

STANDARDIZATION OF PROGRESSIVE MATRICES, 1938

By JOHN C. RAVEN

From the Research Department, Royal Institution, Colchester

I. INTRODUCTION

PROGRESSIVE matrices provide a non-verbal series of tests, suitable for measuring intelligence. From the results of an experimental survey carried out in Colchester⁽¹⁾ a series of sixty matrix tests was prepared for general and clinical use⁽²⁾. Each test consists of a design or 'matrix' from which part has been removed. A testee has to examine the matrix and decide which, of several pieces given below it, is the right one to complete it. Matrix tests can be given successfully to almost any testee irrespective of age, linguistic ability, or physical defects^(4, 5). The printed form of the series is designed to be used as an individual or group test and is suitable for children above six years of age and for adults. It consists of five sets of problems. The initial problems in each set are easy so as to be self-evident. These are followed by problems of increasing difficulty. The sequence of the problems in the series provides the standard training which each testee receives.

The series was standardized for children in the County Borough of Ipswich, by courtesy of the Public Health Department and the Ipswich Education Committee. Individual tests were carried out by Mr R. Cleaver, M.A., and Mr M. Woolner, B.A., between June and December 1938. Group tests were carried out by the writer between June and September 1938. The series was standardized as a group test for men by courtesy of a training depot for militia men. Testing was carried out by the writer and others between April and September 1940.

The success of the standardization for children was due largely to the co-operation of the teachers in the elementary, secondary, public, private and special schools of Ipswich, and to the educational and medical services.

II. STANDARDIZATION

Ipswich is a port, a manufacturing centre, and a market town in an agricultural district. All types of occupation are well represented. As elsewhere, there has been since about 1920 a steady tendency for people to leave the borough and live in surrounding districts, but the borough boundaries have also been extended. In 1931 the population of Ipswich was 87,557, and between 1924 and 1933 approximately 1400 births were

registered annually. Every birth occurring within the borough is entered in the Maternity and Child Welfare Register. The register contains information concerning the parents and notes on the child's welfare, until particulars concerning the child are forwarded to the school medical files where they are kept until the child leaves school or moves out of the borough.

(a) *The individual test*

To obtain a random sample of children born in Ipswich, particulars were taken, from the Maternity and Child Welfare Register, of three in every forty entries, in numerical order, between May 1924 and December 1932.¹ The name, address, and occupation of each child's father was noted, together with the child's early history and school first attended. Where the child remained in the borough, he could be traced to his present school and tested.

Of 919 birth entries noted:

660 children living in the borough were tested,
141 families had moved out of the borough,
30 families had moved and could not be traced,
13 children had left school,
75 children were still-born or had died in infancy.

When the children were grouped according to their father's occupation, it was found that the incidence of families leaving the borough was greater for men in skilled occupations than for unskilled men, and that the percentage of children tested was different for the different groups.

Table 1. *Occupational groups and incidence of children tested*

Occupations of fathers	No. in original sample	Percentage of children	
		Tested	Untested
Owners, professional, clerical, commercial	206	71	29
Skilled and process workers	350	79	21
Unskilled workers	257	84	16
Occupations unknown	31	61	39
Total	844	78	22

The mean age of the children tested was approximately the same for each group but the mean test score was different.

¹ No practical method of standardization is ideal. The method described recommended itself as giving an accurate random sample of children born in the borough. It provided sociological information concerning every individual in the original sample, whether tested or not.

Table 2. *Occupational groups and mean scores of children*

Occupations of fathers	No. tested	Mean age in months	Mean score
Owners, professional, clerical, commercial	147	118.5	27.5
Skilled and process workers	277	118.1	24.2
Unskilled workers	217	119.1	21.6
Occupations unknown	19	112.1	21.0
Total	660	118.3	24.0

To compensate for the percentage of untested children in each group the percentage tested was proportionately weighted. If X children of a class were tested and Y children were not tested, each individual tested was entered on the score distribution as $(X + Y)/X$ individuals. As long as X is sufficiently large, the correction provides a means of estimating the probable score distribution for the whole group. When X is small, the correction tends to give undue weight to chance results.

For each age group, individuals were divided into four classes according to the occupations of their fathers. The number of individuals obtaining any given score was multiplied by the ratio of the whole group to the number tested. Percentile points were calculated from the weighted score distributions. Graph I shows evidence of chance effects, but in publishing the table of norms for the individual test⁽³⁾ it was thought wise to give the actual percentile points obtained and leave the user free to make such approximations as seem necessary.

(b) *Group test standardization for children*

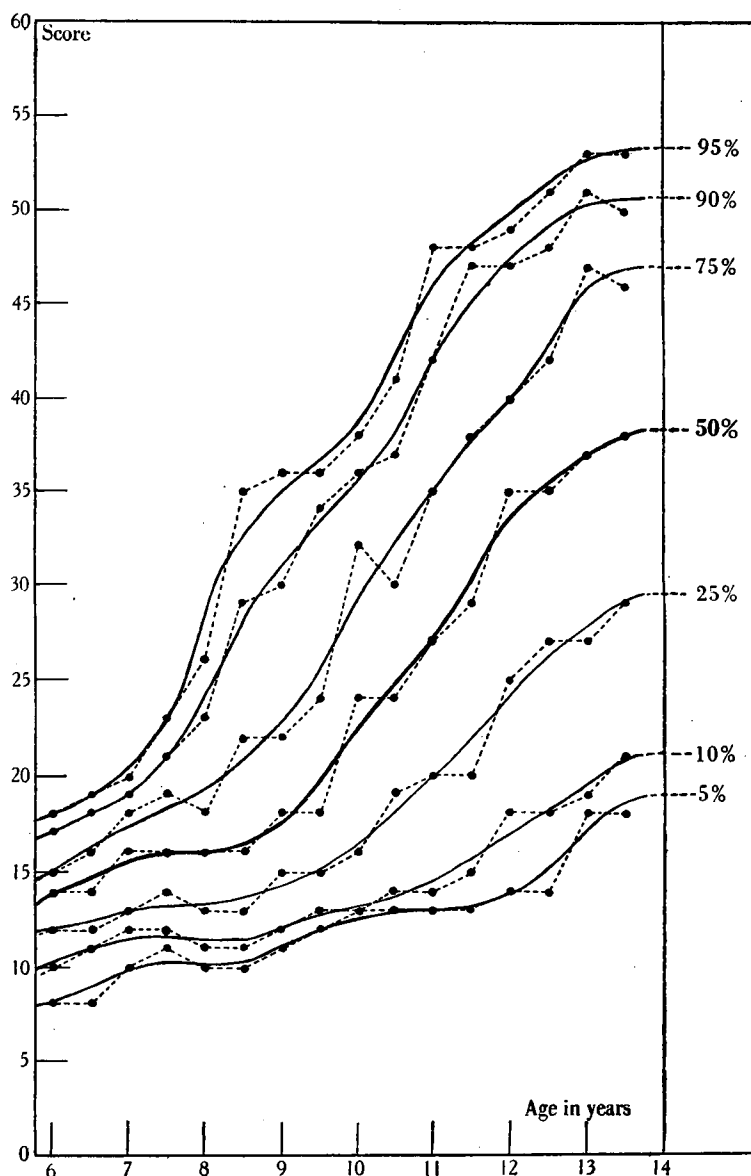
The children tested individually were used as a basis for selecting a random sample of children to be tested in groups. For individual testing practically every school in Ipswich was visited.¹ Attendance returns showed satisfactory agreement between the number of children tested at each school and the total number of children of a similar age attending the school. For each child over eight years of age tested individually, three children in the same class and of the same age were chosen at random to be tested in groups.² Groups varied considerably in size. The standard procedure for group testing was adhered to throughout. One hour was allowed for each group tested.

Elementary, secondary, public, private, and special schools co-operated. 1407 children were tested. Age distributions agreed with the numbers on

¹ Enquiry showed that schools not visited contained less than twelve children over eight years of age.

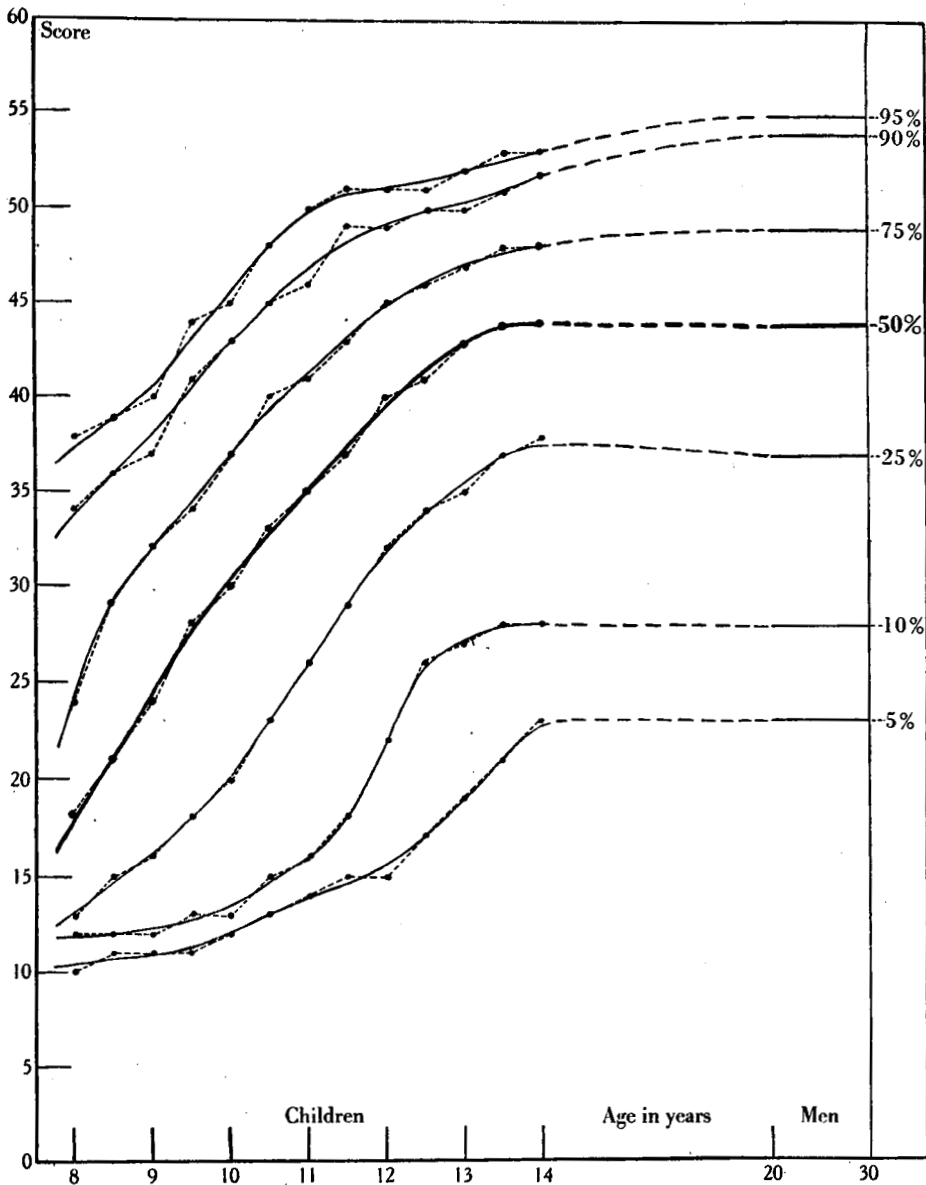
² The method had the advantage that even to the children the selection was obviously arbitrary. There was no feeling that any child had been singled out for a special examination. Suspicion was quickly allayed and co-operation secured.

the school registers. Percentile points were calculated at half-yearly intervals from the natural score distributions of the children tested.



Graph I. *Individual test.* Percentile points calculated from the weighted scores of 660 children

Graph II shows almost no staggering of percentile points due to chance. The published table of norms for the group test gives the percentile points obtained (3). No approximations seem necessary.



Graph II. *Group test*. Percentile points calculated from the natural scores of 1407 children and 3665 men

(c) *Group test standardization for adults*

Men between the ages of twenty and thirty years were tested in groups before they began their military training. The men had been medically examined and were passed as physically fit. No previous mental test had

been given and no selection had been made on evidence of mental ability. 'Reserved' occupations, however, were not represented. Particulars were taken of each man's previous education and civilian occupation. The men were tested in groups of from thirty to thirty-five individuals. The standard group test procedure was adhered to throughout. On some days testing began at 9 a.m. and continued until 6 p.m. There was no significant variation of the score with the time of day. There was a slight variation in the rate of work, the time required to work through the series being a little longer on hot afternoons.

The men were grouped according to their civilian occupations (6). The classification used in grouping the fathers of Ipswich children was adhered to, but 'students' were entered in the professional and clerical group and 'assistants' could not be classified. For each occupational group the percentage of men tested shows close agreement with the percentage obtained in Ipswich by random sampling.

Table 3. *Occupational groups; comparison between a random sample of Ipswich men and militia men*

Occupational groups	Ipswich men %	Militia men	
		%	Median score
Professional, clerical, commercial or students	24	27	50
Skilled workers	42	41	43
Unskilled workers	30	26	36
Unknown	4	—	—
Unclassified assistants	—	6	44

There is marked difference in the median scores for the three occupational groups. The median score for unclassified 'assistants' is 44. This is the same as the median score for the whole group tested. Graph III shows the distributions of scores for each occupational group.

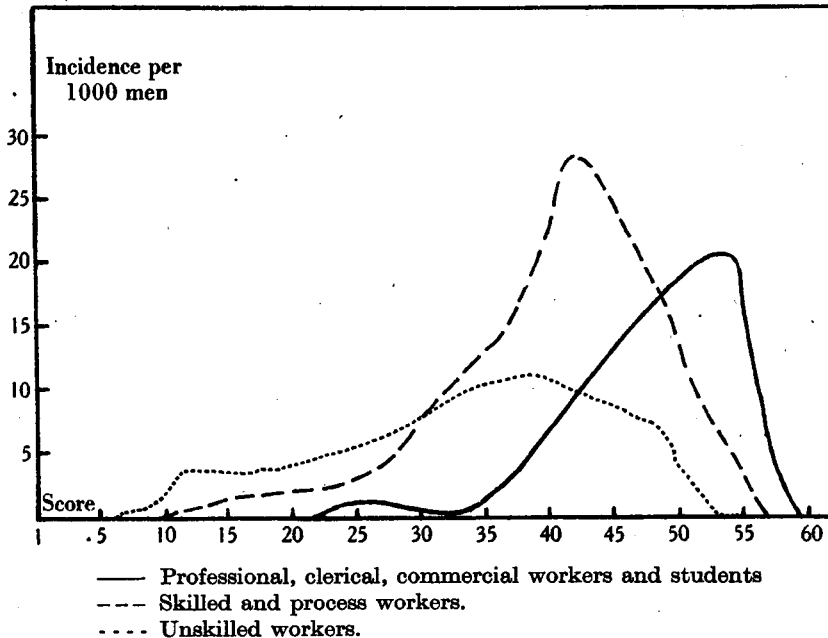
The men were reclassified according to the available information concerning their previous education. 2.5% of the men had had some form of University or Training College education. 20% of the men had received some form of Secondary or Public School education, and 77.5% had attended some form of Elementary School only. Table 4 shows the median score for each of the three groups.

Table 4. *Median scores of militia men for three educational groups*

Educational status reached	% of men	Median score
University	2.5	55
Secondary school	20	50
Elementary school	77.5	41

The available data support the view that the groups tested form an approximately random sample of British men between twenty and thirty years of age. Percentile points for adults were calculated from the natural score distribution for 3665 men.

The relationship between group test percentile points for men and group test percentile points for children is shown in Graph II. The roof of the scale appears to have an influence on scores greater than 55. With this exception scores are satisfactorily distributed. Men obtaining scores equal



Graph III. Distribution of matrix test scores for three occupational groups of men

to or less than the 5 percentile score for the whole group were individually examined by a psychiatrist. There were amongst them a number of rather dull men who had worked under some emotional strain, had failed to grasp the nature of the test, or had shown marked inability to follow instructions. The large majority of the men with scores below the 5 percentile were found to be intellectually defective.

III. PERCENTILE RATING

Variation of ability with age

Scores for the individual test are throughout slightly lower than the equivalent scores obtained when the series is given as a group test. The

individual test involves a social factor which is different when a group test is given. That the difference is due chiefly to differences in the test situation and not to any error of standardization will be seen when Matrix Test individual scores are compared with Binet Intelligence Quotients.

Except where they approach the extremes of the scale the curves shown in Graph II are characteristically similar in form. Scores vary regularly with ability at any given age, irrespective of the position in the scale at which measurements are made. There is no evidence of serious plateaux or irregularities within the test series itself. Median scores show only slow development in eductive ability up to the age of eight years, a fairly rapid development of ability between the ages of eight and thirteen and that, at about the age of thirteen and a half years, innate eductive ability normally reaches its maximum development and remains relatively constant for the next fifteen years or more.¹ Compared with testees of median scores, the eductive ability of testees scoring above the median apparently begins to mature earlier in life, develops at a greater rate and continues to develop over a longer period of years. Maximum development appears to be reached some time after the age of fourteen, the amount of development which takes place after the age of fourteen being proportional to the amount of development which has taken place up to this age. The phenomenon is recognizable in the 95 percentile curve in spite of the fact that the end of the scale tends to prevent testees obtaining scores of more than 55. Conversely, curves for percentile points below the median agree in showing that, for testees of less than average ability, mental development is not only slower, ability to form comparisons and reason by analogy begins to appear later in childhood, continues to develop over a shorter period of years and has ceased to develop before the age of fourteen is reached. Similar results were obtained when the series was given to selected groups of testees (4). They indicate the probable relationship between the apparently conflicting opinions that innate mental ability normally matures at about the age of thirteen and a half years, that mental development may continue up to the age of sixteen years or more, and that, especially when working with dull children, allowance must be made for the 'late developers'.

Mental ages

The age at which the median score is equal to a testee's score may be taken as the testee's mental age. For the purpose of comparing and classifying testees the concept is useful but the rate of development of

¹ No data of equal reliability are as yet available for adults over thirty years of age. Such scanty data as exist suggest that ability to educe relations and correlates from immediate experience declines in later life.

eductive ability in childhood does not seem to be constant and Matrix Test mental ages should not be used like Binet Mental ages for the calculation of Intelligence Quotients.

Percentile grades

Percentile grades have the advantage over I.Q.s that they measure individual ability in terms of the frequency with which an equal degree of ability may be expected to occur in a random group of testees of the same age. A five point percentile grading has been found useful.

A testee can be classified as:

Grade I or *intellectually superior*, if his score lies at or above the 95 percentile score for testees of his own age group.

Grade II or *definitely above average*, if his score lies at or above the 75 percentile score for his age group.

Grade III or *mentally average*, if his score lies between the 25 and 75 percentile scores for his age group. III+ if his score is greater than the median; III- if his score is less than the median.

Grade IV or *definitely below average* if his score lies at or below the 25 percentile for his age group.

Grade V or *intellectually defective* if his score lies at or below the 5 percentile score for his age group.

The standard series does not differentiate clearly between individuals within these groups.¹

The five grades indicated can be differentiated with moderate accuracy. The 95 percentile score for adults is 55. This is the median score for the University men in the sample of militia men and 'intellectual superiority' may be regarded as that degree of eductive ability required for a person to pursue a successful University career or work of similar standing. Scores for the large majority of men following professional, clerical, or commercial occupations are Grade II or 'definitely above average' (Graph III). The scores for the bulk of skilled workers are Grade III or 'average'. Men with scores 'definitely below average' in mental ability are chiefly unskilled workers. The group of unskilled workers shows a rather high incidence of men with 'average ability'. Such data as are available suggest that these men, from social or temperamental causes, fail to maintain regular employment in one occupation.

Testees of Grade V fail in the test at just that stage where it is essential for success to form comparisons and reason by analogy. This ability has

¹ The relative capacities of individuals within a group are best estimated separately. Tests are at present being designed which will differentiate between individuals all of superior intellectual ability. Similar tests could be designed which would differentiate between individuals all intellectually defective.

apparently failed to develop. Testees of Grade V may therefore be described as 'intellectually defective' and 'intellectual defect' may be defined as *a permanent inability to form comparisons and reason by analogy*. It is diagnosed simply from test results. It does not necessarily occur in all persons certified as 'mentally defective', but it is probably the chief cause of social failure. It may exist in persons who, from good repetitive ability, make stable social adjustments. In such cases the certification of mental defect would be unwarranted but the diagnosis of intellectual defect would be justified.

IV. RELATIONSHIP BETWEEN MATRIX TEST PERCENTILE GRADE AND Terman MERRILL INTELLIGENCE QUOTIENT

The 1937 Terman Merrill Revision of the Binet Scale and Progressive Matrices 1938 were given to 301 people referred to London Psychiatric Clinics.¹ The standard deviation of the Terman Merrill I.Q. is given as 17 (7). I.Q.s can therefore be grouped in percentile classes comparable with the Matrix Test Grades. Table 5 shows the distribution of the Matrix Test Percentile Grades for each Terman Merrill I.Q. class.

Table 5. *Relationship between Matrix Test Percentile Grade and Terman Merrill I.Q.*

Matrix test		Terman Merrill I.Q.					Totals
Percentile	Grade	Under 73	Under 89	89-111	Over 111	Over 127	
95 and over	I	—	—	6	6	17	29
75 and over	II	—	3	20	19	14	56
Over 25 and under 75	III	2	20	85	15	10	132
25 and under	IV	9	23	16	—	—	48
5 and under	V	26	9	1	—	—	36
Totals		37	55	128	40	41	301

It is generally found that the incidence of exceptionally dull and also exceptionally bright children attending psychiatric clinics is greater than the incidence of such children in the general population. Both tests show this equally.

In 170 cases Matrix Test Grade agrees with Terman Merrill I.Q. class. In fifty-five cases the matrix score is one grade higher than the I.Q. class; in fifty-four cases it is one grade lower. In eleven cases the Matrix Test

¹ Children were seen at the London Clinic and the Victoria Hospital for Children. Children and adults were seen at the London Hospital, the Tavistock Clinic, and the West End Hospital for Nervous Diseases. Data from clinics concerned chiefly with the diagnosis of mental defect are not included.

score is two grades higher than the I.Q. class, in another eleven cases it is two grades lower than the I.Q. class. There are no differences of more than two grades between the two test results. If the Terman Merrill scale can be regarded as correctly standardized for English children, norms for Progressive Matrices as an individual test are approximately correct.

Discrepancies of two grades between the Matrix Test score and the Terman Merrill I.Q. are interesting:

Cases in which the Terman Merrill I.Q. is two classes higher than the Matrix Test Grade

A. A boy aged 10 years. He was referred for uncontrollable behaviour at home and for being totally ineducable at an ordinary school. His I.Q. was 137, but he was only Grade III according to the Matrix Test. He was verbally fluent. (Terman Vocabulary 18 : 12.)

B. A girl aged 12 was referred for nervous habits and mannerisms. She was excitable and talkative. She attended a private school. Her father was an alcoholic, her mother neurotic. Her I.Q. was 139, but her Matrix Test Grade III. (Terman Vocabulary 27 : 14.)

C. A boy aged 13 was referred for entresis of long standing. His I.Q. was 134, his Matrix Test Grade III. (Vocabulary 27 : 15.)

D. A boy aged 13½ was referred for fits following a bicycle accident. Epilepsy was not confirmed. His I.Q. was 139, his Matrix Test Grade III. (Vocabulary 29 : 15.)

E. A boy aged 13½ was referred for stammering, solitary habits, and inability to concentrate. His I.Q. was 135, his Matrix Test Grade III. (Vocabulary 25 : 15.)

F. A girl aged 15 was referred for stealing at school. She had gained a scholarship but was steadily doing worse work. She had reached the bottom of her class and was difficult to manage. Her I.Q. was 137, her Matrix Test Grade III. (Vocabulary 33 : 20.)

G. A boy aged 15 was referred for stealing at school and for being sullen and intractable at home. He attended a secondary school. He was abnormally studious. His work was inferior and he had become solitary and anti-social. His I.Q. was 140, his Matrix Test Grade III. (Vocabulary 36 : 20.)

H. A girl aged 16. She had won a scholarship to a secondary school but had been expelled for stealing. Her home life was unhappy. She was abnormally anxious to excel at school. Her I.Q. was 132, her Matrix Test Grade III. (Vocabulary 32 : 23.)

I. A girl aged 16½ was referred for inseminia. She showed strong home and social antagonisms. She talked excitedly. Her I.Q. was 132, Matrix Test Grade III. (Vocabulary 31 : 23.)

J. A man aged 20 had had a secondary education, he had secured excellent appointments but was totally unable to keep any sort of work. He was abnormally talkative, had marked sexual aberrations, and was a complete social failure. His I.Q. was 133, his Matrix Test Grade III. (Vocabulary 38 : 26.)

K. A boy aged 10 was referred for psychological examination before emigration. No case notes are available. His I.Q. was 96, his Matrix Test Grade V. (Vocabulary 10 : 11.)

Of this group: two of the younger children showed lack of self-control; one of the elder showed what was probably the early stages of a post-traumatic disturbance, while six had proved failures at school and had

reacted by anti-social behaviour or had become social failures since leaving school. With one exception they were all verbally fluent. They obtained high Terman Merrill I.Q.s. Several had won scholarships to secondary schools but progress had not been maintained. As a group they were characteristically talkative, superficially intelligent but excitable, unstable, lacking in self control. They shone on tests of the Binet type but did badly with more logical problems of the Matrix type. They were bright but lacked genuine intellectual ability.

Cases in which the Terman Merrill I.Q. is two classes lower than the Matrix Test Grade

L. A girl aged 12 was referred for enuresis. She was abnormally slow of speech and showed almost psychotic difficulty in grasping what was said to her. Her I.Q. was 61 but she was Grade III according to the Matrix Tests. (Vocabulary 8 : 14.)

M. A girl aged 10½, an orphan examined before emigration, was slow of speech and emotionally withdrawn. Her I.Q. was 90 but according to the Matrix Test she was Grade I. (Vocabulary 11 : 11.)

N. A boy aged 10½ was referred for speech and reading defects and also for night terrors. His I.Q. was 88, his Matrix Test Grade II. (Vocabulary 10 : 11.) His reading mental age was only 5.9 years.

O. A boy aged 9½ was referred for backwardness at school. His I.Q. was 87. His Matrix Test Grade II. (Vocabulary 10 : 10.) His reading mental age was 5.2 years.

P. A girl aged 9½ was referred for backwardness at school, for nervousness and inability to read. I.Q. 62, Matrix Test Grade III. (Vocabulary 6 : 10.) Reading mental age less than 4 years.

Q. A girl aged 9 was referred for sleep walking. I.Q. 85. Matrix Test Grade II. (Vocabulary 9 : 9.)

R. A girl aged 9, an orphan examined before emigration. She had come from Cornwall. She had difficulty in following instructions, her education had been interrupted, her I.Q. was 109, her Matrix Test Grade I. (Vocabulary 9 : 9.) Her reading mental age was 8.3 years.

S. A girl aged 8½ was an orphan with a history of four unsatisfactory adoptions and no regular education. Her I.Q. was 107, her Matrix Test Grade I. (Vocabulary 12 : 8.) Reading mental age 8.5.

T. A girl aged 7½ was referred for nervous grimacing after cessation of enuresis. She was shy and probably withdrawn. Her I.Q. was 102, Matrix Test Grade I. (Vocabulary 9 : 7.)

U. A boy aged 7 was referred as 'Query M.D.' with the following medical notes: 'His mother hates him and told me that she would do anything to get rid of him... and hoped to have him classed as subnormal mentally so that she might have him put away permanently—she tried to produce this condition in him by keeping him shut up in a bedroom. He was not allowed to go to school... any normal child subject to the treatment this boy has received would show the effects of it.'

His I.Q. was 96 but his Matrix Test Grade I. (Vocabulary 6 : 6.) Totally unable to read.

V. A boy aged 6½ was referred for night terrors. His I.Q. was 103, Matrix Test Grade I. (Vocabulary 7 : 5.) Totally unable to read.

The children were all young. Five were seriously backward in reading. Two had missed regular education, and two were abnormally slow of

speech. Emotional disturbances were not marked but frequently existed in the form of deep anxieties, fears or night terrors. They were unable to respond to tests of the Binet type but felt more confident with tests of the Matrix type in which they needed neither to talk nor to hurry.

The Terman Merrill Revision of the Binet Scale appears to depend for young children too much upon normal school accomplishments; and for older persons too much upon their acquired vocabulary and fluent speech. Tests of the Matrix type probably make slow, self critical, shy or shut in testees appear a little brighter and more intelligent than they are, while quick, impulsive testees are at a disadvantage.

It is sometimes assumed (probably from a resemblance between Matrix tests and tests of the Form Board or Picture Completion type) that the tests are useful chiefly in dealing with persons of inferior intelligence. Progressive Matrices quickly detect intellectual defect but one of the chief merits of the tests is the clarity with which they differentiate between genuine intellectual superiority and superficial brightness due to verbal fluency and quick recall. The clear explicit education of relations between perceived characters is as essential for continued scholastic success as good memory, verbal fluency, and quick recall. The value of Matrix tests for the detection of superior intellectual ability has not, as yet, been fully explored.

V. SUMMARY

A progressive series of sixty matrices has been prepared and standardized for general and clinical purposes. It can be used as an individual or group test for children over six years of age and for adults.

The series gives a five point percentile grading irrespective of the age of the testee. The standard series makes no attempt to differentiate individuals within these groups.

Matrix Test Grades have been compared with Terman Merrill I.Q.s. Case notes show that the series differentiates clearly between genuine intellectual superiority and verbal fluency. Used with defectives the chief merit of the test is that it will differentiate backwardness due to specific defects in reading, speech, or education from genuine intellectual defect.

From its original conception in some experiments of the writer in 1935 to the standardization of the series for adults in 1940, data have been steadily accumulating and thanks are due to the many people who have contributed to the work but especially to the writer's wife and to Dr L. S. Penrose. The work has been assisted by the Medical Research Council, the Darwin Trust, the Child Guidance Council, and the Royal Army Medical Corps.

REFERENCES

- (1) RAVEN, J. C. (1939). The R.E.C.I. series of perceptual tests: An experimental survey. *Brit. J. Med. Psychol.* **18**, 16.
- (2) — (1938). *Progressive Matrices*. London: H. K. Lewis and Co.
- (3) — (1940). *Progressive Matrices; Instructions, Key and Norms*. London: H. K. Lewis and Co.
- (4) RAVEN, J. C. & WAITE, A. Experiments on physically and mentally defective children with perceptual tests. *Brit. J. Med. Psychol.* **18**, 40.
- (5) RAVEN, J. C. (1940). Matrix tests. *Mental Health*, Jan.
- (6) HOGGEN, L. (1938). *Political Arithmetic*. Allen and Unwin.
- (7) TERMAN, L. M. & MERRILL, M. A. (1937). *Measuring Intelligence*. Harrap.