existing in the schools the scripts of 4 boys and 2 girls were rejected. The composition of the sample therefore becomes that shown in Table II.

SCHOOL	BOYS	GIRLS	TOTALS	
Boy's Grammar	11	-	11	24
Girl's Grammar	-	13	13	
Low Fell Mixed	33	15	48	79
Shipcote Boy's	31	-	31	
TOTALS	75	28	103	

TABLE II. SAMPLE POPULATION - AFTER SELECTION.

The proportions of this sample now approximate very closely to those existing in the true population as shown in Table III

	POPULATION	SAMPLE
Selective Schools	23.55%	23.3%
Non-Selective Schools	76.45%	76.7%

TABLE III. COMPARISON OF PERCENTAGES OF POPULATION AND SAMPLE.

The ages of the 103 pupils in the sample on the day the first half of the test was administered were calculated in years and completed months. These ages ranged from 11 years 10 months to 12 years 10 months with an average of 12 years 4 months.

The raw scores obtained by the 103 pupils ranged from 39 to 197, the possible total being 200. These scores were distributed about a Mean of 133.17, with a Standard Deviation of 30.815. The distribution of the scores within the range is shown in Table IV and a Frequency Polygon appears in Figure 11. These show that the distribution is negatively skewed.

CLASS	35/39	40/44	45/49	50/54	55/59	60/64	65/69	70/74	75/79	80/84	85/89	90/94	95/99	100/104
FREQUENCY	1	0	0	1	0	0	3	1	3	3	4	0	3	11

/cont

CLASS	105-109	110-114	115-119	120-124	125-129	130-134	135-139	140-144	145-149	150-154	155-159
FREQUENCY	4	8	4	5	9	6	7	4	4	6	2

CLASS	160-164	165-169	170-174	175-179	180-184	185-189
FREQUENCY	4	3	1	2	3	1

TABLE IV. DISTRIBUTION OF RAW SCORES.

This skewness of the curve can be explained, in part at least, by two characteristics of the sample. First, the sample is rather overbalanced in favour of boys, second the average age of the sample is some 15 - 16 months higher than the age of those for whom the test is intended. As will be seen later, allowance is made for this age difference in the selection of items for the final draft.

The basis for the selection of the items is the analysis of the data obtained from the try-out described above. The lengthy procedure of item analysis demands a chapter to itself; this chapter follows.

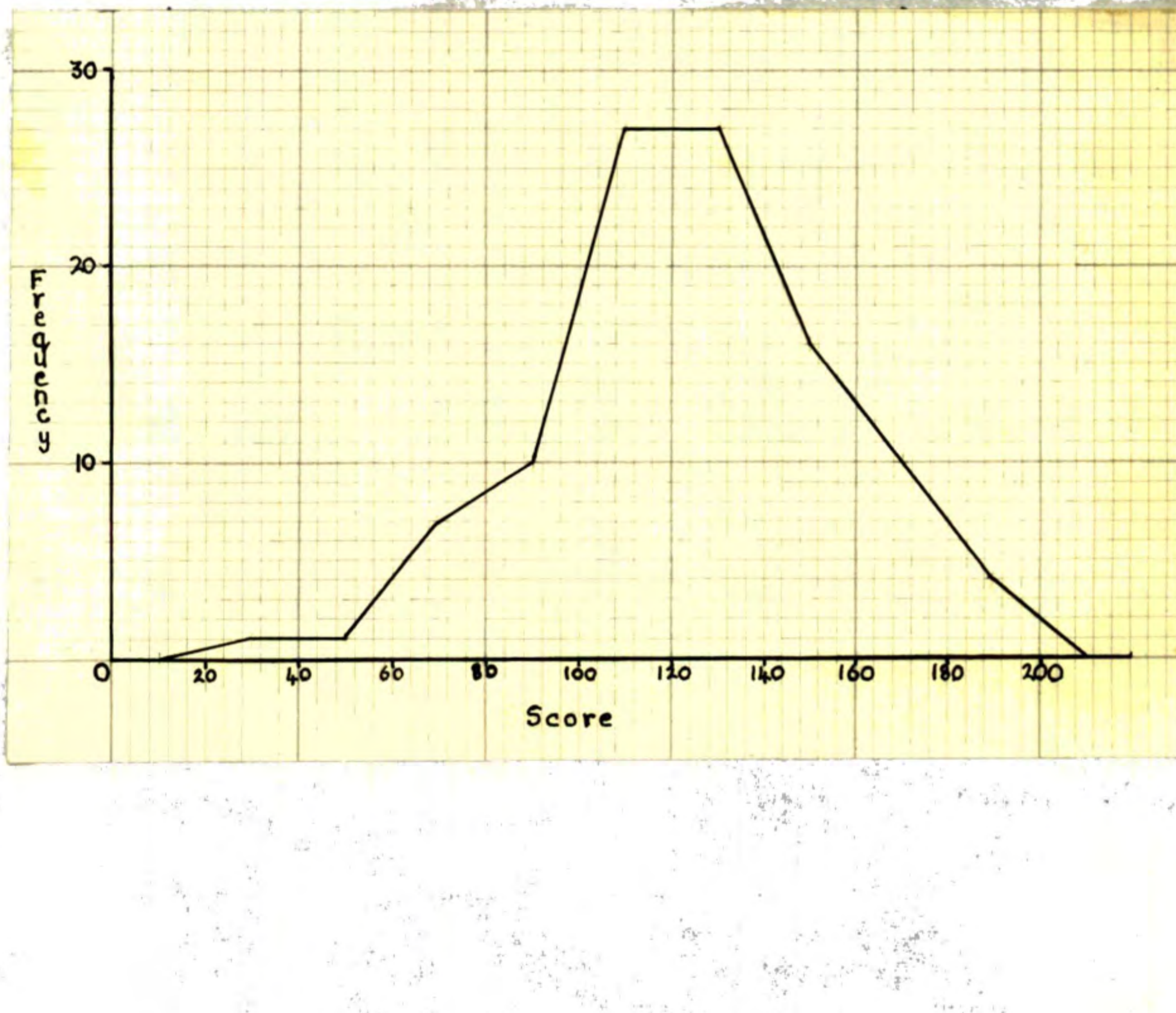


FIG. 11. DISTRIBUTION OF SCORES - GATESHEAD SAMPLE (N=103).

CHAPTER V.

ITEM ANALYSIS OF FIRST TRY-OUT.

The purpose of the analysis was to provide data by which to select those one hundred items which appeared to be most efficient in measuring spatial ability. For the selection two figures are required, an Index of Facility, and an Index of Discrimination.

The Facility of an item is a measure of the ease with which pupils are able to obtain the correct answer. It is usually expressed as the percentage of the sample population which succeeds on that item. A high Facility Index therefore indicates an easy item and a low index an item of some difficulty.

The Discrimination Index is a figure which indicates the degree with which success in a given item can be taken as a sign of possession of the ability under investigation. To obtain this figure the results for each item, in turn, must be judged in relationship to a standard criterion. In theory the most appropriate criterion is an external one such as the score obtained on a similar type of test which is of proved value as an instrument for measuring the ability in question. In practice an internal criterion is more often used, this usually being the total score obtained on the test under construction.

While the first method gives, as the index, a measure of item validity, the latter indicates the degree of internal consistency between items. When a test consists of more or less homogeneous items, internal consistency is a good guide to the more desirable items, although it has to be remembered that the consistency may be due to some factor other than that intended. If however items have been carefully designed at the outset this danger is possibly avoided.

Goodenough, (1950)[1] writes "the use of total score as a criterion for evaluating the separate items can best be justified when the test is of a kind that deals chiefly with the knowledge of facts or the possession of specified skills." With a test of the type described in this thesis the use of the internal criterion is, therefore, not inappropriate. All the items have been devised with the purpose of measuring spatial ability, and so a fair degree of intercorrelation is expected. For this reason, and in the absence of a suitable external criterion total score on the experimental draft of two hundred items has been used in the analysis here described.

Methods of calculating the discrimination index are many and varied but they fall into two contrasting cate-

---

1. Goodenough, F. L. (1950). Mental Testing.

gories, called by Vernon (1948)[1], Grouping Methods, in which criterion scores are divided into two or more categories, and Distribution Methods, in which criterion scores are treated as a continuous distribution. In practice the difference between these two methods is that with Grouping Methods the scripts are taken in order of criterion score and the response for each item tabulated, while in Distribution Methods, the criterion scores of those who pass an item are tabulated and also the scores of those who give other responses; this procedure being repeated for each item. This means that Distribution Methods entail far more clerical work but that they provide more information particularly about incorrect responses.

Experimental data quoted by Vernon in the work cited show that, in fact, Grouping Methods involve only about half the time taken by the other methods. Time available being an important limiting factor in the work here reported this conclusion of Vernon's weighed heavily in favour of using a Grouping Method of calculation.

Other considerations involved in coming to this decision were, (a) since the correct responses to the items are fixed objectively by the nature of the items,

1. Vernon, P. E. (1948) Indices of Item Consistency and Validity. Brit. J. of Psych. Stat. Sect. Vol. I.

detailed analysis was not essential, and (b) correction for guessing is easily applied to grouping methods, and it was felt that such correction might be necessary in certain sub-tests.

Having decided upon 'total test score' as the criterion and Grouping Methods as being most convenient it was then decided that the first analysis should be by the Upper and Lower Thirds method used by Moray House. In this procedure scripts are first divided into three equal groups according to the criterion score, and then for each item the proportion passing that item in the upper third is compared with the proportion passing in the lower third, the Discrimination Index being calculated according to the formula:-

$$D = \frac{U - L}{\frac{N}{3}}$$

where U = Number giving correct response in Upper Group

L = Number giving correct response in Lower Group

N = Total Number of Testees.

Recording the Data.

Where the number of pupils tested is large this use of only the Upper and Lower Third of the distribution results in a considerable reduction in the work involved in the tabulation of scores. With numbers in the region of 300 upwards it would be permissible to calculate the Facility Index on the percentage passes in these two thirds only, but with a total of 103 cases to be considered it is advisable to use all available data in the calculation of Item Facility. Therefore the responses of all the 103 pupils were recorded on score sheets of the type illustrated in Figure 12 one sheet being used for each one-third of the pupils.

NAMES	A. Bee	F. Gee	J. Kay	etc.																TOTALS		
																				R	W	O
Item 1	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/			17		
Item 2	/	/	/	/	W	/	/	/	/	/	W	/	/	/	O	/	/			14	2	1
Item 3	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	W	/			16	1	
Item 4	/	/	/	/	/	O	/	W	/	/	/	/	W	/	W	/	W			12	4	1
Item 5	/	/	W	/	/	/	/	/	/	W	/	/	/	/	/	/	O			14	2	1
Item 6	/	/	/	/	/	/	//	/	O	W	/	/	W	/	/	W	N			12	3	1
Item 7																						

FIGURE 12. SCORE SHEET SHOWING CORRECT RESPONSES (/) WRONG RESPONSES (W) OMITTED ITEMS (O) AND ITEMS NOT ATTEMPTED (N)

The nature of incorrect responses, i.e. a wrong response, an item attempted but omitted, or an item not attempted, was recorded on the sheet to facilitate the application of a guessing correction if this proved necessary. The distinction between an omitted item and one not attempted is that given in Thorndike (1949)[1]:- "assume that all items up to and including the last one for which an answer was marked were attempted and that the first unanswered item after the last answered one was also attempted. We then assume that starting with the second item after the last one answered the remaining items were not tried."

When the score sheets were completed the Total for each item in each third of the sample was transferred to record cards ruled as in Figure 13. Each card bearing the data for ten items.

---

1. Thorndike, R. L. (1949). Personnel Selection.

M - Paper Folding.						
THIRDS	UPPER 33	MIDDLE 36	LOWER 34	TOTAL 103	FACILITY	DISCRIMINATION.
Item 1	$33^{0}/0$	$34^{2}/0$	$30^{3}/1$	97		
2	$26^{7}/0$	$14^{22}/0$	$2^{31}/1$	42		
3	$33^{0}/0$	$30^{6}/0$	$15^{19}/0$	78		

FIGURE 13. RECORD CARD FOR TOTALS IN EACH THIRD. EACH COLUMN RECORDS CORRECT RESPONSES, WRONG RESPONSES AND OMITTED ITEMS IN THE ORDER R^{W}/O.

As Figure 13 shows the number of scripts included in each third was not the same. This situation arose because there were four pupils with a total score of 135. This score was the dividing line between the upper and middle thirds and so the four scripts could be included in either group. It was decided to include them in the middle "third" to give proportions of 33 : 36 : 34 and so avoid the inflation of the discrimination index which would result if they were placed in the upper third with proportions of 37 : 32 : 34.

Calculation of Facility Indices.

From the data collected in this way the Facility Index for each item was first calculated. This was done in the first instance by using the raw scores, that is, expressing the total right responses (R), uncorrected for guessing, as a percentage of 103. This was followed by a second series of calculations using scores to which a guessing correction had been applied, where applicable, and expressing the corrected total right responses (Rc) as a percentage of 103 - NA, where NA is the number of pupils not attempting the item. In terms of formulae the two methods are

(I) $F = \frac{R}{103} \times 100.$ (II) $F = \frac{Rc}{103 - NA} \times 100.$

The guessing correction applied is a version of that given in Thorndike (1949)[1]

$$Rc = R - \frac{W}{n - 1}$$

where Rc = Corrected rights.

R = Rights.

W = Wrongs (and does not include omitted or not attempted items.)

n = number of response alternatives.

---

1. Thorndike, R. L. (1949). Op.cit.

The two sets of indices found by these calculations are set out in Tables Va and Vb. Those given by formula (I) will tend, in multiple choice items at least, to be inflated since no correction has been applied; those found by formula (II) will not suffer from this defect but will be affected by the denominator being the number which attempted the item and not the complete sample. This assumes that if those pupils not attempting items had in fact attempted it, their responses would be in the same proportions as those actually recorded. This is probably a false assumption since the slow pupils not attempting the later questions are often those who have most difficulty with the items. However, using two indices in conjunction will give a fair approximation of the true difficulty level.

On the first examination of these indices two things are apparent; in Sub-test C (Profiles) a large proportion of the items are of considerable difficulty, half being of less than 20% facility; and in Sub-test J (Nails in Blocks) the average difference between the indices found by the two methods is appreciably more than in any other sub-test. This second point is explained by the fact that this is a YES/NO type of test and therefore the "correction-for-guessing" formula becomes simply "Rights-minus-wrongs" and gives a relatively lower corrected total than in any other

ITEM	(I)	(II)	ITEM	(I)	(II)	ITEM	(I)	(II)	ITEM	(I)	(II)
A 1a	73	73	B 6	39	24	C 21	26	16	E 6	87	83
2a	94	94	7	68	60	22	42	24	7	67	52
b	96	96	8	69	62	23	22	9	8	25	7
c	95	95	9	78	76	24	18	10	9	65	50
3a	91	91	10	54	48	25	26	8	10	50	37
b	93	93	C 1	53	31	26	46	43	F 1	48	48
c	43	43	2	61	42	27	23	11	2	37	37
d	84	84	3	66	50	28	18	12	3	41	41
4a	88	88	4	65	49	29	13	3	4	19	19
b	65	65	5	43	21	30	39	47	5	31	31
5a	94	95	6	50	27	D 1	88	84	6	39	39
b	74	75	7	53	31	2	38	20	7	14	14
6a	76	77	8	22	6	3	28	5	8	50	50
b	82	83	9	44	20	4	32	14	9	44	44
c	59	60	10	28	8	5	57	45	10	22	22

ITEM	(I)	(II)	ITEM	(I)	(II)	ITEM	(I)	(II)	ITEM	(I)	(II)
G 1	88	89	I 6	78	78	K 1	91	88	L 16	72	63
2	64	65	7	92	92	2	94	92	17	51	37
3	83	84	8	82	82	3	83	79	18	63	52
4	85	86	9	82	82	4	88	84	19	57	45
5	80	81	10	69	70	5	95	93	20	59	48
6	87	88	J 1	85	71	6	84	80	M 1	94	94
7	86	87	2	85	72	7	76	68	2	41	41
8	86	87	3	87	77	8	64	52	3	76	76
9	80	80	4	62	24	9	75	69	4	50	50
10	92	96	5	92	85	10	81	81	5	59	59
H 1	39	39	6	81	62	L 1	95	93	6	91	91
2	88	88	7	86	73	2	84	78	7	87	87
3	57	57	8	74	49	3	56	36	8	33	33
4	67	67	9	97	93	4	70	55	9	48	47
5	50	50	10	62	25	5	59	40	10	86	87
6	97	97	11	50	7	6	53	37	11	66	65

sub-test. The level of difficulty found in sub-test C suggests that many of these items should be rejected in a revised draft, but this can only be adequately decided by viewing the Facility Indices in relation to those for Discrimination, therefore the question must rest until the latter indices have been recorded.

Discrimination by Upper and Lower Thirds Method.

As stated earlier in this chapter the Discrimination Indices were first calculated by the Upper and Lower Thirds method, using the formula $D = \frac{U - L}{\frac{N}{3}}$ and the uncorrected figures on the record cards. This procedure produced a total of 57 items having Discriminations of .45 or higher, a figure chosen as the selection point for reasons discussed later.

In an effort to find more items that might possibly be of fair discriminating power the analysis was repeated, using the same method but with the formula becoming $D = \frac{Uc - Lc}{\frac{N}{3}}$ When U_c and L_c are total right responses, in Upper and Lower Thirds, after guessing correction has been applied, the method of correction being that already discussed. These calculations resulted in a total of 68 items which reached the .45 level of discrimination.

A complete list of the indices found by these two sets of calculations can be found in TABLE VIII. A summary table showing the distribution of the items fulfilling the criterion is given in TABLE VI.

These figures proved to be insufficient to select one hundred items for the revised draft of the test. The totals quoted, 57 and 68, are totals of those items reaching a discrimination criterion, regardless of difficulty. The figures are of value only in so far as they indicate that certain sub-tests, mainly B. E. F. L. and M. are far superior to others in their discrimination as complete sub-tests.

It was therefore decided to analyse the figures on the score sheets by a method which made more use of the available data. This decision was influenced by the discovery that there were items in which all, or many, of those giving the correct answer were in the upper third, but the Discrimination Index was low because the total numbers of these successes were less than 15. Examples are Items 7, 16 and 20 of sub-test F, the figures for these being:-

	Upper Third	Middle Third	Lower Third	Discrimination.
Item 7	11	3	0	0.32
Item 16	11	3	0	0.32
Item 20	6	0	0	0.17

It was felt that where all correct responses were in the

SUB-TEST.	No. of Items.	No. of ITEMS REACHING CRITERION. 1	2
A - Counting Faces	20	2	N.A.
B - Plan Views	10	4	7
C - Profiles	30	2	4
D - Block Building	10	1	1
E - Overlying Patterns	10	6	6
F - 3-D Networks	20	14	N.A.
G - Front and Side	10	1	N.A.
H - Form Board	10	0	N.A.
I - Three Sections	10	1	N.A.
J - Nails in Blocks	20	0	2
K - Corner Recognition	10	0	0
L - Vertical Sections	20	11	15
M - Paper Folding	20	15	N.A.
TOTAL ITEMS SUCCEEDING		57	68

TABLE VI. DISTRIBUTION OF ITEMS WITH DISCRIMINATION INDICES OF 0.45 OR MORE BY UPPER AND LOWER THIRDS METHOD.

COLUMN 1 - CALCULATED ON RAW SCORES.

COLUMN 2 - CALCULATED ON CORRECTED SCORES.

N.A. = CORRECTION NOT APPLICABLE.

Upper Third then that item was discriminating more successfully than this type of index showed, and that what was required was an index which depended upon the percentage of success in the respective thirds rather than upon the totals succeeding.

Discrimination by Flanagan's Product-Moment Method.

Such an index is that outlined by Flanagan (1939)[1]. This index is obtained from a chart which is based on the findings of Kelley (1939)[2] that upper and lower groups containing twenty-seven per cent of the cases are optimum for certain related estimations. From the chart can be found values of the product-moment coefficient of correlation corresponding to given proportions of success in the upper and lower twenty-seven per cent of the criterion group.

Thorndike (1949)[3] describes this as "the most satisfactory item validity index based on the upper and lower twenty-seven per cent", and prints as an appendix to his

---

1. Flanagan, J. C. (1939). General Considerations in the Selection of Test Items and a Short Method of Estimating the Product-Moment Coefficient from Data at the Tails of the Distribution. J. of Educ. Psych. Vol. XXX.

2. Kelley, T. L. (1939). The Selection of Upper and Lower Groups for the Validation of Test Items. J. of Educ. Psych. Vol. XXX.

3. Thorndike, R. L. (1949) Op.cit.

book tables based on Flanagan's chart. All that is necessary is to convert scores in the groups to percentages and then read off the correlation values by entering rows and columns at the appropriate points.

The score sheets used for the Upper and Lower Thirds method were utilised for obtaining the requisite percentages. On these sheets the names and scores of those 5 and 6 pupils, of the upper and lower thirds respectively, who did not come within the 27 per cent were "blacked out". This left 28 names and scores on each sheet; 27% of 103 being 27.81. The procedure which then followed was both simple and economical in time. For each item the number giving the right response in the upper third was counted and converted to a percentage by slide rule, this was then repeated for the lower third and then the product-moment correlation read off from the table. The results of the analysis were highly satisfying, there being 129 items which had coefficients greater than 0.45, and of these coefficients 102 were greater than 0.5. The distribution of these coefficients within the sub-tests is shown in TABLE VII and Figure 14, the complete list is given in TABLE VIII. Examination of these figures confirms the conclusion of the earlier analysis that sub-tests B, E, F, L and M are most superior in discriminating power and also add sub-tests A and I to the list.

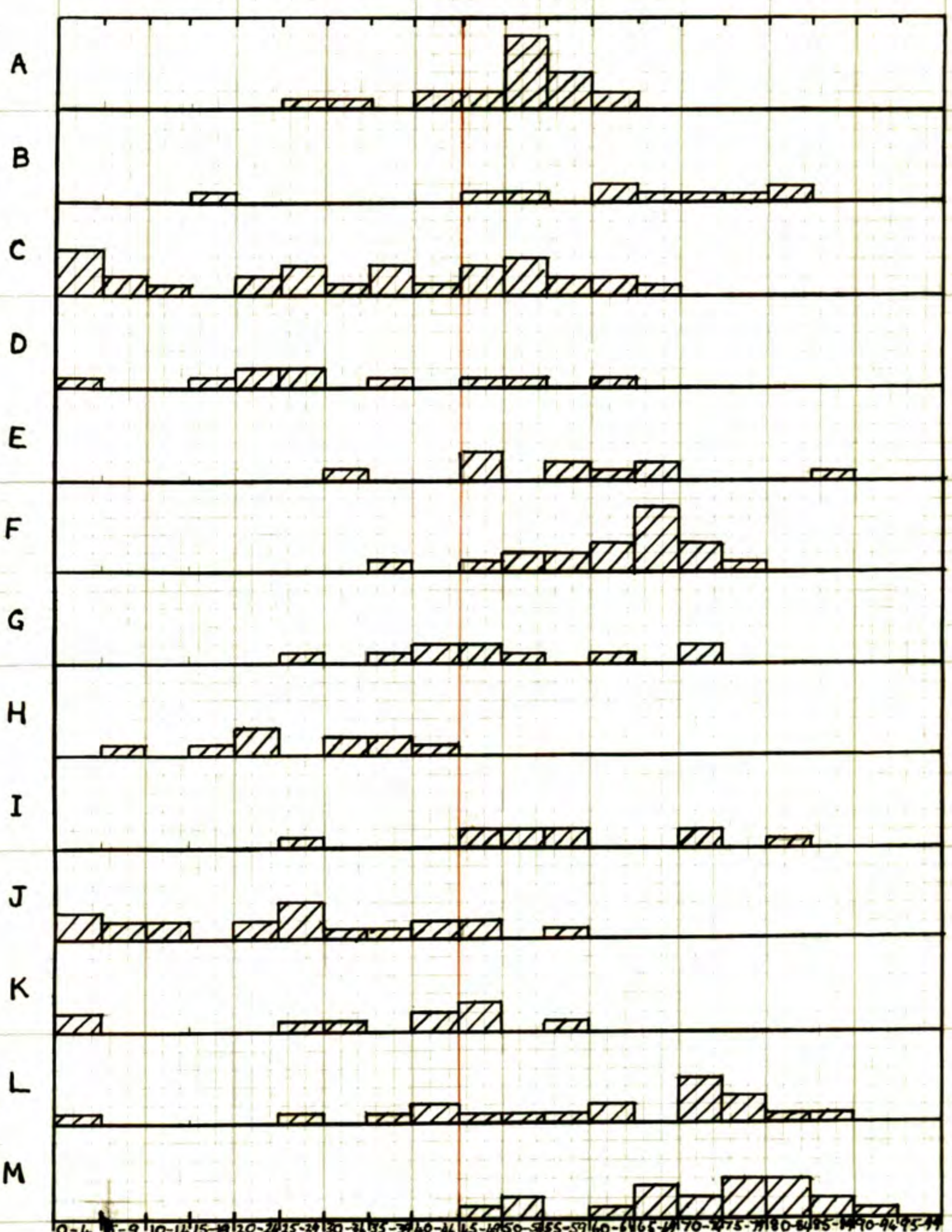


FIGURE 14. HISTOGRAMS SHOWING DISTRIBUTION OF DISCRIMINATION INDICES BY FLANAGAN'S METHOD.

Before discussion of the procedure employed for selecting items on the basis of these figures, some justification of the criterion levels adopted must be presented.

SUB-TEST.	No. of Items.	No. OF COEFFICIENTS.	
		= or > 0.45	= or > 0.50.
A - Counting Faces	20	16	14
B - Plan Views	10	9	8
C - Profiles	30	12	9
D - Block Building	10	3	1
E - Overlying Patterns	10	9	6
F - 3-D Networks	20	19	18
G - Front and Side	10	8	4
H - Form Board	10	0	0
I - Three Sections	10	9	7
J - Nails in Blocks	20	3	1
K - Corner Recognition	10	4	1
L - Vertical Sections	20	17	14
M - Paper Folding	20	20	19
TOTALS		129	102

TABLE VII. DISTRIBUTION OF DISCRIMINATION INDICES OBTAINED BY FLANAGAN'S PRODUCT-MOMENT METHOD.

Justification of Criterion Level.

In the selection of items two methods can be used. Either the items are placed in order according to their

Discrimination Indices and the top one hundred taken to form the revised test regardless of the point at which the line dividing this hundred from the rest comes, or items can be selected which reach, or exceed, a predetermined level of discrimination. This latter method has been used here, with a discrimination index of 0.45 taken as the criterion level. An attempt will now be made to show why this level was chosen.

Goodenough (1950)[1] suggests that "some point in the distribution of the 't' statistic will be chosen as the dividing line" and goes on to say that "Most people have found that a value of 'r' or of 't' that reaches the 25 per cent level of confidence is sufficient to justify the inclusion of an item in a test made up of not fewer than 50 equally weighted items."

In "Upper and Lower Third" indices the criterion commonly used by Moray House is 0.45. This can be shown amply to fulfil Goodenough's conditions. The "t" statistic or Critical Ratio of a sample statistic is usually determined by dividing the observed value by the Standard Error of that value. The Standard Error of a U - L index can be calculated by the following formula of Johnson, A.P. (1951)[2]

---

1. Goodenough, F. L. (1950). Mental Testing.
2. Johnson, A. P. (1951). Notes on a Suggested Index of Item Validity: the U-L. Index. J. of Educ. Psych. XLII.

$$\text{S.E. of U - L Index} = \frac{1}{f}\sqrt{RU + RL - \left(\frac{RU^2 + RL^2}{f}\right)}$$

where RU = Rights in Upper Group.

RL = Rights in Lower Group.

f = size of group $(\frac{N}{3})$

Substituting in this formula the values which would give an index of 0.45 viz. RU = 30, R_L = 14.5, f = 34.33 we get:-

S.E. of Index = 0.1016.

This figure will give a Critical Ratio of 4.427 which far exceeds the figure required for 25% level of confidence and, in fact, easily satisfies the 1% level of confidence. The tables supplied by Fisher[1] show that for N = 30 the Critical Ratio at the 1% level of confidence is 2.750, the value for N = 103 will be lower than this.

Turning now to the Product-Moment Correlation of Flanagan, two formulae will be used to show that here again indices of 0.45 and above satisfy the 1% level of confidence. The first is taken from Chambers' (1952)[2] who states that for small samples the significance of a product moment coefficient should be assessed by the formula

$$t = \frac{r\sqrt{N - 2}}{\sqrt{1 - r^2}}$$

1. Fisher, R. A. (1946). Statistical Methods for Research Workers. 10th edition.

2. Chambers, E. G. (1952). Statistical Calculations for Beginners.

substituting $r = 0.45$ and $N = 103$

$$t = 5.063$$

which easily satisfies the conditions.

The second formula is taken from Zubin (1936)[1] who quotes it as being "the usual formula" and uses it in an example about the selection of test items that differentiate significantly between two groups. It is:-

$$C.R. = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{N_1} + \frac{P_2 Q_2}{N_2}}}$$

Where P_1 = Proportion succeeding in Upper Group.

P_2 = Proportion succeeding in Lower Group.

Q_1 = Proportion failing in Upper Group.

Q_2 = Proportion failing in Lower Group.

N_1 = Number in Upper Group.

N_2 = Number in Lower Group.

This formula can be shown to be another variation of

$$\frac{\text{Observed Value}}{\text{S.E. of Observed Value.}}$$

The full proof is set out in Appendix A.

Substituting in this formula the following data of item K.9 which has an index of 0.45.

$P_1 = 90,$ $Q_1 = 10.$ $N_1 = 28.$

$P_2 = 54,$ $Q_2 = 46.$ $N_2 = 28.$

1. Zubin, J. (1936). Note on a Graphic Method of Determining the Significance of the Difference between Group Frequencies. J.of Ed.Psych.XXVII.

we get

$$\text{Critical Ratio} = 3.274.$$

which once again more than adequately fulfills the demands of a 1% level of confidence.

On all this evidence it can be concluded that with the two types of Discrimination Indices here employed a criterion level of 0.45 can safely be used for selecting suitable items.

Selection of Items for Revised Draft.

In a test of the type described in this thesis, where the items are grouped into sub-tests and where the sub-tests are preceded by practice tests, it is not practical to select items solely on their showing in the lists of Facility and Discrimination Indices. The selection must be on the basis of statistics plus practical considerations. It would, for instance, be highly impracticable to include the one item from Sub-test D which satisfies the numerical requirements for this would entail the printing of practice test and instructions for one solitary item. What has to be looked for are sub-tests which as a whole satisfy the requirements or from which sufficient items can be extracted to form a sub-test of suitable length.

Apart from the consideration of minimum size for any given sub-test there is also a practical limit to the number of sub-tests. A disproportionately large number of sub-tests will increase the time required for the test with consequent danger of loss of attention and interest. Printing costs also have to be considered, a six sub-test booklet requires 24 pages, which must be viewed as the upper limit.

With these points in mind, and setting statistical requirements of (a) Discrimination Indices to be equal or greater than 0.45 and (b) Facility Indices to be between 35% and 95%, the data set out in Table VIII was examined.

The limits to the Facility Indices were set at 35% and 95% to allow for the fact that the age range of the sample population was fourteen months higher than that of the population for which the test is ultimately intended.

From the data of Tables VI, VII and VIII and Figure 14, it was concluded that the following sub-tests were most suitable for inclusion in the revised draft, with the proviso that those items which failed to reach the required statistical levels should be either completely replaced or revised to bring their Facility Indices within the required range.

Sub-test A. Counting Faces.

B. Plan Views.

E. Overlying Patterns.

F. 3-D Networks.

I. Three Sections

L. Vertical Sections.

M. Paper Folding.

The manner in which these tests were organised into the revised draft is discussed in the next chapter.

ITEM	DISCRIMINATION			FACILITY		ITEM	DISCRIMINATION			FACILITY	
NOS.	1	2	3	I	II	NOS.	1	2	3	I	II
A 1a	0.47		0.64	73	73	C 21	0.52	0.44	0.66	26	16
2a	0.12		0.51	94	94	22	0.32	0.44	0.39	42	24
b	0.06		0.40	96	96	23	0.32	0.22	0.54	22	9
c	0.09		0.46	95	95	24	0.38	0.23	0.54	18	10
3a	0.12		0.51	91	91	25	0.15	0.19	0.20	26	8
b	0.15		0.54	93	93	26	-0.19	-0.19	-0.15	46	43
c	0.32		0.46	43	43	27	0.29	0.25	0.35	23	11
d	0.17		0.32	84	84	28	0.35	0.26	0.54	18	12
4a	0.17		0.40	88	88	29	0.17	0.06	0.27	13	3
b	0.41		0.53	65	65	30	0.09	0.17	0.08	39	47
5a	0.15		0.54	95	95	D 1	0.12	0.20	0.27	88	84
b	0.38		0.57	74	75	2	0.32	0.34	0.36	38	20
6a	0.32		0.57	76	77	3	0.12	0.15	0.15	28	5
b	0.12		0.28	82	83	4	0.38	0.42	0.50	32	14
c	0.44		0.54	59	60	5	-0.08	-0.07	-0.03	57	45
7a	0.44		0.54	51	51	6	0.12	0.04	0.25	23	4
b	0.32		0.51	76	77	7	0.47	0.60	0.64	61	52
c	0.15		0.58	90	91	8	0.09	0.14	0.23	62	56
8a	0.52		0.61	58	58	9	0.17	0.23	0.22	70	70
b	0.17		0.58	90	91	10	0.23	0.15	0.45	23	9
B 1	0.29	0.40	0.66	88	84	E 1	0.35	0.39	0.48	76	75
2	0.38	0.50	0.71	83	78	2	0.47	0.68	0.65	78	68
3	0.06	0.08	0.15	32	10	3	0.09	0.15	0.46	93	90
4	0.44	0.55	0.61	56	43	4	0.52	0.66	0.61	49	28
5	0.41	0.53	0.51	55	44	5	0.52	0.48	0.59	37	17
6	0.47	0.50	0.60	39	24	6	0.20	0.29	0.48	87	83
7	0.55	0.72	0.82	68	60	7	0.50	0.73	0.66	67	52
8	0.35	0.43	0.48	69	62	8	0.26	0.22	0.31	25	7
9	0.50	0.62	0.78	78	76	9	0.47	0.67	0.58	65	50
10	0.76	0.85	0.82	54	48	10	0.70	0.79	0.85	50	37
C 1	0.41	0.61	0.48	53	31	F 1	0.52		0.63	48	48
2	0.32	0.50	0.42	61	42	2	0.70		0.79	37	37
3	0.20	0.32	0.26	66	50	3	0.61		0.65	41	41
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]

ITEM		DISCRIMINATION			FACILITY		ITEM		DISCRIMINATION			FACILITY	
NOS.		1	2	3	I	II	NOS.		1	2	3	I	II
G	1	0.20		0.63	88	89	K	1	0.15	0.20	0.41	91	88
	2	0.55		0.70	64	65		2	-0.03	-0.03	0.00	94	92
	3	0.35		0.71	83	84		3	0.15	0.27	0.40	83	79
	4	0.23		0.51	85	86		4	0.17	0.23	0.58	88	84
	5	0.15		0.25	80	81		5	0.03	0.05	0.00	95	93
	6	0.15		0.44	87	88		6	0.26	0.35	0.48	84	80
	7	0.15		0.35	86	87		7	0.20	0.28	0.29	76	68
	8	0.20		0.48	86	87		8	0.23	0.31	0.32	64	52
	9	0.29		0.49	80	80		9	0.32	0.42	0.45	75	69
	10	0.06		0.40	92	96		10	0.26	0.31	0.48	81	81
H	1	0.17		0.19	39	39	L	1	0.12	0.19	0.51	95	93
	2	0.03		0.08	88	88		2	0.26	0.41	0.63	84	78
	3	0.38		0.42	57	57		3	0.38	0.57	0.44	56	36
	4	0.20		0.23	67	67		4	0.61	0.92	0.73	70	55
	5	0.23		0.22	50	50		5	0.61	0.92	0.73	59	40
	6	0.03		0.33	97	97		6	0.64	0.79	0.70	53	37
	7	0.15		0.37	88	88		7	0.55	0.84	0.78	72	58
	8	0.12		0.37	84	84		8	0.64	0.83	0.73	52	34
	9	0.09		0.23	89	89		9	0.70	0.96	0.84	72	60
	10	0.15		0.32	78	78		10	0.29	0.45	0.45	67	52
I	1	0.32		0.49	79	79		11	-0.06	-0.07	-0.04	28	27
	2	0.38		0.70	81	81		12	0.41	0.53	0.44	42	23
	3	0.44		0.51	53	53		13	0.09	0.12	0.27	83	77
	4	0.20		0.58	92	92		14	0.32	0.44	0.37	63	50
	5	0.55		0.82	71	71		15	0.61	0.77	0.71	46	29
	6	0.26		0.48	78	78		16	0.38	0.50	0.59	72	63
	7	0.17		0.51	92	92		17	0.67	0.84	0.79	51	37
	8	0.32		0.57	82	82		18	0.67	0.89	0.87	63	52
	9	0.41		0.72	82	82		19	0.61	0.81	0.75	57	45
	10	0.23		0.29	69	70		20	0.61	0.81	0.64	59	48
J	1	0.12	0.26	0.23	85	71	M	1	0.09		0.46	94	94
	2	0.15	0.29	0.40	85	72		2	0.70		0.77	41	41
	3	0.09	0.20	0.28	87	77		3	0.52		0.78	76	76
	4	0.00	0.03	0.08	62	24		4	0.50		0.54	50	50
	5	0.17	0.35	0.58	92	85		5	0.58		0.67	59	59

CHAPTER VI.

THE REVISED DRAFT.

As stated at the end of Chapter V, the selection of items can be approached from two points of view, theoretical grounds and practical considerations. Theoretical grounds are based on the statistics of Facility and Discrimination, and these have already been utilised to find those items which satisfy the criteria of

a) Facility between 35 and 95.

b) Discrimination .45 or greater.

A further theoretical consideration which has important practical consequences is the type of distribution of scores which will be obtained when the test is applied to a population of reasonable size. The distribution of raw-scores in the First Try-Out approximated to a Normal curve. For a test, such as this, intended for selection, a raw-score distribution which approaches the Rectangular is of greater service; since a 'cut-off' can be made at any point. Therefore when considering the items from the point of view of Facility, an even spread over the whole range should be aimed at.

The practical considerations already mentioned are printing costs, dependent on the number of sub-tests included; and length of working-time when the test is administered. It was decided that twenty-four was the maximum number of pages which could be entertained. Working as the basis that the minimum space necessary for a sub-test preceded by a practice test is 4 pages (2 sheets), this booklet size therefore limits the number of sub-tests to six. Within a sub-test the lowest "economical" number of items was fixed at 10. ("Economical" being used here to indicate the ratio of items to practice test.) The total number of items aimed at was one hundred.

From the figures of Chapter V, the sub-tests selected as possibles on the statistical grounds numbered seven which together contained 110 items:-

A - Counting Faces.	20 items
B - Plan Views.	10 items
E - Overlying Patterns.	10 items
F - 3-D Networks.	20 items
I - Three Sections.	10 items
L - Vertical Sections.	20 items
M - Paper Folding.	20 items

The first step was therefore to reduce the number of sub-tests to six and the number of items to one hundred. Inspection of the seven sub-tests showed that sub-test E is the only one which does not employ three-dimensional diagrams, relies for its response on some knowledge of "equation-form" and is possibly the one test of the seven which most relies on 'g' rather than 'k'. This test was therefore ruled out thereby reducing both the number of tests and the number of items to the required levels.

The remaining six tests, hereafter called the 'selected tests', obviously contain some items which fail by wide margins to reach the statistical requirements. The efforts made to improve or replace these items will be detailed, but first a slight digression will be made.

The six selected tests were all, in the Try Out Draft, preceded by practice tests. Is this due to chance or can some reason be found for it? First it must be pointed out that while the unselected group contained all those tests (H, J, K) not preceded by practice it also included four tests (C, D, E, G) which had previous practice. That is, tests which had practice were also rejected, some narrowly such as E and some at an early stage of the analysis, e.g. sub-tests C and D. In other words those tests provided with practice items form a series of suitability ranging from the very poor to the good, and some therefore fall in the selected group and the rest in the unselected. Turning to the unpractised tests (H, J and K) it is seen that these items satisfy on Facility but have been rejected on Discrimination. In H and K, particularly, the Facilities are high and thus the items would fail to discriminate. Possibly these sub-tests would be more useful with a lower age-group. The same comments can be made, to a certain extent, about sub-test J, but here there might be some other factor inherent in the nature of the test or due to the draughtsmanship of the drawings which would present problems inconsistent with those posed in the other sub-tests. From this it can be concluded that there are a number of reasons apart from the possession or non-possession of a practice test which affect the selection of the tests. It may be noted here that Bain (1946)[1], found no significant difference between the scores obtained by a group receiving instruction and the scores of a group receiving no instruction except that which was included within the test material.

Returning now to the re-organisation of the selected tests to form a new draft reference will be made to the figures shown in Table IX. The first step was to revise or replace unsatisfactory items. Starting with Test A, the figures show four items which do not fall within the criteria limits. Of these, two have Discrimination Indices of 0.40 which, while not fulfilling the statistical requirements set, are not grounds for complete rejection of the

---

1. Bain, J. T. (1946). The Construction of a Space Test. B.Ed. Thesis. Edinburgh.

items since the 0.45 Discrimination level has been shown to be more than adequate at the 1% level of confidence and it is usual to accept an index satisfying the 25% level of confidence. Furthermore, this Test, A, is of a type where a number of items are based on one diagram and efforts to improve one item by redrawing may easily have adverse effects on several items. Therefore these two items are retained. For this latter reason the other two items, whose indices fall below requirements, are also retained.

Test B contains one item only which fails, on both Facility and Discrimination. This item was therefore re-drawn. The test consists of identifying the plan view of blocks placed together. The unsatisfactory item was the only one which had the blocks placed together so that one was seemingly suspended in the air; it was therefore re-drawn with the blocks placed 'on the ground'. (Fig.15)

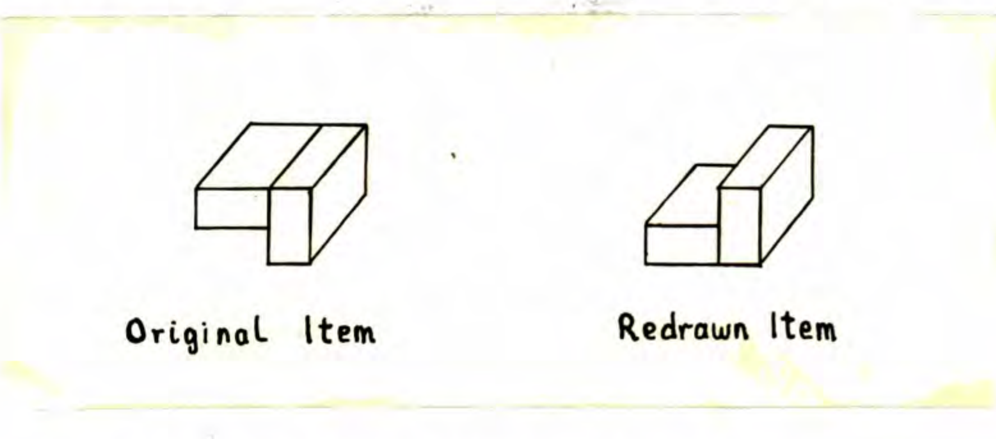


FIG. 15 SHOWING REVISION OF ITEM B.3.

A-Counting Faces.			B-Plan Views.			F-3-D Net-works.			L-Vertical Sections.			M-Paper Folding.		
Item	F	D	Item	F	D	Item	F	D	Item	F	D	Item	F	D
1a	73	0.64	1	88	0.66	11	42	0.39	1	95	0.51	1	94	0.46
2a	94	0.51	2	83	0.71	12	48	0.67	2	84	0.63	2	41	0.77
b	96	0.40	3	32	0.15	17	57	0.69	3	56	0.44	3	76	0.78
c	95	0.46	4	56	0.61	19	30	0.63	4	70	0.73	4	50	0.54
3a	91	0.51	5	55	0.51	15	33	0.52	5	59	0.73	5	59	0.67
b	93	0.54	6	39	0.60	16	14	0.66	6	53	0.70	6	91	0.54
c	43	0.46	7	68	0.82	17	17	0.60	7	72	0.78	7	87	0.63
d	84	0.32	8	69	0.48	18	22	0.74	8	52	0.73	8	33	0.79
4a	88	0.40	9	78	0.78	19	44	0.67	9	72	0.84	9	48	0.91
b	65	0.53	10	54	0.82	20	6	0.51	10	67	0.45	10	86	0.70
5a	95	0.54	F-3-D Net-works.			I-Three Sec-tions.			11	28	-0.04	11	66	0.84
b	74	0.57	1	48	0.63	1	79	0.49	12	42	0.44	12	50	0.82
6a	76	0.57	2	37	0.79	2	81	0.70	13	83	0.27	13	80	0.78
b	82	0.28	3	41	0.65	3	53	0.51	14	63	0.37	14	60	0.82
c	59	0.54	4	19	0.71	4	92	0.58	15	46	0.71	15	55	0.87
7a	51	0.54	5	31	0.59	5	71	0.82	16	72	0.59	16	60	0.83
b	76	0.51	6	39	0.48	6	78	0.48	17	51	0.79	17	80	0.68
c	90	0.58	7	14	0.68	7	92	0.51	18	63	0.87	18	44	0.89
8a	58	0.61	8	50	0.74	8	82	0.57	19	57	0.75	19	71	0.65
b	90	0.58	9	44	0.59	9	82	0.72	20	59	0.64	20	79	0.70
			10	22	0.68	10	69	0.29						

TABLE IX. FACILITY AND DISCRIMINATION INDICES OF SELECTED TESTS.

FACILITY EXPRESSED AS % OF 103

DISCRIMINATION BY FLANAGAN'S METHOD.

INDICES FULFILLING CRITERIA PRINTED IN RED.

Test F required much reorganisation. This test supplied items with good discrimination but ten items had low Facility indices. Of these ten, four were retained and placed toward the end of the test while the other six were re-drawn. This redrawing consisted of simplification by reducing the number of corners on the solid shapes and thus the number of circles required in the response (see Fig. 16)

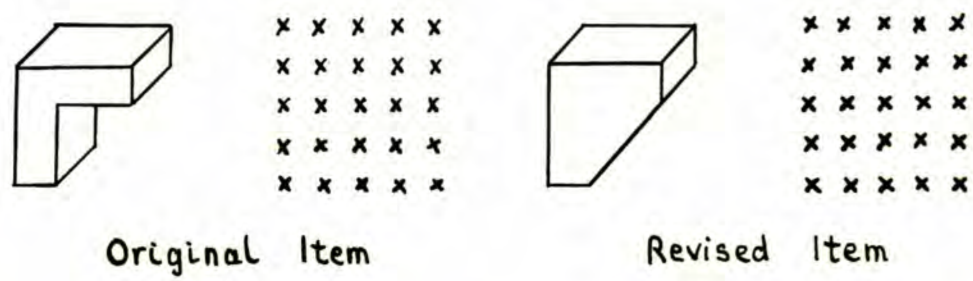


FIG. 16. REVISION OF ITEM F.16.

Test I required only slight alteration at one or two points where the 'response' diagrams might be misleading in their proportions, but in Test L one item, No. 11, required complete replacement and No. 13, simplification. The replaced item was again the only one which was not drawn as though lying on a horizontal surface. (Fig. 17)

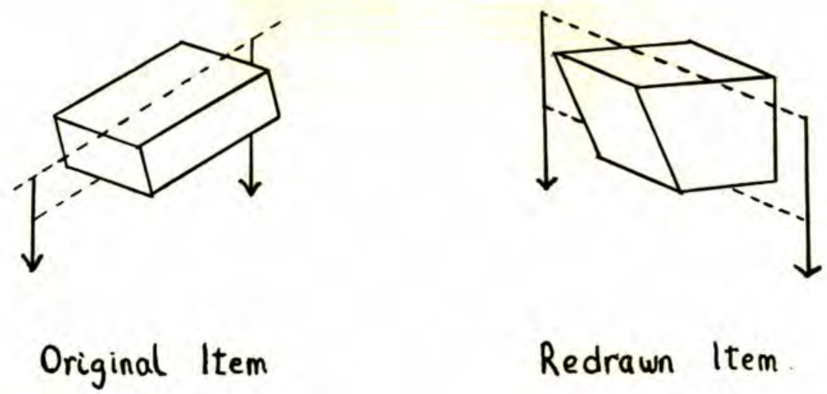


Fig. 17. REVISION OF ITEM L.11.

Finally, in Test M, no alteration was made.

After these revisions had been made the six tests were arranged in order of increasing difficulty except that Test A which requires only one page for its 20 items was placed last to form the back page of the booklet. Each test was preceded by a page of instruction and practice items, in substance those used in the Gateshead try-out with slight revisions made in the text where experience in the try-out indicated possible improvements. The revised form of the test thereby became:-

Test 1,	Three Sections (formerly I)	10 items
Test 2,	Plan Views (formerly B)	10 items
Test 3,	Paper Folding (formerly M)	20 items
Test 4,	Vertical Sections (formerly L)	20 items
Test 5,	3-D Networks (formerly F)	20 items
Test 6,	Counting Faces (formerly A)	20 items

The revised draft was subjected to an investigation to find out how far it conforms to the requirements regarding Rectangular distribution discussed above. For this investigation two distributions were plotted, that of the Facility Indices and that of the scores obtained by the 103 pupils of the First Try Out when the scripts were rescored for the selected tests only. Both these distributions, of course, include the unsatisfactory items of the selected tests, but nevertheless are useful pointers to the ultimate success of the test.

The Facility Indices of the 100 items ranged from 6% to 96% around a mean of 61.75 with the distribution shown in Figure 18. Of the 13 indices not reaching the level of 35% set in Chapter V, eight are of items which were redrawn, four are between 30% and 35% and the items retained, and the remaining index of 22% refers to the last item in Test 5, the most difficult item of the most difficult test. The mean of the Facilities is high since the Try Out Group was of an average age fourteen months higher than the age group for whom the test is intended.

The rescored scripts of the 103 pupils show a distribution (Fig. 19) which approaches the rectangular although skewed negatively. The scores range from 16 to 96 about a mean of 61.81 and with a standard deviation of 10.105.

FIG. 18. DISTRIBUTION OF FACILITIES - SELECTED ITEMS.

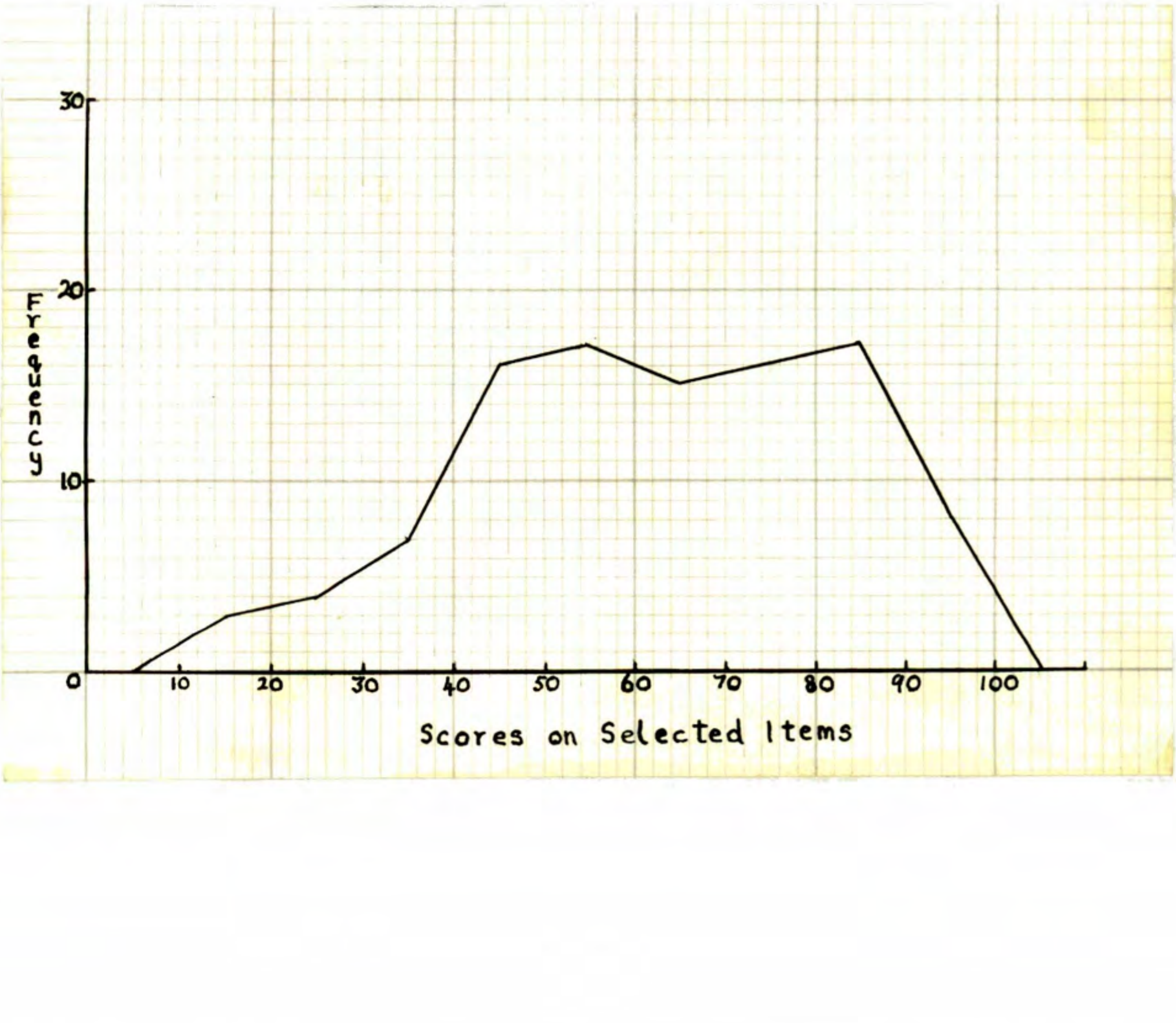


FIG. 19. DISTRIBUTION OF SCORES ON SCRIPTS RESCORED FOR SELECTED ITEMS ONLY (N=103).

The skewness and the high average can again be accounted for by the age of the Try-Out Group.

As a further confirmation of the development of the test along the right lines the scores obtained by the 11 boys and 13 girls from grammar schools were compared. The means for the two groups:-

Boys 87.64 Girls 78.92

show a difference of 8.72 which when subjected to the "t" test for small samples as described by Chambers (1952)[1] can be shown to be significant at the 1% level of confidence. This significant difference is a fair indication that the test is in fact measuring spatial ability.

This revised draft was printed by King's College Printing Section, Newcastle upon Tyne, from plates prepared photographically from the original drawings of the author. A copy of the revised test is bound into the present volume as Appendix C. Since this test was to be administered to a complete age group involving a large number of schools, complete detailed instructions were drawn up in a manual reprinted in Appendix B. This manual is based on the instructions used in the First Try-Out and includes time limits for each test also based on the experiences obtained with the Gateshead sample.

---

1. Chambers, E. G. (1952) Statistical Calculation for Beginners.

CHAPTER VII.

THE CARLISLE SAMPLE AND RESULTING STATISTICS.

To standardise a test it is necessary to administer it to the whole of the required age-group in any given area. On November 12th, 1958, the test in its revised form was taken by the entire 10 year old age-group of the City of Carlisle as part of the first stage of the selection procedure of that Authority. In all there were 1,045 pupils whose ages were within the range 10 years 1 month to 11 years 2 months, with an average age of 10 years 8 months; the ages calculated in completed months on the day of the examination. This total was made up of 537 boys and 508 girls.

The scripts of these pupils were all marked and checked by the author and his wife as in the Gateshead Try-Out. The scores obtained range from 1 to 99 about a mean of 42.1 and with a Standard Deviation of 20.745. The distribution of these scores is shown in Fig. 20, where the distribution for Boys and Girls are shown separately and in Fig. 21, where they are combined. From the performance of the Carlisle pupils a number of statistics and items of information were obtained.

Score-Age Distribution.

For a test to be of practical use it is advisable that the score of any pupil should be expressed by a number which

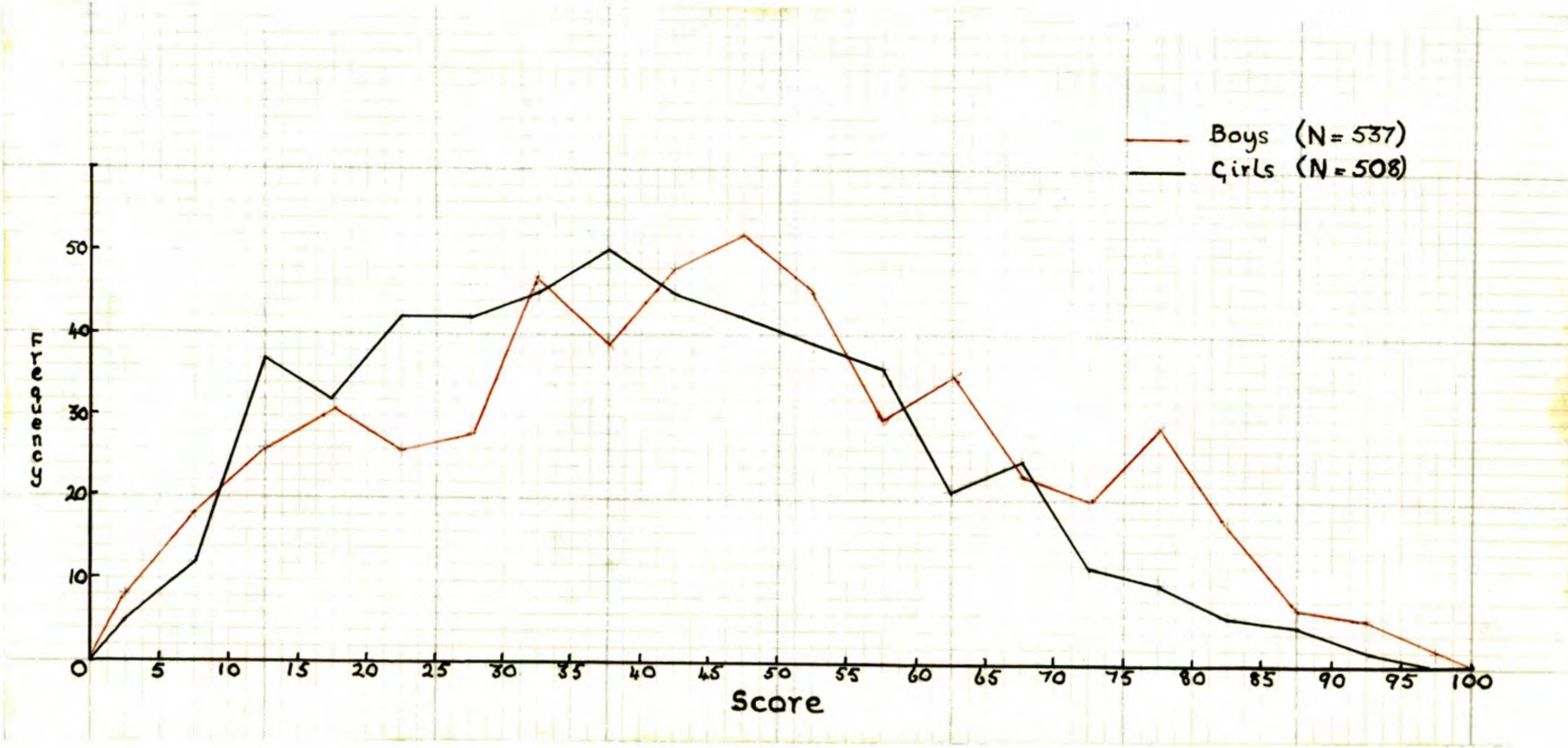


Fig. 20. CARLISLE SAMPLE. DISTRIBUTION OF RAW SCORES.

makes due allowance for age, the pupils position being assessed by comparison with a representative sample of children of exactly the same age. This can be done by constructing a Conversion Table from which a child's standardised score can be obtained if his raw score and age is known. The standardised scores from such a table are so arranged that they have a mean of 100 and a standard deviation of 15, and in this respect are similar to Intelligence Quotients.

A Conversion Table is constructed by classifying all the scores according to each month of age and then submitting the data to certain statistical procedures. To provide the necessary data the scores of the 1,045 pupils were recorded on Score-Age distribution sheets, classified according to ages in vertical columns and according to scores in horizontal rows. The two distributions, for boys and for girls are shown in Tables Xa and Xb respectively. These data were passed to the National Foundation for Educational Research for the necessary statistical treatment.

<u>Difference of Means.</u>

From the entire sample, a measure of the difference in mean scores for boys and for girls was obtained, to confirm the earlier finding that there is for this test, as for other Spatial Tests, a significant difference between

SCORE	10.1	10.2	10.3	10.4	10.5	10.6	AGE 10.7	10.8	10.9	10.10	10.11	11.0	11.1	11.2	TOTALS
95 - 100									1	1					2
90 - 94			2						1	1	1		1		6
85 - 89		1		2	1			1	1					1	7
80 - 84			3		1	2	3		1	1	1	1	4		17
75 - 79		3	2	1	3	3	1	2	3		1	7	1	2	29
70 - 74		1	1		1	2	4	2	2	3		1	2	1	20
65 - 69				6	1	2	1	2	5	1	2		3		23
60 - 64		2	4	1	1	1	2	3	7	5		6	3		35
55 - 59	1	1	1	2	2	1		4	4	4	1	1	6	2	30
50 - 54		1	5	5		4	3	5	4	6	3	3	4	2	45
45 - 49		1	5	3	2	9	6	2	3	2	3	8	7	1	52
40 - 44	1	3	6	3	3	6	3	2	2	5	6	5	1	2	48
35 - 39			3	3	3		6	3	4	5	5	1	6		39
30 - 34		3	2	5	4	4	2	3	4	7		10	3		47
25 - 29		1	3	4	3	3	4	1	2	4	1	2			28
20 - 24			2	3	2	3	2	4	2		1	3	4		26
15 - 19		2	3	5	2	4	2	2	2	2	3	3	1		31
10 - 14		1	5	3	1	2	2	1	3	1	1	3	2	1	26
5 - 9		1	2	3	2	2	3		1	1		1	2		18
0 - 4			2	2	1		2					1			8
TOTALS	2	21	51	51	33	48	46	37	52	49	29	56	50	12	537

TABLE Xa. SCORE-AGE DISTRIBUTION - BOYS

SCORE	10.1	10.2	10.3	10.4	10.5	10.6	10.7	10.8	10.9	10.10	10.11	11.0	11.1	11.2	TOTALS
							AGE								
95 - 100															
90 - 94								1		1					2
85 - 89			1				1				1	2			5
80 - 84				1				1			2		1	1	6
75 - 79		1						2	1			2	3	1	10
70 - 74			2	1					1	2	4	1	1		12
65 - 89			2	2	2	3	1	2	1	2	4	2	2	2	25
60 - 64			3		2	2	1	2	3	1		3	4		21
55 - 59		4	1	1	3		7	5	4	5	2		1	3	36
50 - 54		3	1	3	4	3	3	3	5	5		4	3	2	39
45 - 49			4	2	2	3	4	5	3	8	5	5	1		42
40 - 44		1	2	2	2	5	7	4	6	4		2	8	2	45
35 - 39		4	5	3	5	5		3	7	3	2	4	7	2	50
30 - 34		1	4	4	3	5	5	2	4	6	5	2	3	1	45
25 - 29		4	5	3	3	2	4	3	5	1	2	5	4	1	42
20 - 24		3	3	6	6	6		5	2	2	2	3	2	2	42
15 - 19		1	3	3	2	4	4	4	3	1	1	1	1	4	32
10 - 14			3	4	5	4	2	3	4		6	4	1	1	37
5 - 9		1	2	1	2	2			1			3			12
0 - 4		1	2	1			1								5
TOTALS		24	43	37	41	44	40	45	50	41	36	43	42	22	508

TABLE Xb. SCORE-AGE DISTRIBUTION - GIRLS the performance of the two sexes. The relevant figures are:-

Boys (N = 537) Mean = 44.504 S.D. = 21.725

Girls (N = 508) Mean = 39.56 S.D. = 19.335

Difference of Means = 4.944.

Applying the "t" test to this difference a Critical Ratio of ~~2.873~~ 3.89 is obtained which shows that the difference is significant at the 1% level.

Item Analysis.

The item analysis carried out on the two hundred items of the first draft was based on a relatively small sample, 103 pupils, and one whose average age was considerably higher than the range proposed in Chapter One. For these reasons, and to obtain figures for those items which had been radically revised, indices of facility and discrimination were calculated, from the Carlisle scores, for the one hundred items of the Revised Draft. These indices were obtained from a smaller sample drawn from the 1,045 scripts. This smaller sample totalled 185, a number which was decided upon because this would provide upper and lower 27% groups of 50 scripts and so facilitate the calculation of percentages. To retain the same proportion of Boys + Girls in the sample as existed in the total population 95 scripts of boys and 90 of girls were selected. These scripts were selected by random numbers, using the tables

Fig. 21. FREQUENCY POLYGONS OF CARLISLE POPULATION SCORES AND SAMPLE SCORES

provided by Lindquist (1940)[1].

The sample which was extracted closely paralleled the total population as the figures in Table XI and the frequency polygons in Fig. 21 show.

	POPULATION	SAMPLE
Boys	537	95
Girls	508	90
Mean Age	10 yrs.8 mth.	10 yrs. 8 mth.
Range of Scores	1 - 99	2 - 91
Mean of Scores	42.1	42.57
S.D. of Scores	20.745	20.59

TABLE XI. Comparison of Carlisle Population and Analysis Sample.

After the scripts for the sample had been extracted they were arranged in order of total score and the top 50 and bottom 50 taken to form the upper 27% and lower 27% respectively. The upper group consisted of all those scoring 55 or higher and the lower group included all those with scores less than 30. The responses for each item on these one hundred scripts were then entered on score sheets as described in Chapter Five. Again, as in the

1. Lindquist, E. F. (1940). Statistical Analysis in Educational Research.

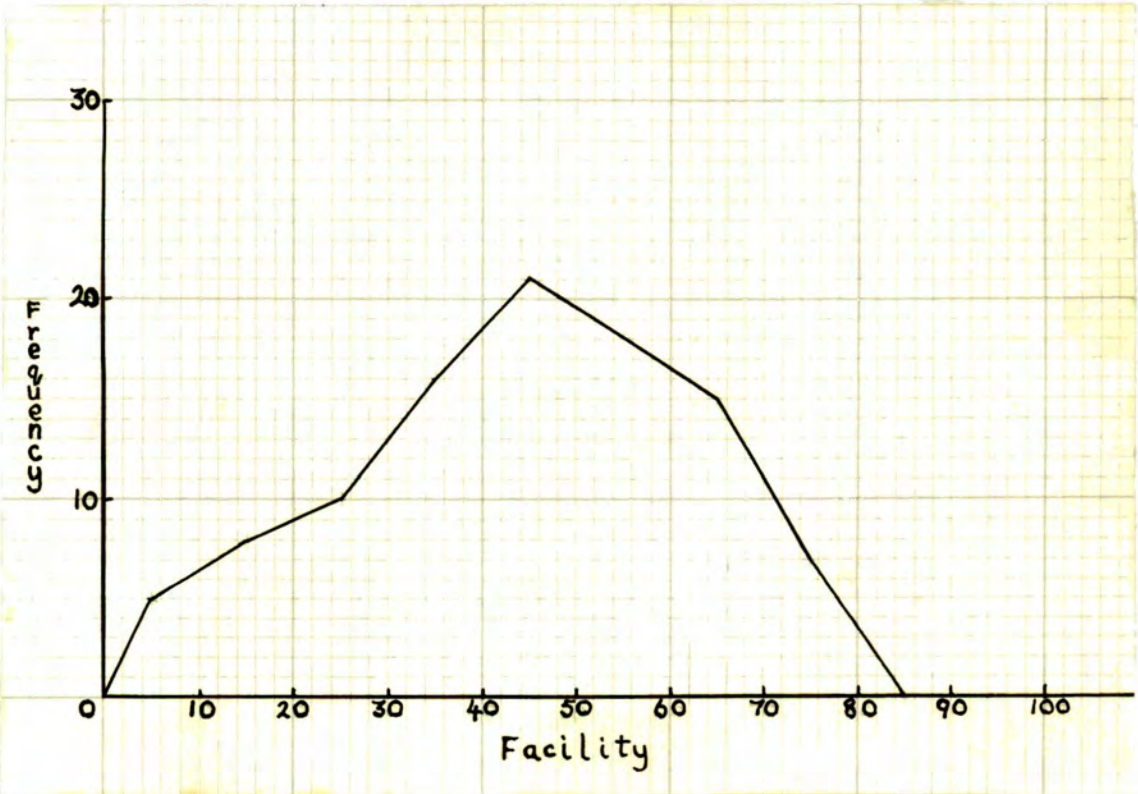


FIG. 22. DISTRIBUTION OF FACILITIES - REVISED DRAFT.

first analysis, the totals passing in each item in the Upper 27% and in the Lower 27% were recorded on cards and these figures used for the calculations.

The Facility Indices were obtained by adding the total right responses in the Upper group to the total right responses in the Lower group. This gave a figure which can be regarded as a percentage, since the two groups together contain 100 scripts. The indices so obtained range from 4 - 77, with a mean of 42.85 and the distribution shown in Fig. 22. These indices are listed in Table XII, together with the Discrimination Indices.

The Discrimination figures were again obtained by Flanagan's Method, where the percentage passing in the Upper 27% and the percentage passing in the Lower 27% are used to read off from tables an estimated Product Moment Coefficient. The required percentages were calculated rapidly since all that this entailed was the doubling of the figures on the record cards. The resulting coefficients are shown in Table XII and their distribution within each test in Fig. 23. They range from 0.21 to 0.89, with a mean discrimination of 0.60.

These new indices, of Facility and of Discrimination, show no grounds for radical alteration of the test, they confirm the ascending order of difficulty of the sub-tests,

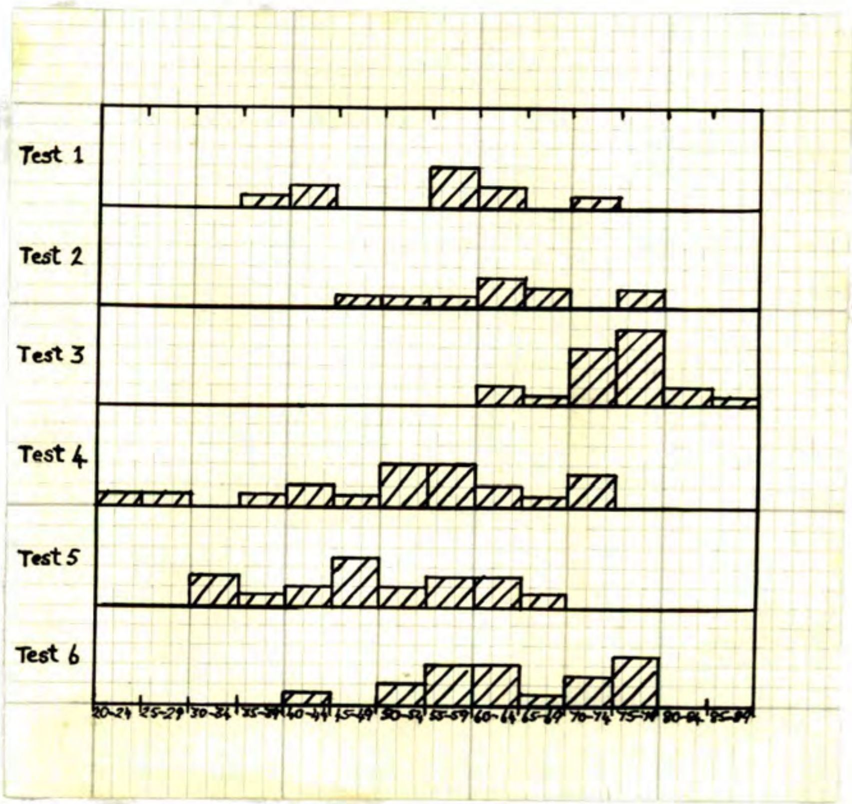


FIG. 23. HISTOGRAMS SHOWING DISTRIBUTION OF DISCRIMINATION INDICES - REVISED DRAFT.

ITEM	F	D	ITEM	F	D	ITEM	F	D	ITEM	F	D
Three 1	37	0.58	26	59	0.76	51	40	0.71	76	4	0.35
Sect- 2	47	0.58	27	62	0.73	52	36	0.61	77	7	0.46
ions. 3	42	0.37	28	24	0.72	53	64	0.52	78	6	0.30
4	73	0.61	29	35	0.70	54	59	0.44	79	11	0.46
5	47	0.42	30	66	0.75	55	31	0.50	80	6	0.43
6	57	0.43	31	34	0.69	56	61	0.40	Coun- 1a	51	0.79
7	57	0.74	32	37	0.78	57	40	0.50	ting 2a	71	0.76
8	48	0.56	33	52	0.74	58	46	0.71	Faces 2b	71	0.76
9	52	0.63	34	43	0.70	59	40	0.47	2c	70	0.77
10	43	0.59	35	40	0.80	60	39	0.65	3a	66	0.59
Plan 11	68	0.78	36	53	0.76	3-D 61	24	0.57	3b	73	0.69
Views12	64	0.77	37	58	0.63	Net- 62	34	0.54	3c	27	0.43
13	69	0.60	38	32	0.78	works63	25	0.58	3d	53	0.50
14	43	0.47	39	44	0.64	64	29	0.63	4a	64	0.57
15	60	0.66	40	46	0.75	65	14	0.61	4b	41	0.64
16	25	0.52	Vert- 41	77	0.36	66	33	0.58	5a	73	0.74
17	53	0.69	ical 42	54	0.25	67	16	0.44	5b	52	0.63
18	62	0.60	Sect- 43	30	0.33	68	17	0.65	6a	54	0.60
19	52	0.56	44	43	0.59	69	11	0.33	6b	61	0.57
20	40	0.63	45	32	0.56	70	32	0.47	6c	37	0.63
Paper21	62	0.79	46	23	0.64	71	15	0.62	7a	21	0.53
Fold-22	37	0.82	47	43	0.59	72	33	0.34	7b	54	0.56
ing 23	50	0.77	48	31	0.73	73	10	0.53	7c	64	0.77
24	41	0.72	49	47	0.58	74	17	0.46	8a	28	0.70
25	46	0.89	50	58	0.21	75	8	0.49	8b	68	0.73

TABLE XII. REVISED DRAFT. FACILITY INDICES (F) AND DISCRIMINATION INDICES (D) with the proviso that Test 6, Counting Faces, is placed last for practical reasons rather than difficulty, and show that Test 5, is more than twice as difficult as any other sub-test. Comparison of the indices obtained for revised and re-drawn items with the indices of the items they have replaced shows that a considerable improvement has resulted (Table XIII).

Original Item			Revised Item		
No.	F	D	No.	F	D
B 3	32%	0.15	13	69%	0.60
I 10	69%	0.29	10	43%	0.59
L 11	28%	-0.04	51	40%	0.71
L 13	83%	0.27	53	64%	0.52

TABLE XIII. To Show Improvement in Revised Items.

This improvement is even pronounced when viewed in the knowledge that the Facility Indices in general are about two thirds of those obtained in the Gateshead Try Out, a drop which can be explained by the big difference in average age of the two groups, a difference of twenty months. There is not in Test 5 (3-D Networks) so definite an improvement in revised items, even when allowance is made for the fact that the Facility Indices of unchanged items in this test are, on average, one half of those of the Gateshead analysis. From the figures (Table XIV) one can conclude no more than that, of the new items, three show improvement and the others are no worse than those they replace.

Original Item			Revised Item		
No.	F	D	No.	F	D
F 4	19%	0.71	72	33%	0.34
F 5	31%	0.59	78	6%	0.30
F 7	14%	0.68	76	4%	0.35
F 10	22%	0.68	73	10%	0.53
F 16	14%	0.66	74	17%	0.46
F 17	17%	0.60	75	8%	0.49
F 20	6%	0.51	77	7%	0.46

TABLE XIV. Comparison of Original and Revised Items of Test 5, 3-D Networks.

Reliability Coefficient.

Reliability is the degree of accuracy with which a test measures what it does measure. A test with one hundred per cent reliability would give exactly the same scores if administered a second time to a given population. It is not always possible to obtain figures for such a re-test but there are two other techniques in common use for obtaining a reliability coefficient. One is to correlate scores on a test with the scores obtained, by the same group, on an equivalent test, the second is to correlate the scores obtained on one half of the test with the scores from the other half. This Split-Half method has been used here.

The total score obtained on the odd numbered items was taken to form one half of the test and the score on the even numbers taken as a second, parallel form of the test. The 185 scripts of the item analysis sample were scored on this basis and the resulting scores plotted on the scattergram reproduced in Fig. 24. From the distribution on this scattergram a product moment coefficient of correlation was calculated with a result of

$$r = \cancel{0.8861}\ 0.9315$$

This is, of course, a correlation between two tests each 50 items in length. A longer test would be expected to show a higher correlation. This correlation can be predicted by applying the Spearman-Brown Prophecy Formula:-

$$R = \frac{nr}{1 + (n-1)r}$$

where R = expected coefficient

r = obtained coefficient

n = multiple by which test is to be lengthened.

In this case the formula becomes

$$R = \frac{2r}{1 + r}$$

ODDS

EVENS	0-2	3-5	6-8	9-11	12-14	15-17	18-20	21-23	24-26	27-29	30-32	33-35	36-38	39-41	42-44	45-47
45-47															1	
42-44														1	1	1
39-41												1	3	3	2	
36-38											2	2	4	3		
33-35										1	1	3	2			
30-32								1	2	1	3	2				
27-29								1	5	7	1					
24-26						1	1	5	4	2	1	1				
21-23						1	8	6	2	6	2					
18-20					1	6	6	5	2							
15-17			1	1	5	3	4	6	1							
12-14			2	3	2	3	2	1	1							
9-11		1	4	4	4											
6-8		2	4	2	1											
3-5	2	4	5	1												
0-2	1	1														

FIG. 24. SCATTERGRAM OF SCORES ON TOTAL ODD ITEMS (50) AND TOTAL EVEN ITEMS (50)

and gives a coefficient of

$$R = 0.9642.$$

This means that, taking note of the fact that a Split Half coefficient tends to be a little higher than a Re-test coefficient, the reliability is high enough for the test to be of practical use.

CHAPTER VIII

Investigation of the Test

For a test to be truly 'spatial' in character it should possess those characteristics common to recognised spatial tests which have been revealed by the investigations detailed in Chapter I. Of these properties the two most important are that the test should be a measure of the K-factor and that it should correlate positively with school subjects of a technical/practical/scientific bias. Investigations along these two lines have been carried out with the present test.

During the summer term of 1960 the following tests, in the order shown, were administered to all the boys in the 1st and 2nd forms of the Matthew Arnold Secondary School, Oxford.

(1) Spatial Test 1, N.F.E.R.

(2) Verbal Test (Adv.) 1, N.F.E.R.

(3) "Experimental" Spatial Test

Omitting those absent for one or other test, scores were obtained for 85 boys whose ages ranged from 11.11 to 13.10 about a mean of 12.11. Since the upper age range is beyond the furthest limit of the conversion tables for the new spatial test all scores were expressed as raw scores.

Product moment correlation coefficients between each test and the other two tests were calculated and are shown in Table XV.

	Verbal	NFER Sp. 1	Spatial
Spatial	0.5789	0.7811	-
NFER Sp. 1	0.4793	-	0.7811
Verbal	-	0.4793	0.5789

Table XV. Inter-Test Correlations. (N = 85)

These correlations all of which are highly significant, show that the experimental test has more in common with the established spatial test, which is known to have a high K-loading, than with the verbal test.

The scores on the three tests were also correlated with school subject marks but before detailed description of these, reference must be made to an earlier, limited investigation carried out at Gateshead Grammar School in the spring of 1959. The investigation is described as limited from both a quantitative and a qualitative point of view. The number of boys involved was but 20 and this constituted a 'technical stream' formed, in the main, from those members of forms 3B and 3C, age range 13.6 to 14.6, who had not been "creamed off" for more academic scientific or literary/linguistic courses. For this small group school subject marks taken from an internal half-yearly examination were correlated with scores on the experimental spatial test and on Verbal Test (Adv.) 1. The resulting product moment correlations are shown in Table XVI.

	Woodwork	Eng. Draw.	Mechanics	English
Verbal	-0.3054	-0.1177	-0.1411	+0.1254
Spatial	+0.1393	+0.1056	-0.4262	+0.2107

Correlation between Spatial and Verbal, r = -0.1742

Table XVI. Product Moment Coefficients - Grammar School (N = 20).

No conclusion can be drawn from these figures but they are reported for comparison with the Secondary Modern results.

The results obtained in the Secondary Modern School were also based on examination marks obtained in the annual internal examinations held by the school. These examinations were marked 'within forms' and therefore marks could not be pooled to obtain estimates for the whole of the 2nd form and the whole of the 1st form. Product moment coefficients were therefore calculated within each form and, since the A and B streams followed the same syllabus and took the same examination, although marked separately, the coefficients for 2A and 2B, and for 1A and 1B, were combined by transforming r to z, pooling the z values and then re-converting.[1] Forms

1. See Tippett, L.H.C. (1948). The Methods of Statistics. Moroney, M.J. (1956). Facts from Figures.

2A and 2B. N = 29.

	Metal Wk.	Eng. Draw.	Woodwork	Science	Maths.	English
Verbal	+0.0809	-	-	+0.4203	+0.5027	+0.4930
Spatial	+0.4774	-	-	+0.2676	+0.1240	+0.4643
NF. Sp.1.	+0.5235	-	-	+0.1823	+0.2968	+0.5419

2C. N = 14.

Verbal	-	-0.1374	+0.3465	+0.4365	+0.6822	-0.0615
Spatial	-	+0.0429	+0.1691	+0.3300	+0.1710	+0.3643
NF. Sp.1.	-	+0.2041	+0.1930	+0.3771	+0.3362	+0.2651

1A and 1B. N = 26.

Verbal	-	-	+0.4682	+0.3216	+0.4317	+0.6748
Spatial	-	-	+0.5132	+0.2973	+0.3070	+0.0310
NF. Sp.1.	-	-	+0.2790	+0.2879	+0.3790	+0.1370

1C. N = 15.

Verbal	-	-	-0.0463	+0.1290	-0.2396	+0.4113
Spatial	-	-	-0.0759	+0.0864	-0.1736	+0.2333
NF. Sp. 1.	-	-	-0.0457	+0.2395	-0.0883	+0.5685

Table XVII. Product Moment Correlations - Modern School.

2C and 1C followed different syllabuses with different examinations and the correlations for these forms are therefore quoted separately. All these coefficients are set out in Table XVII.

Set out in this way the pattern of correlations between the Spatial Test and the school subjects can be compared to that existing between N.F.E.R. Spatial 1 and the subjects and it can be seen also that, in general, these figures show the Spatial Test correlating positively with performance in school, the two exceptions (I.C. Woodwork and Maths) occurring where Verbal (Adv.)1 and N.F.E.R. Sp. 1. correlations are also negative. However it is probably more instructive to examine the figures by 'Subjects' rather than by 'Forms'.

Metalwork

The correlations for this subject were obtained from forms 2A and 2B in the Secondary Modern School.

(a)

	N	Verbal	Spatial	NF.Sp.1.
Sec. Mod. 2 A.B.	29	+0.0809	+0.4774	+0.5235

(b)

Sec. Mod. 2A	16	+0.0851	+0.5081	+0.5962
" " 2B	13	+0.0755	+0.4355	+0.4163

Table XVIII. Metalwork and Test Scores.

The combined correlations shown in Table XVIII (a) were obtained from the correlations for the separate forms 2A and 2B, which are shown in XVIII (b), and the substantial differences between the verbal/metalwork correlations and the spatial/metalwork correlations suggest that, in this instance, the spatial test is preferable to the verbal test as a predictor of success in metalwork.

Engineering Drawing.

Correlations from two sources, the 3rd form technical stream of the grammar school and the 2nd form C-stream of the secondary modern school show the same pattern, small negative correlations for metalwork marks and verbal test scores and small positive correlations with the spatial scores but in neither case are the correlations large enough to allow any conclusion to be drawn.

	N	Verbal	Spatial	NF.Sp.1.
Grammar (Tech.)	20	-0.1177	+0.1056	-
Sec. Mod. 2C	14	-0.1374	+0.0429	+0.2041

Table XIX. Engineering Drawing and Test Scores.

Woodwork

From information obtained from four separate samples of pupils no definite pattern emerges.

	N	Verbal	Spatial	NF.Sp. 1.
Grammar (Tech.)	20	-0.3054	+0.1393	-
Sec. Mod. 2C	14	+0.3465	+0.1691	+0.1930
" " 1 A.B.	26	+0.4682	+0.5123	+0.2790
" " 1C	15	-0.0463	-0.0759	-0.0457

Table XX. Woodwork and Test Scores.

Mathematics

Here there is a consistent picture with the verbal test showing greater correlation with mathematics than does either of the two spatial tests. For this secondary modern sample the most probable conclusion is that the verbal test is a better predictor of success in mathematics than is the spatial test.

	N	Verbal	Spatial	NF.Sp. 1.
Sec. Mod. 2 A.B.	29	+0.5027	+0.1240	+0.2968
" " 2C	14	+0.6822	+0.1710	+0.3362
" " 1 A.B.	26	+0.4317	+0.3070	+0.3790
" " 1C	15	-0.2396	-0.1736	-0.0883

Table XXI. Mathematics and Test Scores.

Science

The correlations with science marks, as for mathematics, are higher for the verbal test than for the spatial tests, although the differences between the verbal/science and spatial/science correlations are not so marked as the corresponding differences for mathematics.

	N	Verbal	Spatial	NF.Sp. 1.
Sec. Mod. 2 A.B.	29	+0.4203	+0.2676	+0.1823
" " 2C	14	+0.4365	+0.3300	+0.3771
" " 1 A.B.	26	+0.3216	+0.2973	+0.2879
" " 1C	15	+0.1290	+0.0864	+0.2395

Table XXII. Science and Test Scores.

English

English was included in the list of subjects with which the tests were correlated in the expectation that the verbal test would show clear advantages as a predictor of success in this subject. The results in the table show this not to be so.

	N	Verbal	Spatial	NF.Sp. 1.
Grammar (Tech.)	20	+0.1254	+0.2107	-
Sec. Mod. 2 A.B.	29	+0.4930	+0.4643	+0.5419
" " 2C	14	-0.0615	+0.3643	+0.2651
" " 1 A.B.	26	+0.6748	+0.0310	-0.1370
" " 1C	15	+0.4113	+0.2333	+0.5685

Table XXIII. English and Test Scores.

One possible explanation of this confused pattern is that the English marks are taken from school examinations, a type of measure which has time and again been shown to be unreliable.[1]

This factor of unreliability in the marks used has, of course, to be borne in mind with each of the subjects here discussed but the correlations reported suggest that both the experimental spatial test and N.F.E.R. Spatial 1 are better measures of the ability involved in Metalwork than is Verbal Test (Adv.) 1, while the verbal test is of greater value than either of the spatial tests as a predictor of success in Mathematics.

---

1. See for example, Vernon, P.E. and Parry, J.B. (1949), Personal Selection in the British Forces, and Hartog, P.J. and Rhodes, E.C. (1936), The Marks of Examiners.

CHAPTER IX

Retrospect and Prospect

The aims stated in the title of this work and expanded in Chapters I and II may be summarised as follows.

1. The construction of a test using diagrammatic material based on projections and sections of solid objects.
2. The test so constructed to be standardised for use with pupils in the approximate age range of $10\frac{1}{2}$ to $11\frac{1}{2}$ years.
3. The test to be an instrument for measuring spatial ability.
4. The test to be of value in predicting performance in schools.

How far have these aims been accomplished?

The Construction of the Test

The final form of the test, as it appears in Appendix C, shows that a test has in fact been constructed which utilises in all it sub-tests drawings of solid objects; the one sub-test of the First Draft which did not conform to this having been eliminated in the analysis following the First Try Out.

That projections and sections are used is not as obvious since these terms have to a large extent been avoided in the instructions but Test 1, Three Sections, and Test 4, Vertical Sections, do employ a straightforward use of the section concept. Test 3, Paper Folding, and Test 5, 3-D Networks, can be regarded as requiring the projection of three dimensional structures on to plane surfaces, while Test 2, Plan Views, and Test 6, Counting Faces, require the identification of faces other than those presented for direct viewing, these faces being of the same shape as a section or projection taken parallel to the face concerned.

Age-Range and Standardisation

Two trials of the test were carried out during the construction, one using the First Draft with a sample of pupils having an average age of 12 years 4 months and the second using the Revised Draft with a complete age group averaging 10 years 8 months. The mean scores obtained were; for the First Draft re-scored for selected items: 61.81; for the Revised Draft: 42.1. It can therefore be seen that of the two groups, the first has a higher average age and the second a slightly lower average age than that proposed, and also that the mean scores show correspondingly higher and lower levels than that usually set in mental tests. The conclusion can be drawn that the difficulty level is such as would make the test suitable for pupils intermediate in age to the two experimental groups, that is of an age of the order of 11 years. This in fact would have been the age of the Carlisle group if the try out had been inserted into the second part of the battery in February/March rather than in the first part taken in November.

The standardised instructions and the conversion tables based on this sample therefore equip the test for use with pupils nearing the end of their primary school careers, that is at the '11+ examination' age, and the test has in fact been published for this purpose by the National Foundation for Educational Research as Spatial Test 3.

Measure of Spatial Ability

In the absence of a factor analysis the suggestion that the test is 'spatial' in character rests upon two pieces of evidence.

Firstly there is the evidence provided by the difference between the mean scores for boys and girls. At two stages during the development of the test this difference was calculated. In both cases the difference proved to be significant at the 1% level of confidence.

The second calculation can be particularly relied upon since (a) it is based on the scores obtained with a complete age group, (b) scores obtained on the Moray House Verbal Reasoning Test 59 by the same group show a significant sex-difference in the opposite direction, that is mean score for girls significantly greater than mean score for boys. This difference in favour of girls was also found with N.F.E.R. Verbal Test 8A, although in this case the difference is barely significant at 5% level. The relevant figures are collected in Table XXIV, where all the figures are based on standardised scores in contrast to Chapter VII where the difference for the spatial test was based on raw scores.

Test	Mean (Boys)	Mean (Girls)	Difference	Crit. Ratio
'Spatial' Test	101.036	97.693	3.343	3.76
MH Verb. Reas. 59	101.814	105.936	-4.122	4.42
NFER Verbal 8A	102.335	103.873	-1.538	1.97

Table XXIV. Differences in Mean Scores - Carlisle Sample.

(Note - a minus sign in column 4 indicates a difference in favour of girls).

These figures would seem to indicate that the difference in favour of the boys in the spatial test is a feature of that test and not due to some property of the sample population, in which case the sex difference can be interpreted as an indication that the test involves spatial ability.

Secondly the inter-test correlations obtained from the results of the Oxford investigation with 85 boys show that the test has more in common with N.F.E.R. Spatial Test 1, than with Verbal Test (Adv.) 1. Spatial Test 1, has, to quote the Manual supplied with it, "a high loading in the spatial factor (k), as well as in the general factor (g)", and for some years has been used for selection and has been shown to be of some value in this context.[1]

Predictive Value

Chapter VIII contains details of an investigation carried out in a Secondary Modern School. It must be emphasised that this is the first stage in a much longer investigation in which it is hoped to follow up the performance of the pupils in the various subjects throughout their school careers and eventually to correlate the test scores obtained in the 1st and 2nd forms with subsequent degree of success in external examinations taken during the last year at school.

Nevertheless, from what has already been done using internal examination marks as the criterion a few tentative suggestions about the experimental test can be made. When the correlations between the test and subject marks are viewed in isolation it is seen that the majority of these correlations are positive and of these that between the test and metalwork (2A, B, N = 29) is significant at the 1% level, that between the test and woodwork (1A, B, n= 26) is significant at the same level and that between the test and English (2A, B, N = 29) is significant at the 5% level of confidence.

1. Smith, I. Macfarlane, (1960). The Validity of Tests of Spatial Ability as Predictors of success on Technical Courses. Brit. J. Ed. Psych. XXX, Pt. 2.

When compared with the correlations obtained between a verbal test and the subject marks it is found that, with this secondary modern school population at this age range, the Spatial test would seem to be the better predictor of success in Metalwork, while the verbal test is the better predictor of performance in Mathematics.

Future Developments

The work described above is but the early stages in the development of a test. For a test to be of real value the collection of data concerning it must continue for a long time. Further investigation can be along two lines; extended "predictive" trials and factor analysis.

It has already been noted that it is hoped to follow up the careers of the 85 modern school boys and to measure ultimate success by the criterion of external examinations. Similarly, and on a much wider scale, the entire school population of the age-range tested in Carlisle can be followed up in the variety of schools and courses to which they have been allocated. Both of these investigations require a minimum of four to five years from the initial testing for their completion and in the case of pupils taking examinations of higher levels the period is nearer seven to eight years.

A factor analysis would provide the best evidence of the degree to which the test measures spatial ability and the possibility also arises, since the test involves three-dimensional material, of adding to the information available as to the relationship between two-dimmansional and three-dimensional test items. There is at present no great agreement on these points. Renshaw's (1950)[1.] investigation showed slight indications that two-dimensional tests were more closely linked with 'K' than were three-dimensional tests. Macfarlane Smith (1954)[2.] writing of Spatial Test 1 states that his results "do not support the view that tests consisting of drawings of 3-dimensional objects measure a different factor from tests containing two-dimensional figures", while Pichot (1956)[3.] writes, "Il est possible que ce facteur (S) inclue un sous-facteur portant sur les configurations tri-dimensionelles."

This is in the future, for the present it is suggested that the four-fold aim stated above has been accomplished to the following extent.

---

1. Renshaw, T. (1950). A Factorial Study of Two and Three Dimensional Space Tests. Ph.D. Thesis, Edinburgh.

2. Smith, I. Macfarlane, (1954). The Development of a Spatial Test. Durham Research Review 5.

3. Pichot, P. (1956). Les Tests Mentaux.

A test which fits the requirements as to subject-matter has been constructed. This test has been standardised and supplied with instructions and conversion tables and has reached publication stage. The evidence of sex difference in mean scores and of correlation with recognised tests is reasonable indication that the test involves spatial ability and the limited suggestion is made that the test can be of greater use than a verbal test in predicting performance in metalwork.

Appendix A.

Justification of Formula Used in Chapter V.

In Chapter V the following formula was used to find the Critical Ratio of the Discrimination index:-

$$C.R. = \frac{P_1 - P_2}{\sqrt{\frac{P_1Q_1}{N_1} + \frac{P_2Q_2}{N_2}}}$$

where P_1 = Proportion passing in Upper Group.

P_2 = Proportion passing in Lower Group.

Q_1 = Proportion failing in Upper Group

Q_2 = Proportion failing in Lower Group

N_1 = Number in Upper Group.

N_2 = Number in Lower Group.

It was there stated that this formula is a variation of:-

$$C.R. = \frac{\text{Observed Value}}{\text{Standard Error of Observed Value}}$$

These two formulae can be shown to be equivalent as follows

(i) The Discrimination Indices used are calculated on the percentage passes in the Upper and Lower Groups. Thus the numeration of the two formulae, $P_1 - P_2$ and "Observed Value", can be taken as equivalent.

It therefore remains to show that the denominator $\sqrt{\frac{P_1Q_1}{N_1} + \frac{P_2Q_2}{N_2}}$, of the formula is equivalent to the "Standard Error of the Observed Value".

(ii) The S.E. of a difference is usually obtained from the formula

$$S.E._{P_1 - P_2} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

where σ_1 = S.D. of Upper Group

σ_2 = S.D. of Lower Group

N_1 = Number in Upper Group

N_2 = Number in Lower Group

Therefore the problem becomes:-

to prove that

$$\sqrt{\frac{P_1Q_1}{N_1} + \frac{P_2Q_2}{N_2}} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

which in its simplest form becomes

<u>to prove that</u>

$$P_1Q_1 = \sigma_1^2$$

$$P_2Q_2 = \sigma_2^2$$

(iii) <u>Proof</u>

Let number of pupils in Upper Group be x

Let number of passes in Upper Group be y.

then proportion passing $= \frac{y}{x}$

and proportion failing $= \frac{x - y}{x}$

$$\therefore \; P_1Q_1 = \frac{y}{x} \times \frac{x - y}{x}$$

$$= \frac{xy - y^2}{x^2} \qquad \text{- - - - - - - - - (a)}$$

Using the arbitrary mean method to calculate the Standard Deviation, the formula is

$$\sigma = \sqrt{\frac{\sum d^2}{N} - \left(\frac{\sum d}{N}\right)^2}$$

where d = deviation from mean.

But all scores on any one item are either Pass or Fail, scored 1 or 0.

Taking arbitrary mean to be 1 the deviating scores are those which are 0 and therefore each deviation = 1.

$$\therefore \quad \sum d = x - y$$

$$\text{and} \quad \sum d^2 = x - y$$

substituting

$$\sigma_1 = \sqrt{\frac{x-y}{x} - \left(\frac{x-y}{x}\right)^2}$$

$$= \sqrt{\frac{x-y}{x} - \frac{x^2 + y^2 - 2xy}{x^2}}$$

$$= \sqrt{\frac{x^2 - xy - x^2 - y^2 + 2xy}{x^2}}$$

$$= \sqrt{\frac{xy - y^2}{x^2}}$$

therefore $\sigma_1^2 = \dfrac{xy - y^2}{x^2}$ - - - - - - - - - - - - - (b)

From results (a) and (b)

$$P_1Q_1 = \frac{xy - y^2}{x^2}$$

$$\sigma_1^2 = \frac{xy - y^2}{x^2}$$

$$\therefore \quad P_1Q_1 = \sigma_1^{\,2}$$

Similarly $\quad P_2Q_2 = \sigma_2^{\,2}$

Thus
$$\frac{P_1 - P_2}{\sqrt{\frac{P_1Q_1}{N_1} + \frac{P_2Q_2}{N_2}}} = \frac{\text{Observed Value}}{\text{S.E. of Observed Value}}$$

Appendix B.

Instructions for administration of the test (revised draft) as used in the Carlisle try-out and the validation experiments.

GENERAL DIRECTIONS.

1. It is ESSENTIAL that the procedure here outlined is followed exactly. No deviations, however slight, from the oral instructions are permissible. The greatest care must be taken to ensure that the six sub-tests are correctly timed. For this purpose it is necessary to use a watch with a seconds hand. ON NO ACCOUNT must a watch or clock without a seconds hand be used. If a stop watch is used, its accuracy should first be checked by comparing it with an ordinary watch with a seconds hand, as serious errors in timing sometime occur when stop- watches are used. It is desirable to have a spare watch in case of accidents.

2. The pupils write their answers in PENCIL. The supervisor should ensure that each child has two sharpened pencils before the test begins. If this is not possible, a supply of spare pencils should be kept at hand in case any pupil should break his pencil during the test. No materials other than pencils should be provided. Pens, rulers, erasers must NOT be used.

3. It is desirable that there be two invigilators to each room. One of these, the supervisor, should be res-

ponsible for the timing of the sub-tests. He should stand at the desk, facing the children, reading the instructions as here outlined, and keeping time with a watch. He should guard against having his attention distracted in any way whatever. The times allowed for each sub-test are short, and it is easy to over-run the allotted periods inadvertently.

The second invigilator should patrol the room quietly. He should be responsible for the prevention of copying and for ensuring that the children turn over the pages correctly, and that they follow the instructions given at the foot of each page.

4. Since copying is comparatively easy with a test of this type, the supervisor should arrange that the children sit at separate desks, and as far away from one another as possible. While the test is being worked, the second invigilator should see that they write the answers in the correct places and in the correct way. Thus, if he sees that a child is using a wrong METHOD of answering, for example, underlining instead of crossing out, he should correct the child by pointing to the words in the instructions at the top of the page. APART FROM THIS, NO ASSISTANCE WHATEVER SHOULD BE GIVEN DURING THE ACTUAL WORKING OF THE TEST. An opportunity for asking questions will be given at the end of each Practice Test. The supervisor

should answer these BRIEFLY IF THEY ARE RELEVANT. He should not allow himself to be drawn into a discussion, or to prolong the time taken to administer the test.

5. There are six sub-tests, each preceded by a practice test. The timing is as follows:-

Test 1	Pages 4 and 5	4 minutes
Test 2	Page 8	$3\frac{1}{2}$ minutes
Test 3	Pages 12 and 13	$6\frac{1}{2}$ minutes
Test 4	Pages 16 and 17	5 minutes
Test 5	Pages 20 and 21	16 minutes
Test 6	Page 24	4 minutes

The total working time, including the time required for the practice tests, is approximately one hour.

PROCEDURE.

1. The oral instructions should be read from this booklet of instructions, and NOT from a test booklet. THIS IS MOST IMPORTANT because the instructions contain answers to the practice tests and these are not given in the test booklets. Also, in a few instances, the oral instructions are slightly more detailed than the printed instructions in the test booklets.

2. Having arranged the desks so that the children are seated as far away from one another as possible, the supervisor should warn the pupils NOT TO OPEN THE BOOKLETS UNTIL THEY ARE TOLD. Then he should distribute the booklets

WITH THE FRONT PAGE UPPERMOST.

When ready he should say: "Print your last name in the space after the word 'surname', and your Christian name (or names) in the line below. Then fill in the other spaces. In the space for date of your birthday, write the day, month and year of your birth. Write very plainly. When you have finished filling in the particulars, lay your pencils down. DO NOT OPEN OR TURN OVER THE BOOKLETS."

3. When all the pupils have finished, the supervisor says, "Now, everyone attend carefully while I read the rules of the test". He should hold up the test booklet and point to the instructions and say "Do you see where it says 'Read the following carefully'? I shall read these rules to you, and you will follow as I read them." The supervisor then reads the rules, clearly and without haste.

4. As soon as the rules are read, the supervisor says "The test is in six sections. You will all begin each section at the same time and stop at the same time. At the start of each section, you will be told the number of the test, and the number of the page, so that you can make sure that you are working at the right page."

"Before you start each section, you will do a Practice Test, to make sure that you understand what is to be done in the actual test. The Practice Test is not marked, so your score in it will not change your final mark."

"Remember you must not ask any questions during the working of the test. You will be allowed to ask questions at the end of each Practice Test."

5. "Now turn over to page 3. This is Practice Test 1. I shall read the instructions and you follow them. The three drawings at the top of the page represent the end views and the middle section of a solid object. They are in the order end, middle, end as marked. Imagine them placed as in the next diagram and you see that if they were joined up they form the object in the third diagram. Try to imagine the object formed by the next three shapes. Is it A or B or C? Put your answer in the space provided. Do it now." The supervisor pauses for 15 seconds and then says "The answer is 'C'. If you have any questions, ask them now". The supervisor pauses and briefly answers any questions concerning the answers in Practice Test 1.

"Now turn over to page 4. Make sure you are at page 4 and Test 1. On this page are sets of three drawings showing end, middle, end. In the spaces provided put the letter of the object on page 5 which fits each set of drawings. You will have 4 minutes to do this test. Are you ready? BEGIN." The supervisor writes down the exact time when he said "Begin" and also what the time will be 4 minutes later. After the test has started he should answer NO QUESTION WHATSOEVER.

6. After 4 minutes, he says "STOP WORKING, PENCILS DOWN". Pausing, he says "Now turn over to page 7, to Practice Test 2. Follow the instructions as I read them. Below is a model made of blocks placed together, and next to it are four drawings A, B, C and D. One of these drawings shows the view of the model looking down on it from above.

A cross has been placed on diagram C because it shows the view looking down on the model from above.

In the Trial Questions you have to place a cross on the diagram A, B, C or D which you think shows the model if you were looking down on it from above. Make sure you put the cross ON the diagram. Do these two questions now." The supervisor pauses for 30 seconds and then says "The answer to No. 1 is B, the answer to No. 2 is D. If you have any questions, ask them now". The supervisor pauses, and briefly answers any questions concerning the answers in Practice Test 2.

"Now turn over to page 8. Make sure you are at page 8 and Test 2. There is only one page to this test and you will have $3\frac{1}{2}$ minutes to do it. Are you ready? BEGIN." The supervisor writes down the exact time when he said "Begin" and also what the time will be $3\frac{1}{2}$ minutes later.

7. After $3\frac{1}{2}$ minutes he says "STOP WORKING, PENCILS DOWN." Continuing he says "Now turn over to page 11 to Practice Test 3. The diagram at the top on the left is the shape of a piece of paper which when folded forms the model shown in the second diagram. A and B are folded to form the walls and C forms the roof.

Below this are the same shape and a model with part of its sides shaded. Look at the shape. Lines have been drawn on it to show where you would cut to remove the parts shown shaded on the model.

In the trial questions you have to look at the shaded parts on the model and then draw lines on the SHAPE to show where you would cut to remove those parts.

Do these two questions now." The supervisor pauses 45 seconds and then says "The correct answers are, in No. 1 you should have drawn a line joining the point marked 2 to the point marked 14, in No. 2 your line should join points 1 and 8. If you have any questions, ask them now." The supervisor pauses and briefly answers any questions concerning the answers in Practice Test 3.

"Now turn over to page 12. Make sure you are at page 12 and Test 3. This test occupies two pages. When you reach the foot of page 12 go on to page 13. You will have

$6\frac{1}{2}$ minutes to do both pages. Are you ready? BEGIN." The supervisor writes down the exact time when he said "Begin" and also what the time will be $6\frac{1}{2}$ minutes later.

8. After $6\frac{1}{2}$ minutes he says "STOP WORKING, PENCILS DOWN." Continuing he says "Turn over to page 15 to Practice Test 4. The first diagram represents a wooden block which is to be sawn vertically into two parts in the place shown by the dotted lines. In the next diagram the two halves are moved apart to show the cut fact which is shaded. The shape of this cut face is the same as one of the three drawings which are shown next. Drawing B is the correct shape and so a cross has been placed on it. In the Trial Question you are shown another block to be sawn and you have to put a cross on one of the three drawings A, B and C which shows the shape of the cut face. Do it now." The supervisor pauses and then says "The correct answer is C. Are there any questions?" The supervisor pauses and briefly answers any questions concerning the answers to Practice Test 4.

"Now turn over to page 16. Make sure you are at page 16 and Test 4. This Test occupies two pages. When you reach the foot of page 16 go on to page 17. You will have 5 minutes for both pages. Make sure you put your crosses ON the diagrams. Are you ready? BEGIN." The supervisor notes the time when he said "Begin", and also what the time will be 5 minutes later.

9. After 5 minutes he says "STOP WORKING, PENCILS DOWN". Continuing he says "Now turn over to page 19 to Practice Test 5. Shape A shows a solid block which has been drawn by joining up the crosses which are at its corners. It could be copied on to Framework B by joining the crosses with circles round them. Look at Shape C, this could be drawn on Framework D by joining the crosses which have circles round them. Now look at the Trial Question, at Shape E. You have to put circles round the crosses in Framework F which you would join to make that shape. Do it now." The supervisor pauses 60 seconds and then says "Your circles should be round crosses 3 and 5 in the top line, crosses 1, 3 and 5 in the middle line, and crosses 1 and 3 in the bottom line. Any questions?" The supervisor pauses and briefly answers any questions concerning the answers in Practice Test 5.

"Now turn over to page 20. Make sure you are at page 20 and Test 5. This test occupies two pages. When you reach the foot of page 20 go on to page 21. You will have 16 minutes for both pages. Are you ready? BEGIN." The supervisor notes the time when he said "Begin" and also what the time will be 16 minutes later.

10. After 8 minutes he says "Remember there are two pages in this test, pages 20 and 21. You have 8 minutes more."

At the end of 16 minutes he says "STOP WORKING, PENCILS DOWN." Continuing he says "Turn over to page 23 to Practice Test 6. In this test each question shows a model built from the sides shown next to it. These sides are labelled a, b, c ...... etc. No other sides are used but some of the ones shown are used more than once. Put on each of these shapes a figure showing the number of times it has been used to make the model. You must remember to count the sides and base of the model which do not show in the drawing.

Look at the Example. It is made of 4 triangle sides and 1 square base and so these numbers have been marked on the shapes, 4 on 'a' and 1 on 'b'.

Do the two trial questions now." The supervisor then pauses for 60 seconds and says "The correct answers are No. 1 a - 2, b - 3; No. 2 a - 1, b - 2, c - 2. Are there any questions?" The supervisor pauses and briefly answers any questions concerning the answers in Practice Test 6.

"Now,turn over to Page 24. Make sure you are at Page 24 and Test 6. When you have finished this page do NOT turn back to earlier pages. You will have 4 minutes to do this test. Are you ready? BEGIN." The supervisor notes the time when he said "Begin" and also what the time will be 4 minutes later.

11. After 4 minutes he says, "STOP WORKING, PENCILS DOWN." The test booklets are then collected.

Appendix C.

The Test - Revised Draft.

Note: In addition to the name of the author the printed form of the test bears the name of his supervisor, Dr. I. Macfarlane Smith.

NEWCASTLE SPATIAL TEST

by

I. Macfarlane Smith M.A., B.Sc., Ed.B., Ph.D. & J.S. Lawes B.Sc

DO NOT TURN OVER OR OPEN THIS BOOK UNTIL YOU ARE TOLD

FILL IN THE FOLLOWING PARTICULARS:-

SURNAME ....

CHRISTIAN NAME(S) ....

NAME OF YOUR SCHOOL ....

YOUR AGE .... YEARS .... MONTHS

DATE OF YOUR BIRTHDAY ....

TODAY'S DATE ....

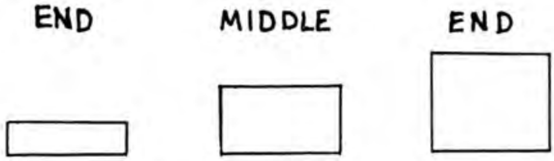
READ THE FOLLOWING CAREFULLY:-

1. Do not open this book until you are told to do so.
2. The test is in sections. You will be told how much time is allowed for each section.
3. When you come to the end of a page, FOLLOW THE INSTRUCTIONS given at the bottom.
4. Each time you are told to stop, STOP WORKING AT ONCE.
5. Work as quickly and as carefully as you can.
6. If, when you try a question you find you cannot do it, DO NOT WASTE TIME BUT GO ON TO THE NEXT.
7. Make any alterations in your answers CLEARLY.
8. ASK NO QUESTIONS AT ALL DURING THE TEST.
9. If you should require another pencil, put up your hand.

Not to be filled in by Scholar

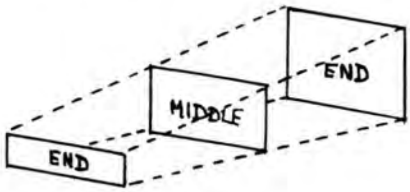
Page	Item Nos.	Score
4	1-10	
8	11-20	
12	21-30	
13	31-40	
16	41-50	
17	51-60	
20	61-70	
21	71-80	
24	81-100	
Total		
Age	Years	Months
Standardised Score		

PRACTICE TEST 1.

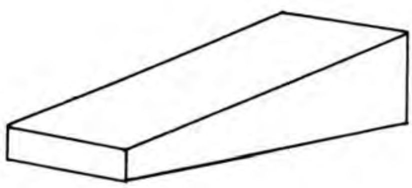


These three drawings represent the end views and middle section of a solid object, in the order - end, middle, end, as marked.

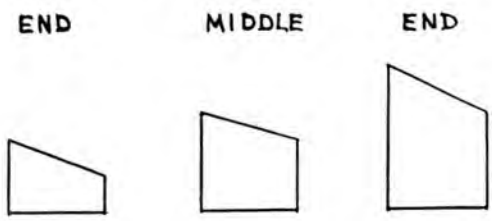
Imagine them placed like this:-



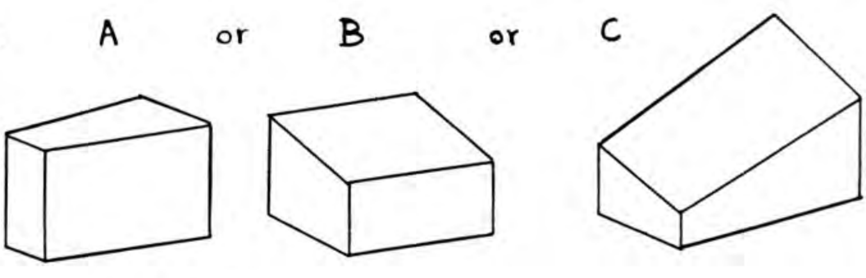
and you see that they form this object:-



Try to imagine the object formed by these shapes:-



Is it A or B or C

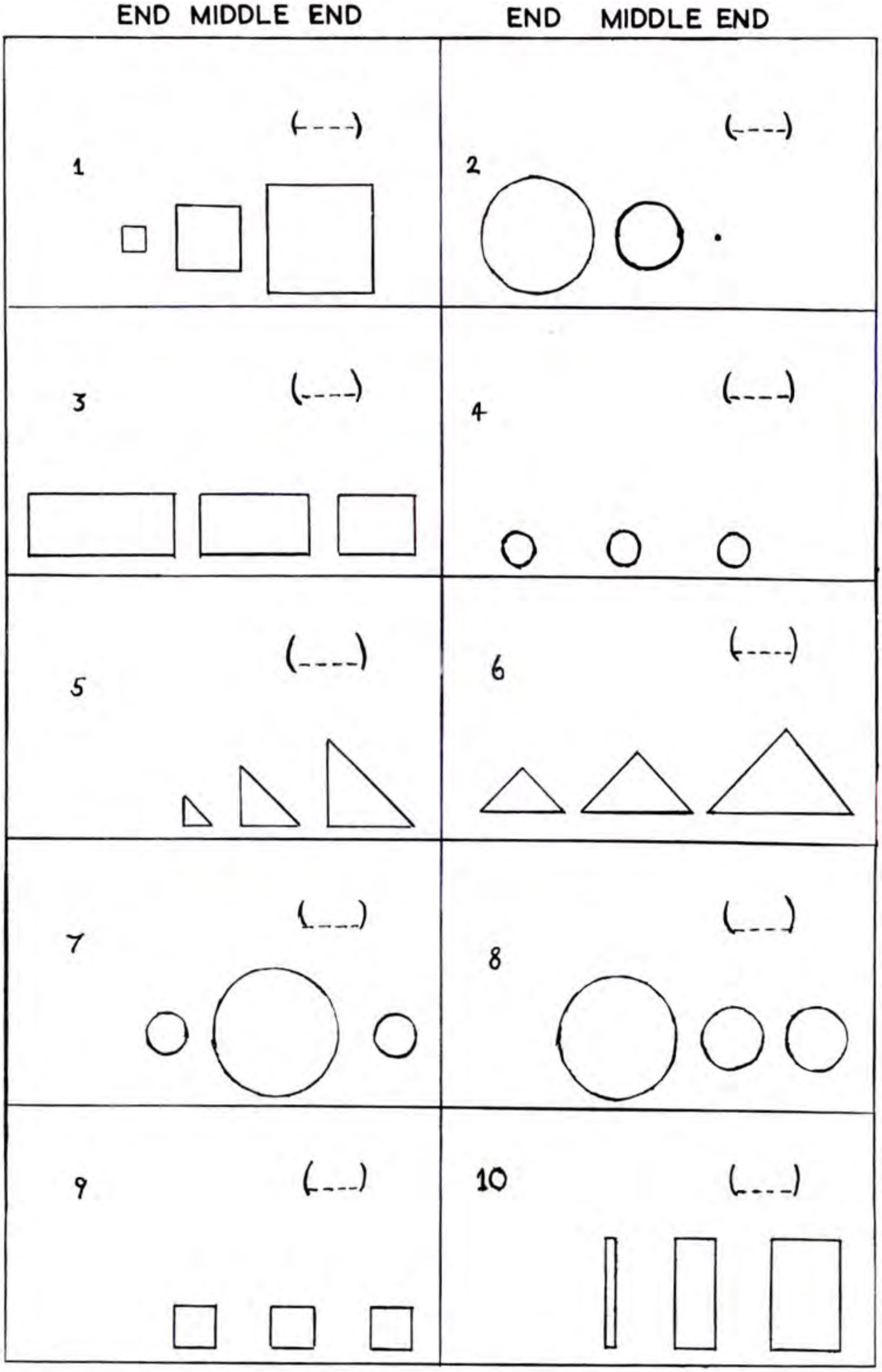


Put your answer here (.....)

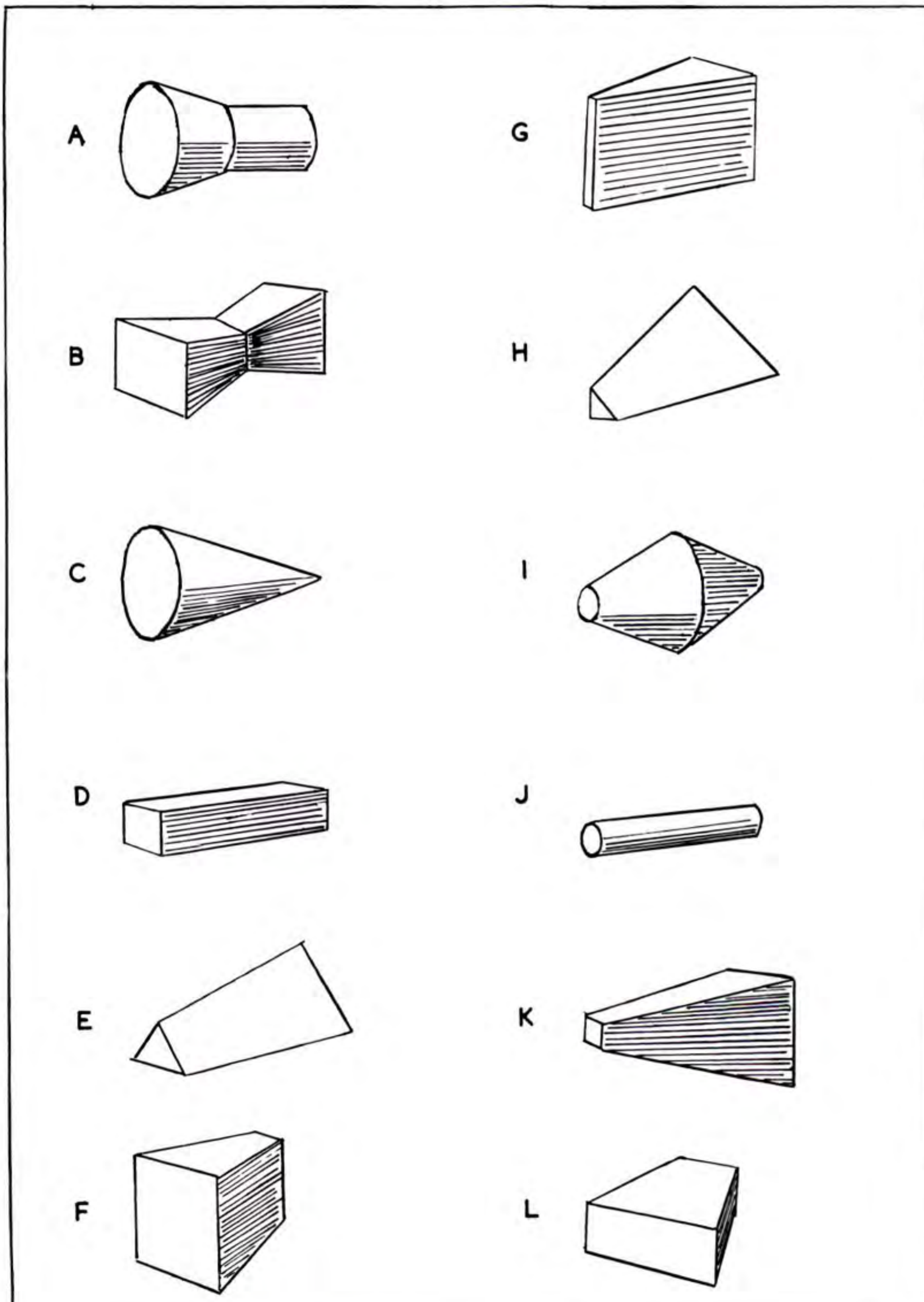
DO NOT TURN OVER UNTIL YOU ARE TOLD.

TEST 1.

The drawings on this page are in sets of three. Each set shows the end views and middle section of a solid object in the order - end, middle, end. In the spaces provided (....) put the letter of the object on the opposite page which fits each set of drawings.



These are the objects to match up with the drawings on the opposite page.

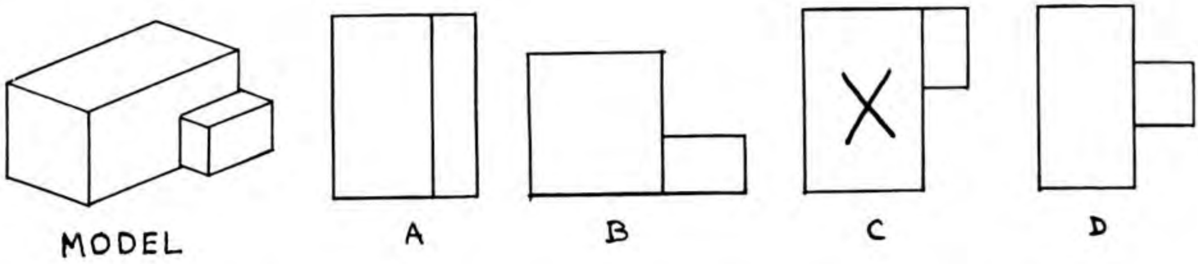


DO NOT TURN OVER UNTIL YOU ARE TOLD.

DO NOT TURN BACK TO EARLIER PAGES.

PRACTICE TEST 2.

Below is a model made of blocks placed together and next to it are 4 drawings. One of these drawings shows the view of the model looking down on it from above.

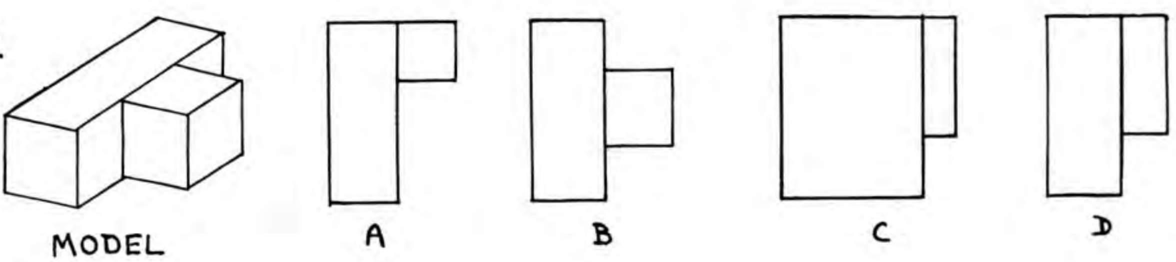


A cross (X) has been placed on diagram C because it shows the view looking down on the model from above.

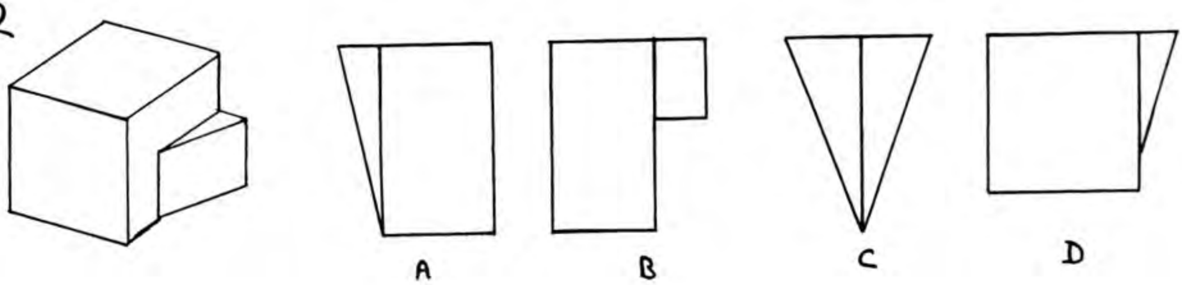
TRIAL QUESTIONS

Place a cross (X) on the diagram (A, B, C or D) which you think shows the model if you were looking down on it from above.

1



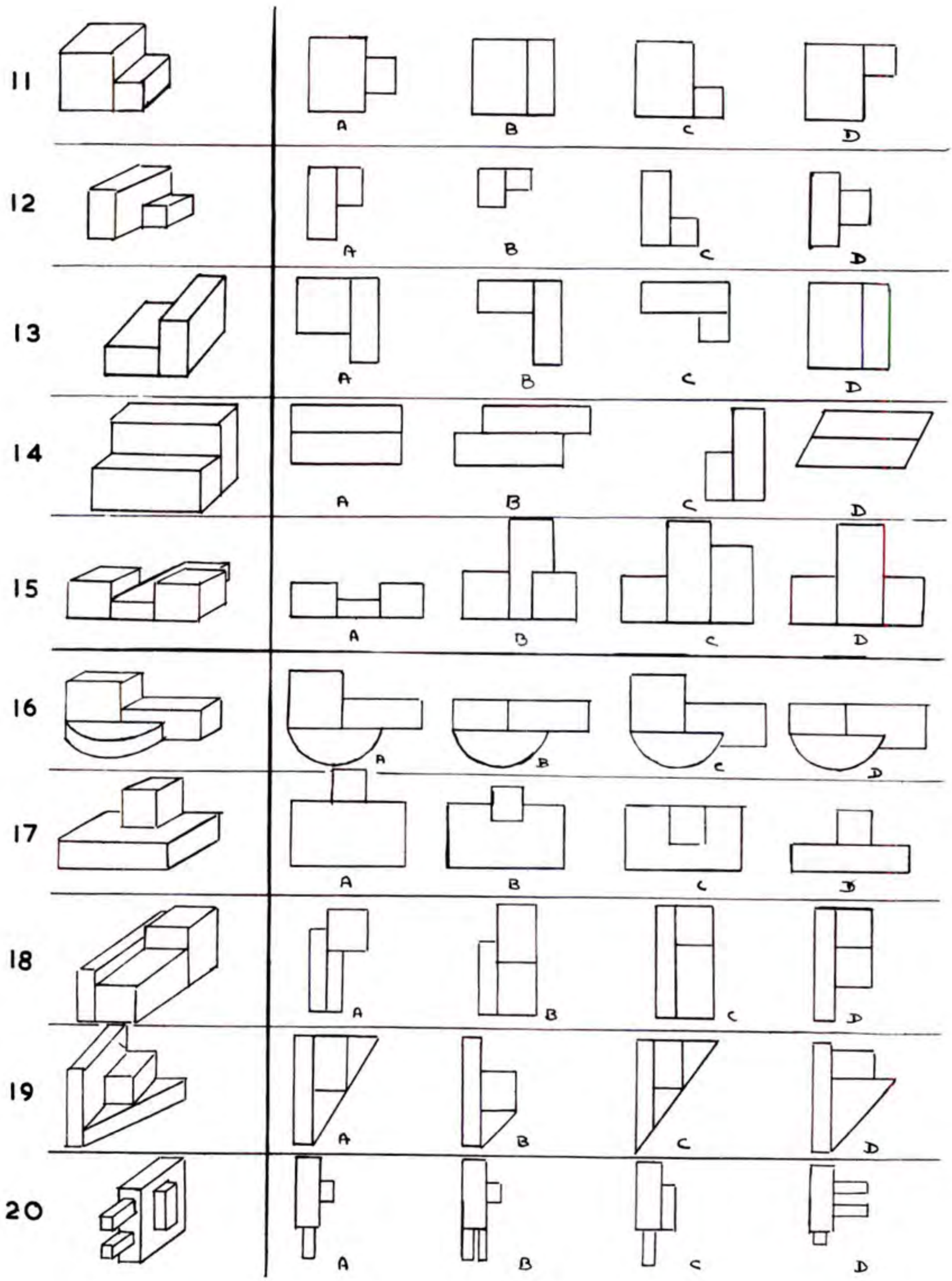
2



DO NOT TURN OVER UNTIL YOU ARE TOLD.

TEST 2.

On each line the left hand drawing shows several blocks placed together. Put a cross, "X", on the one of the four drawings on the right of the thick black line which shows the view looking down on the blocks.



DO NOT TURN BACK TO EARLIER PAGES.

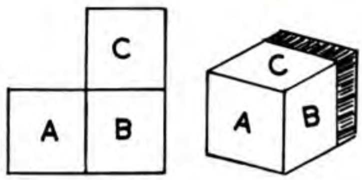
DO NOT TURN OVER UNTIL YOU ARE TOLD.

TEST 3.

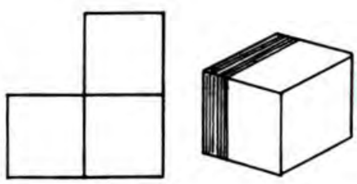
On the Shape draw lines to show where you would cut to remove the parts which would be the shaded portions if the paper was folded to form the Model.

SHAPE **MODEL** **SHAPE** **MODEL**

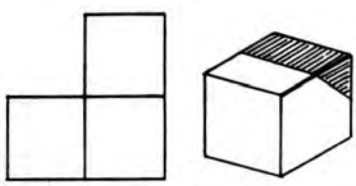
21



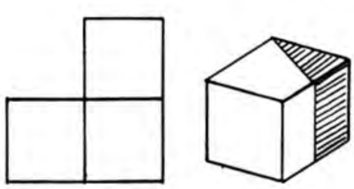
22



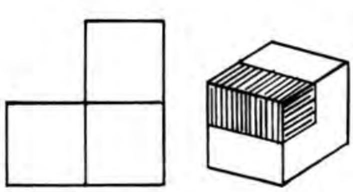
23



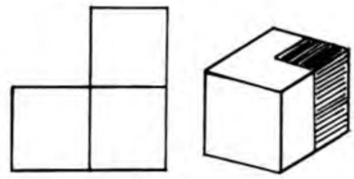
24



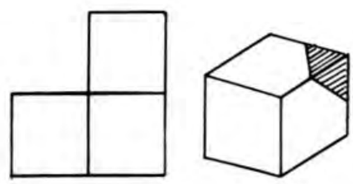
25



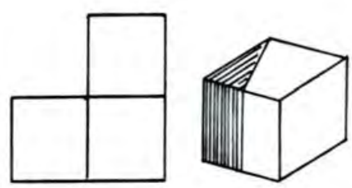
26



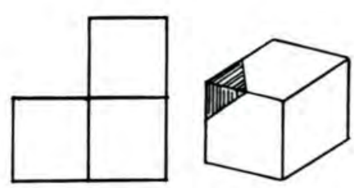
27



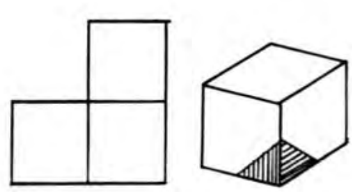
28



29

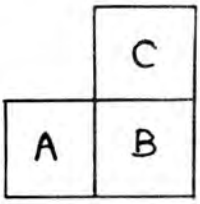


30

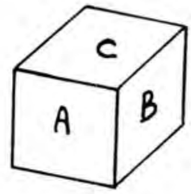


CARRY STRAIGHT ON TO THE NEXT PAGE

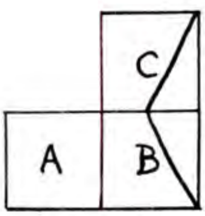
PRACTICE TEST 3.



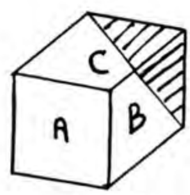
This diagram is the shape of a piece of paper which when folded forms this model:-



Below are the same Shape and Model with part of the Model shaded. On the Shape lines have been drawn to show where you would cut to remove the parts shown shaded on the Model.



SHAPE



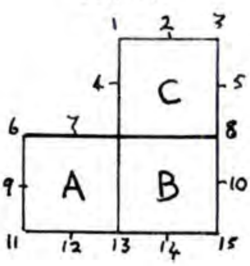
MODEL

TRIAL QUESTIONS

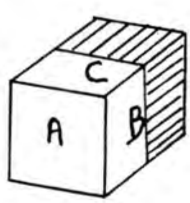
Draw lines on the Shape to show where you would cut to remove the parts shown shaded on the Model.

SHAPE

1

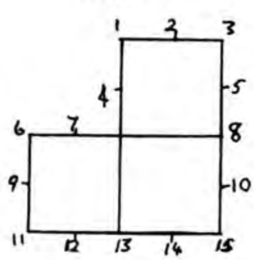


MODEL

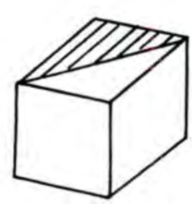


SHAPE

2



MODEL



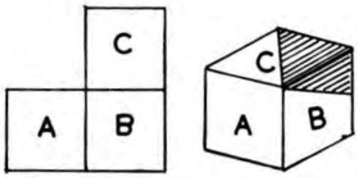
DO NOT TURN OVER UNTIL YOU ARE TOLD.

DO NOT TURN BACK TO EARLIER PAGES.

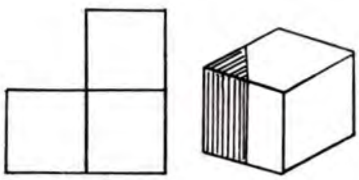
CONTINUE AS ON THE LAST PAGE

SHAPE MODEL

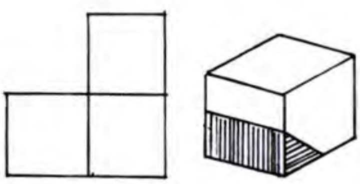
31



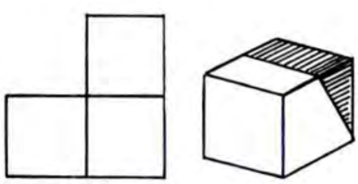
32·



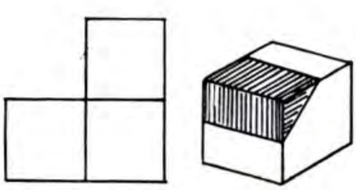
33



34

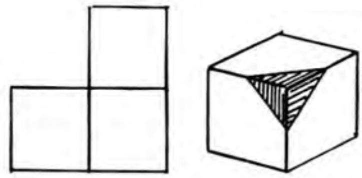


35

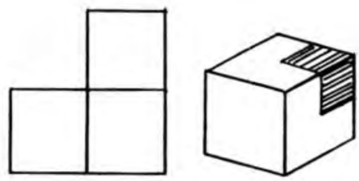


SHAPE MODEL

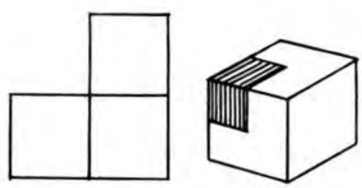
36



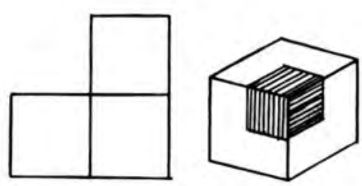
37



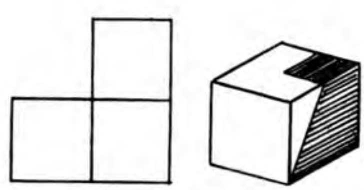
38



39



40

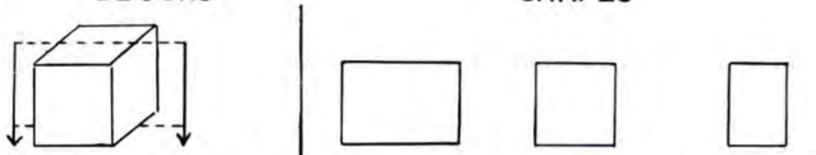


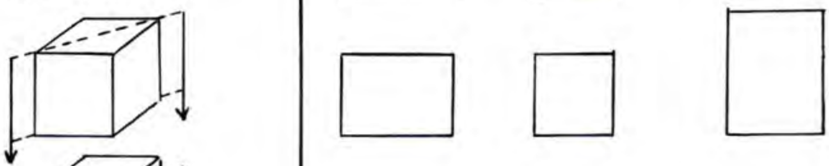
DO NOT TURN OVER UNTIL YOU ARE TOLD.

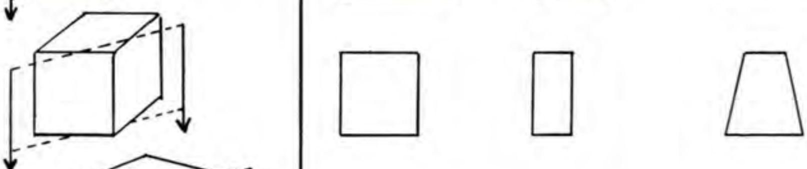
TEST 4.

Each question shows a block of wood. Imagine a cut made where shown by the lines. Place a cross (X) on the one of the three drawings on the right which shows the shape of the cut face.

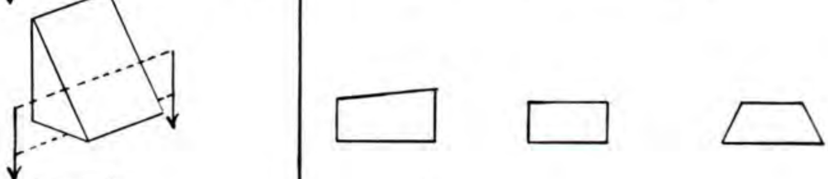
BLOCKS **SHAPES**

41


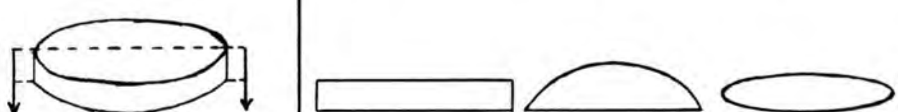
42


43


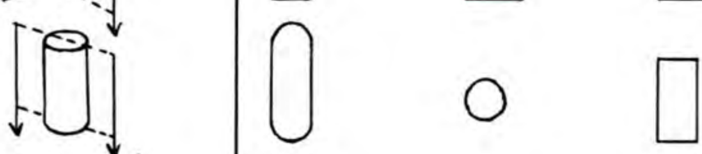
44

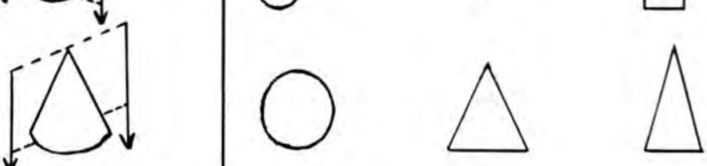

45


46


47


48

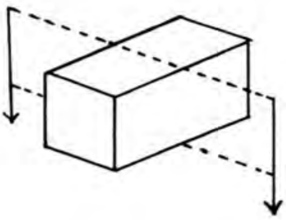

49


50


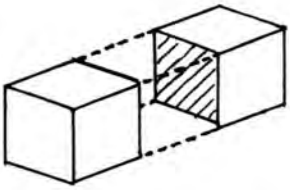
CARRY STRAIGHT ON TO THE NEXT PAGE.

PRACTICE TEST 4.

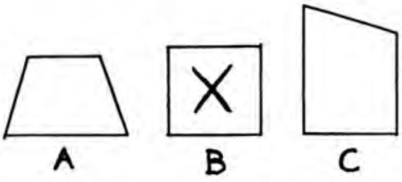
This diagram represents a wooden block which is to be sawn vertically into two parts in the place shown by the dotted lines.



In this next diagram the two halves are moved apart to show the cut face. This cut face is shaded.



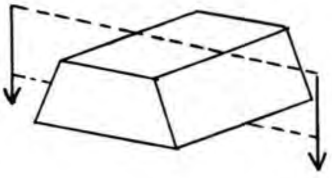
The cut face is the same shape as one of these three drawings.

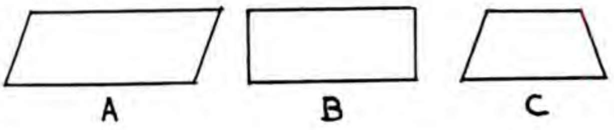


B is the correct shape and so is marked with a X.

TRIAL QUESTION.

What would be the shape of the cut face if the block shown below is cut where shown? Put a cross (X) on the drawing which shows the correct shape.

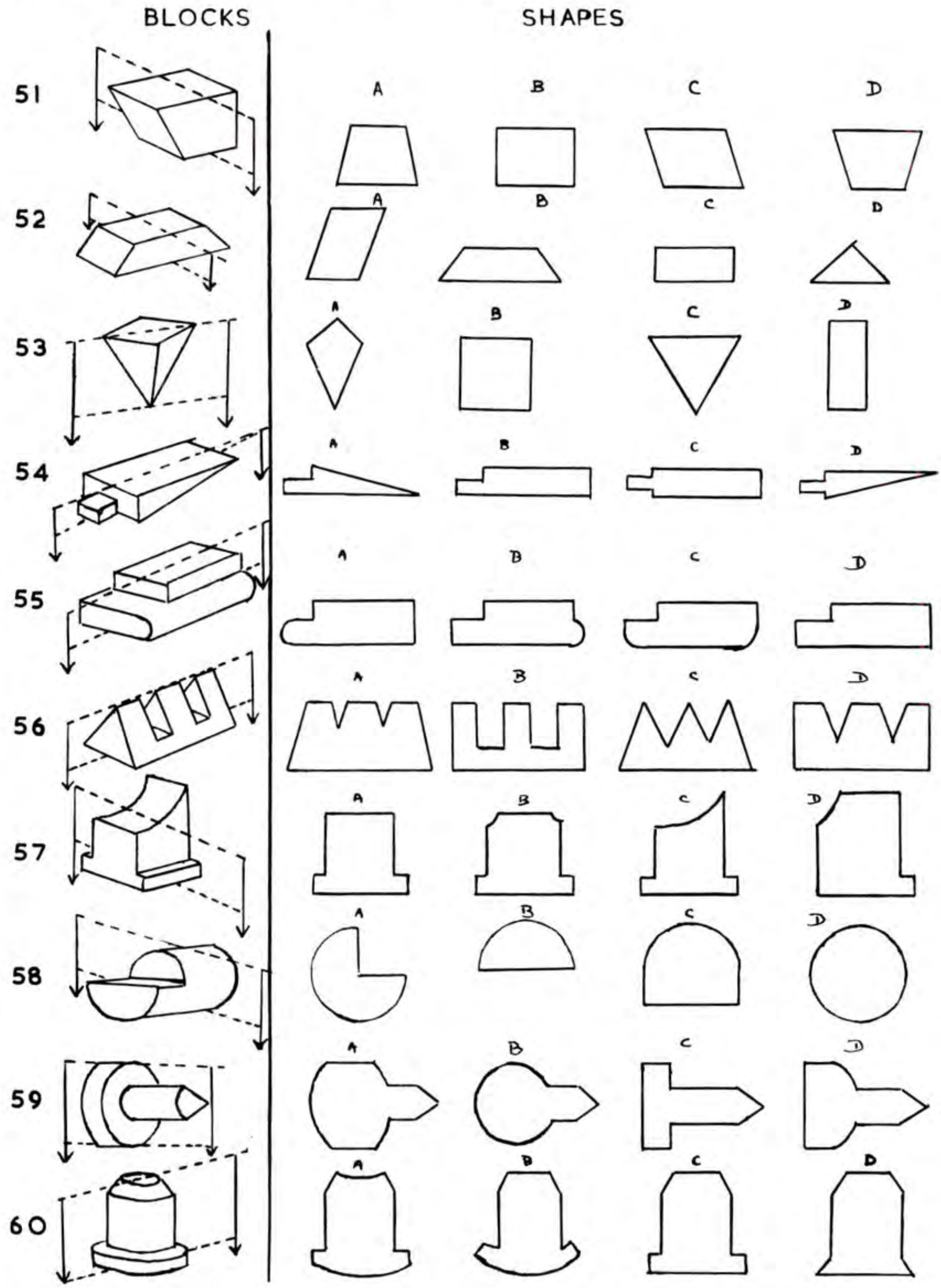




DO NOT TURN OVER UNTIL YOU ARE TOLD.

DO NOT TURN BACK TO EARLIER PAGES.

CONTINUE AS ON THE LAST PAGE
This time there are four shapes to choose from.

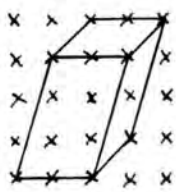


DO NOT TURN OVER UNTIL YOU ARE TOLD.

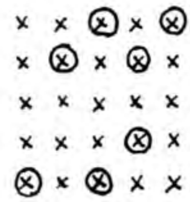
PRACTICE TEST 5.

Shape A shows a solid block which has been drawn by joining up the crosses which are at its corners. It could be copied on to Framework B by joining the crosses with circles round them.

SHAPE A

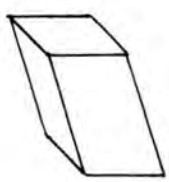


FRAMEWORK B



In the same way Shape C could be drawn on Framework D by joining the crosses which have circles round them.

SHAPE C

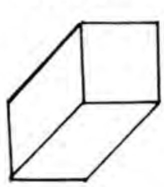


FRAMEWORK D

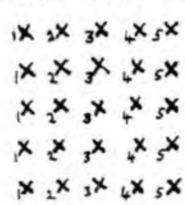
TRIAL QUESTION.

On Framework F put circles round the crosses which could be joined to make Shape E.

SHAPE E



FRAMEWORK F



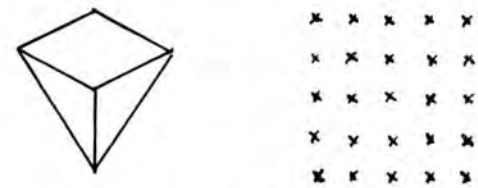
DO NOT TURN OVER UNTIL YOU ARE TOLD.

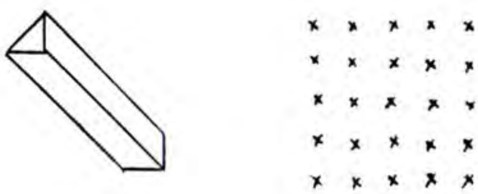
TEST 5.

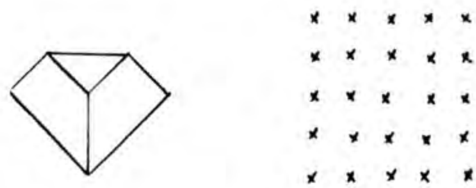
Each question shows the shape of a solid block and a framework of crosses. On each Framework put circles round the crosses you would join up to make the shape shown.

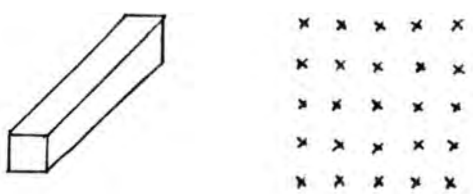
SHAPE	FRAMEWORK	SHAPE	FRAMEWORK

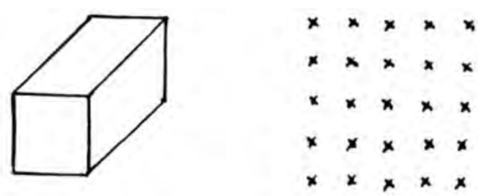
61 

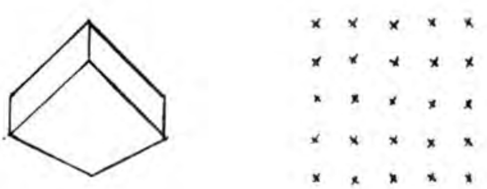
66 

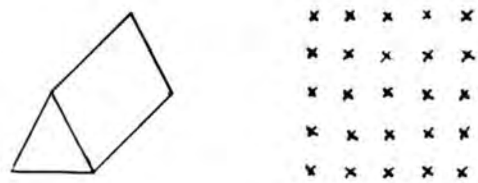
62 

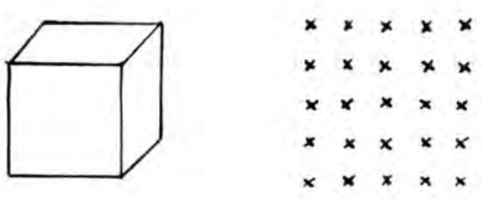
67 

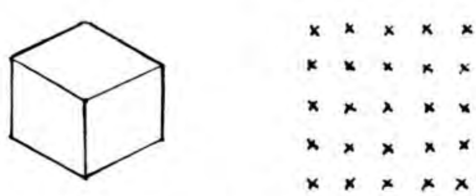
63 

68 

64 

69 

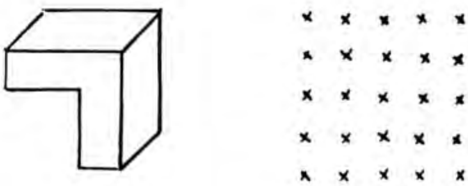
65 

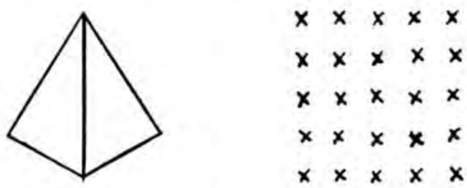
70 

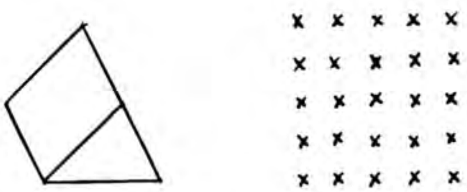
CARRY STRAIGHT ON TO THE NEXT PAGE

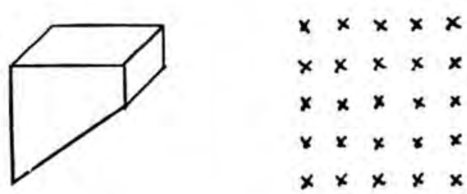
CONTINUE AS ON THE LAST PAGE.

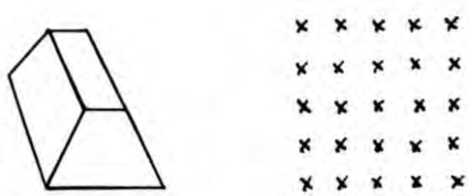
SHAPE	FRAMEWORK	SHAPE	FRAMEWORK

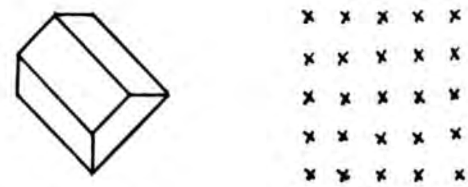
71 

72 

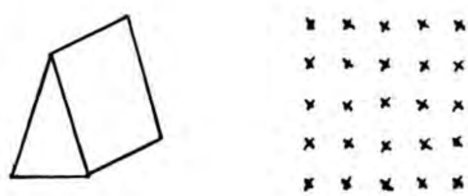
73 

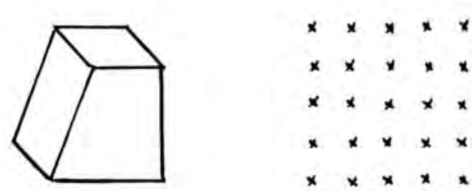
74 

75 

76 

77 

78 

79 

80 

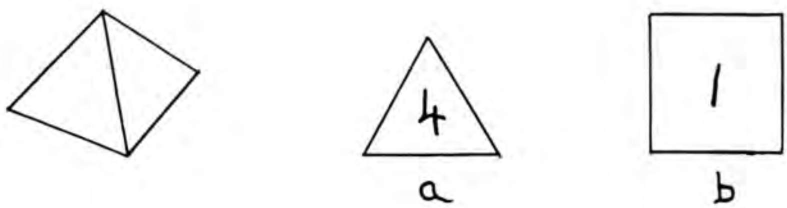
DO NOT TURN OVER UNTIL YOU ARE TOLD.

DO NOT TURN BACK TO EARLIER PAGES.

PRACTICE TEST 6.

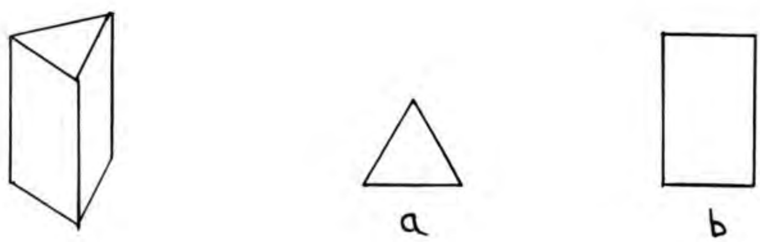
Each question shows a model built from the sides shown next to it. These sides are labelled a, b, c, ... etc. NO OTHER SIDES ARE USED but some of the ones shown are used more than once.
Put on each of these shapes a figure showing the number of times it has been used to make the model.
You must remember to count the sides and base of the model which do not show in the drawing.

EXAMPLE.

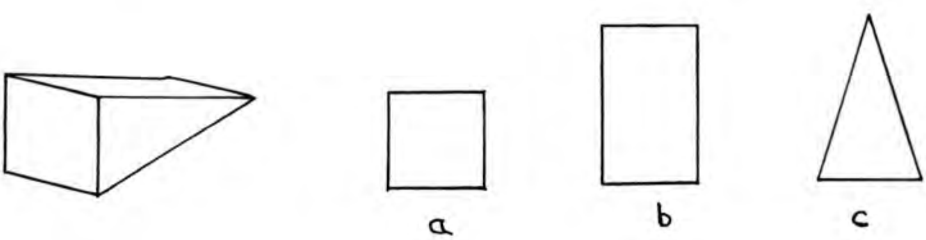


TRIAL QUESTIONS.

1



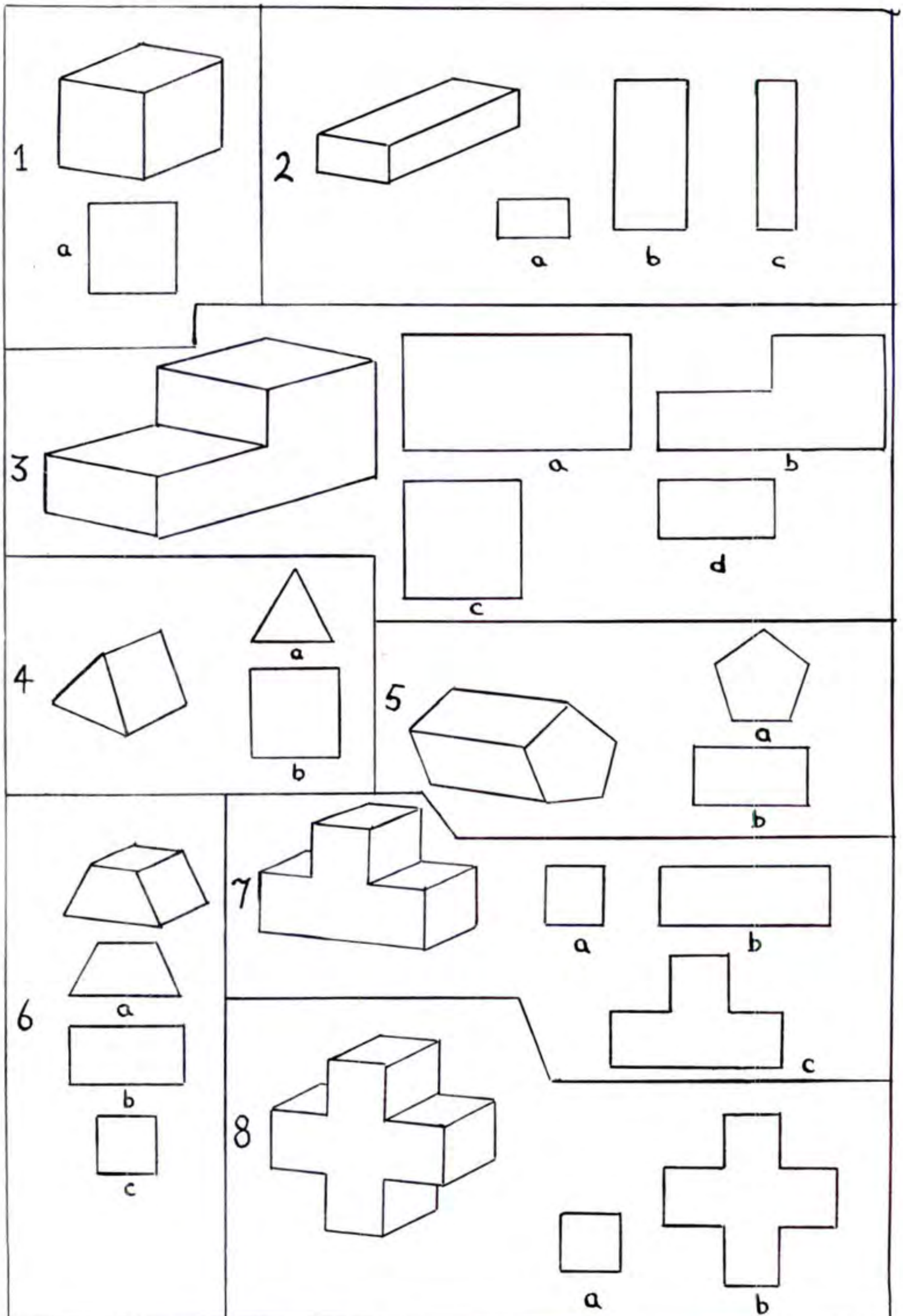
2



DO NOT TURN OVER UNTIL YOU ARE TOLD.

TEST 6.

Each question shows a model and the shapes of its sides labelled a, b, c, ... etc. On each shape put a number showing how many times that side has been used to build the model.



DO NOT TURN BACK TO EARLIER PAGES.

Printed by King's College Printing Section, Newcastle upon Tyne

KCPS 716.1400.10.58

BIBLIOGRAPHY

The following abbreviations are used:-

Brit. J. Ed. Psych.	British Journal of Educational Psychology.
Brit. J. of Psych.	British Journal of Psychology.
J. of Ed. Psych.	Journal of Educational Psychology.
Occup. Psych.	Occupational Psychology.

ADCOCK, C. (1948) A Re-analysis of Slater's Spatial Judgment Research. Occup. Psych. XXII, No. 4.

ALEXANDER, W.P. (1935) Intelligence, Concrete and Abstract. Brit. J. of Psych. Monograph. Supplement XIX.

BAIN, J.T. (1946) The Construction of a Space Test. B. Ed. Thesis, Edinburgh.

BARAKAT, M.K. (1951) A Factorial Study of Mathematical Abilities. Brit. J. of Psych. Statistical Section IV, Pt. 3.

BLACKWELL, A.M. (1940) A Comparative Investigation into the Factors Involved in Mathematical Ability of Boys and Girls. Brit. J. Ed. Psych. X, Pt. 2.

BURT, C. (1917) The Distribution and Relations of Educational Abilities. King, London.

CHAMBERS, E.G. (1952) Statistical Calculations for Beginners. Cambridge University Press, London.

COX, J.W. (1928) Mechanical Aptitude. Methuen, London.

DEMPSTER, J.B.B. (1948) The Selectors Point of View. Brit. J. Ed. Psych. XVIII, Pt. 3.

DREW, L.J. (1947) An Investigation into the Measurement of Technical Ability. Occup. Psych. XXI, No. 1.

EL KOUSSY, A.A.M. (1935) The Visual Perception of Space. Brit. J. of Psych. Monograph Supplement XX.

EL.KOUSSY, A.A.M. (1948) A Further Examination of the K-factor. XIIIth International Congress of Psychology.

EMMETT, W.G. (1949) Evidence of a Space Factor at 11+ and Earlier. Brit. J. of Psych. Statistical Section, Vol. II, Pt. 1.

FISHER, R.A. (1946) Statistical Methods for Research Workers (10th Edition). Oliver and Boyd, Edinburgh.

FLANAGAN, J.C. (1939) General Considerations in the Selection of Test Items and a Short Method of Estimating the Product-Moment Coefficient from Data at the Tails of the Distribution. J. of Ed. Psych. XXX.

GALTON, F. (1869) Hereditary Genius. Macmillan, London.

GALTON, F. (1883) Inquiries into Human Faculty. Dent, London.

GOODENOUGH, F.L. (1950) Mental Testing. Staples, London.

HARTOG, P.J. and RHODES, E.C. (1936) The Marks of Examiners. Macmillan, London.

HOLLIDAY, F. (1940) An Investigation into the Selection of Apprentices for the Engineering Industry. Occup. Psych. XIV, No. 2.

HOLLIDAY, F. (1941) A Further Investigation into the Selection of Apprentices for the Engineering Industry. Occup. Psych. XV, No. 4.

HOLLIDAY, F. (1943) The Relation Between Psychological Test Scores and Subsequent Proficiency of Apprentices in the Engineering Industry. Occup. Psych. XVII, No. 3.

HOLZINGER, K.J. and SWINEFORD, F. (1946) The Relation of Two Bi-factors to Achievement in Geometry and Other Subjects. J. of Ed. Psych. XXXVII.

JOHNSON, A.P. (1951) Notes on a Suggested Index of Item Validity: the U-L Index. J. of Ed. Psych. XLII.

KELLEY, T.L. (1928) Crossroads in the Mind of Man. Stanford University Press, Stanford, Calif.

KELLEY, T.L. (1939) The Selection of Upper and Lower Groups for the Validation of Test Items. J. of Ed. Psych. XXX.

KERR, G. (1942) Aptitude Testing for Secondary Courses. Occup. Psych. XVI, No. 2.

LINDQUIST, E.F. (1940) Statistical Analysis in Educational Research. Harrap, London.

MILLS, L.F. (1947) An Investigation into the Properties of a Space Test. B. Ed. Thesis, Edinburgh.

MORONEY, M.J. (1956) Facts from Figures. Penguin Books, Harmondsworth.

PEEL, E.A. (1949) Evidence of a Practical Factor at the Age of Eleven. Brit. J. Ed. Psych. XIX, Pt. 1.

PICHOT, P. (1956) Les Tests Mentaux. Presses Universitaires de France, Paris.

PRICE, E.J.J. (1940) The Nature of the Practical Factor (F).
Brit. J. of Psych. XXX, Pt. 4.

RENSHAW, T. (1950) A Factorial Study of Two and Three Dimensional Space Tests.
Ph.D. Thesis, Edinburgh.

SHUTTLEWORTH, C.W. (1942) Tests of Technical Aptitude.
Occup. Psych. XVI, No. 4.

SLATER, P. (1940) Some Group Tests of Spatial Judgment or Practical Abilities.
Occup. Psych. XIV, No. 1.

SLATER, P. (1941) Tests for Selecting Secondary and Technical School Children.
Occup. Psych. XV, No. 1.

SLATER, P. and BENNETT, E. (1943) The Development of Spatial Judgment and its Relation to Some Educational Problems.
Occup. Psych. XVII, No. 3.

SMITH, I. MACFARLANE (1948) Measuring Spatial Ability in School Pupils.
Occup. Psych. XXII, No. 3.

SMITH, I. MACFARLANE (1954) The Development of a Spatial Test.
Durham Research Review 5.

SMITH, I. MACFARLANE (1959) Selection for Technical Education.
Journal of the Institute of Education of Durham University.
Vol. 10, No. 51.

SMITH, I. MACFARLANE, (1960) The Validity of Tests of Spatial Ability as Predictors of Success on Technical Courses.
Brit. J. Ed. Psych. XXX, Pt. 2.

SPEARMAN, C. (1932) The Abilities of Man.
Macmillan, London.

SPEARMAN, C. and WYNN JONES, Ll. (1950) Human Ability.
Macmillan, London.

STEPHENSON, W. (1931) Tetrad Differences for Non-Verbal Sub-tests. Tetrad Differences for Verbal Sub-test. Tetrad Differences for Verbal Sub-tests relative to Non-Verbal Sub-tests.
J. of Ed. Psych. XXII.

THORNDIKE, R.L. (1949) Personnel Selection.
Wiley, New York.

THURSTONE, L.L. (1938) Primary Mental Abilities.
University of Chicago Press, Chicago.

TIPPETT, L.H.C. (1948) The Methods of Statistics.
Williams and Norgate, London.

VERNON, P.E. (1948) Indices of Item Consistency and Validity.
Brit. J. of Psych. Statistical Section, Vol. I, Pt. 3.

VERNON, P.E. (1950) The Structure of Human Abilities.
Methuen, London.

VERNON, P.E. and PARRY, J.B. (1949) Personnel Selection in the British Forces.
University of London Press, London.

WATTS, A.F. and SLATER, P. (1950) The Allocation of Primary School Leavers to Courses of Secondary Education.
1st Interim Report, N.F.E.R.
Newnes, London.

WRIGLEY, J. (1958) The Factorial Nature of Ability in Elementary Mathematics.
Brit. J. Ed. Psych. XXVIII, Pt. 1.

YATES, A. and PIDGEON, D.A. (1957) Admission to Grammar Schools.
3rd Interim Report, N.F.E.R.,
Newnes, London.

ZUBIN, J. (1936) Note on a Graphic Method of Determining the Significance of the Difference between Group Frequencies.
J. of Ed. Psych. XXVII.

Spatial Test 1	N.F.E.R.
Verbal Test 8A	N.F.E.R.
Verbal Test (Adv.) 1	N.F.E.R.
Verbal Reasoning Test 59	Moray House.

ACKNOWLEDGEMENTS

The author's sincere thanks are due to:-

the administrators, headteachers and staffs of the various schools in which the test has been used,

the Institute of Education of Durham University for a research grant to defray the costs of printing,

Mrs. D.E. Lawes, M.A., for considerable assistance with marking and checking,

and most particularly to:-

Dr. I. Macfarlane Smith for unlimited encouragement and advice throughout the work involved.