THE ABILITIES OF ADOLESCENT AND ADULT HIGH-GRADE MALE DEFECTIVES.

By J. TIZARD, M.A., B.Litt.,

Medical Research Council, Occupational Psychiatry Research Unit, Maudsley Hospital, Denmark Hill, London, S.E.5;

in collaboration with

N. O'CONNOR, M.A., and J. M. CRAWFORD, M.D., D.P.M.

[Received 1 July, 1950.]

CONTENTS.

I. INTRODUCTION.

DURING the first half of 1949 a battery of cognitive and personality tests was given to a sample of 104 adolescent and adult high-grade male defectives resident in Darenth Park, Dartford. The testing was carried out by N. O'Connor and the writer, in collaboration with Dr. J. M. Crawford, Deputy Superintendent of Darenth Park. The following report presents a summary of the cognitive and motor test findings, and discusses some of their implications.

In the Appendix are included factual and statistical data not included in the body of the report.

II. REVIEW OF LITERATURE BEARING ON INTELLIGENCE AND COGNITIVE ABILITIES OF HIGH-GRADE FEEBLE-MINDED ADULTS.

(a) *The " Upper Limit " of Defective Intelligence.*

The Mental Deficiency Act of 1927, which states that " mental defectiveness means a condition of arrested or incomplete development of mind existing

https://doi.org/10.1192/bjp.96.405.889 Published online by Cambridge University Press

before the age of 18 years," lays down no criteria by which mental development may be assessed. In principle, however, it is generally agreed that before an individual can be judged to be mentally defective he should be shown to be both cognitively defective, and in need of care, supervision, and control. A few individuals, it is true, are certified on " moral " or social grounds ; but in such cases defects of personality or aberrations in conduct are added to deficiency in intelligence. Individuals of normal or even dull intelligence are not usually certified when they break the law or offend against the social code. Cognitive ability is usually assessed by mental tests.

In mental deficiency practice the early work of Goddard (1914), Terman (1916), Burt (1921) and others following up the pioneer researches of Binet suggested certain critical levels in intelligence, below which individuals might be considered cognitively defective. The usefulness of the Binet-Simon test in enabling psychologists to differentiate between normal and defective children led some to define mental defect exclusively in terms of scores on this test. Thus, Terman stated that, in the diagnosis of mental defect,

> ". . . nothing else is as significant as the I.Q. All those who test below 70 I.Q. by the Stanford revision of the Binet-Simon scale should be considered feeble-minded, and it is an open question whether it would not be justifiable to consider 75 I.Q. as the lower limit of 'normal' intelligence. Certainly a large proportion falling between 70 and 75 can hardly be classed as other than feeble-minded, even according to social criteria."

Although psychologists would not to-day give unqualified acceptance to such an extreme statement of the psychometric definition of mental defect, none the less, in psychiatric writings, the concept of 70 I.Q. is still widely accepted as marking the upper limit of defective intelligence. Henderson and Gillespie (1944) state that an I.Q. of less than 70 (on the 1937 Revision of the Stanford-Binet test) " corresponds roughly to clinical feeble-mindedness, and of below 40 to imbecility." Penrose (1949) repeats the traditional classification of grades of mental defect in terms of " Binet intelligence test level," though he recognizes its limitations. (According to this, the feeble-minded have, as children, I.Q.'s of 50–69, and as adults, mental ages of 7–10 ; imbeciles have, as children, I.Q.'s of 20–49, and as adults, mental ages of from 3–6 ; idiots have I.Q.'s under 20, or mental ages under 3 years.) Curran and Guttmann (1946) state that " very special reasons must be present before it is possible to certify as mentally defective an individual with an I.Q. above 75." Similar statements could be quoted from other text-books which include chapters on mental deficiency.

A notable opponent of this point of view is Tredgold, who has always maintained that social adaptation, and not cognitive ability, is the decisive factor in deciding whether or not an individual is to be deemed feeble-minded. However, as Porteus (1941) points out, though Tredgold rejects the intelligence quotient as the basis of a satisfactory definition, yet assessment of the kind and degree of incomplete mental development used by Tredgold as criterion " must certainly be affected by the use of mental tests."

Summing up the evidence, it might be said that over the last 40 years psychiatric opinion has inclined to the opinion (i) that intelligence tests are necessary for diagnostic purposes in all but the most severe forms of mental deficiency; (ii) that for the feeble-minded, as opposed to imbeciles and idiots, intelligence is only one factor to be considered, deficiencies in temperament and character being equally important, and in some cases of decisive importance; (iii) that although dull (I.Q. 70–85), or even normal (I.Q. 85 plus) individuals may be certified, only "very exceptional circumstances" will justify such certification. One would, in short, expect to find very few individuals in M.D. institutions with I.Q.'s above 75.

(b) *Other Abilities of Defectives.*

The use of intelligence tests as diagnostic instruments has deflected interest from other cognitive and motor abilities and their assessment. It is well known that defectives tend to be clumsy, to lack fine sensory discrimination, and to be slow in movement. A review of studies dealing with these "special abilities" has been published elsewhere (Tizard and O'Connor, 1950), and the evidence will not be repeated here. It may, however, be said that little systematic study of the abilities of defectives which are not highly correlated with test intelligence, has been made, at least in recent years, using adequately standardized tests.

It was to throw some light on the distribution of abilities in high-grade adult defectives (i.e. over 16 years old), so that at a later date the tests might be used for vocational guidance purposes, that a testing programme was carried out in Darenth Park. In this paper we shall be concerned only with scores on tests of cognitive and motor abilities and their intercorrelation, among what we believe to be a fairly typical sample of high-grade adult and adolescent male defectives recently admitted to an institution in the London area.

III. Empirical Findings with Adult Defectives.

Five Intelligence tests, 12 of the 15 subtests devised by the United States Employment Service for vocational guidance purposes, and a Rail-walking test, devised by S. Roy Heath, jun., as a test of general locomotor co-ordination, were given to 104 adult high-grade males in the institution. The average age was nearly 21 years; only 3 patients were over 30 years of age, and the youngest patients were nearly 16. Testing occupied several months.

Results of the testing are presented below, Intelligence test results being presented in Section III (a), other cognitive test results being presented in Section III (b).

(a) *Intelligence Test Findings.*

Tests Given.

The following tests were given:

Kohs Blocks, Alexander Version.
Progressive Matrices, 1938, untimed.

Binet Vocabulary subtest of 1937, Terman Merrill Revision, Form L.
Porteus Maze test, Vineland Revision.
Cattell Non-Verbal Intelligence test. Form I.B.

Some relevant information regarding the method of administration of these tests is presented in the Appendix.

Results.

Tables of results are presented below. In Table I are given frequency distributions of the I.Q.'s on each of the tests, together with the mean and median I.Q.'s and the standard deviations. Table II gives the intercorrelations between the tests, as well as the means and standard deviations of raw scores on all of the tests included in the battery.

Table III presents the number of patients with one or more I.Q.'s above 75 on each of the five tests.

The following points are selected for discussion :

(i) *The Distribution of I.Q.'s.*

TABLE I.—*Frequency Distributions of I.Q.'s on Five Intelligence Tests not Corrected for Differences in S.D.* $N = 104$.

I.Q.	Kohs.	Matrices.	Binet Voc.	Porteus I.Q.	Cattell.*
Median I.Q.	75·36	72·33	71·10	85·69	65·10
Mean I.Q.	75·41	74·55	71·38	82·56	63·78
S.D. of I.Q.'s	17·00	12·73	14·89	22·00	14·07
120+	1	0	2	3	0
110—119	2	1	0	11	0
100—109	5	2	2	4	3
90—99	13	12	5	18	1
80—89	23	16	11	29	4
70—79	20	33	40	13	22
60—69	15	28	23	8	43
50—59	19	10	18	10	15
Less than 50	6	2	3	8	16

* Published figures by Cattell (3) give the standard deviation of this test as between 20 and 25 points. A " corrected " I.Q. obtained by reducing the standard deviation from 25 to 16 points gives a median I.Q. of 72·08 and a mean of 73·42.

Table I shows that there were considerable differences in the comparative difficulty of the different tests. The Porteus Maze test, on the one hand, with a mean I.Q. of 82, was relatively easy ; the Cattell test, on the other hand, with a mean I.Q. of 63, was relatively difficult. The other three tests, Kohs Blocks, Matrices, and Binet Vocabulary, lay in between, giving mean I.Q.'s between 71 and 75.

(ii) *The Intercorrelations Between Tests.*

In Table II are given the intercorrelations between all of the tests included in the battery. An inspection of the table of correlations shows that, whereas

https://doi.org/10.1192/bjp.96.405.889 Published online by Cambridge University Press

TABLE II.—*Correlations Between* 18 *Cognitive Tests, Together with First Factor Loadings and Communalities* (h^2) $N = 104$. (*Decimal Points Removed.*)

Std. dev. (S)	18·90	9·36	5·37	21·67	17·14	6·31	10·49	4·44	6·58	30·85	5·71	10·18	7·82	11·68	11·28	4·20	3·90	30·01
Mean score	12·59	24·20	12·96	82·56	48·98	14·48	25·95	6·10	11·10	97·51	10·88	48·77	13·79	75·40	79·92	20·70	21·14	82·89
Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Kohs 1	..	65	−06	54	62	59	30	34	61	36	59	28	63	34	31	39	44	40
Matrices 2	..	..	−04	52	61	50	31	39	47	26	48	29	57	36	34	38	40	24
Binet Voc. 3	..	..	..	−07	−01	−13	−12	14	−16	−02	−17	−16	−18	−07	−12	−15	−05	02
Porteus I.Q. 4	..	..	..	..	62	43	23	31	41	24	38	34	55	34	34	35	47	36
Cattell 5	..	..	..	..	..	71	33	44	64	33	50	51	65	36	37	45	59	48
G.A.T.B. A 6	..	..	..	..	..	..	33	36	67	34	54	47	65	34	26	35	45	27
,, C 7	..	..	..	..	..	..	..	13	37	43	30	56	41	41	40	44	52	26
,, D 8	..	..	..	..	..	..	..	..	29	10	36	24	27	08	03	06	21	−02
,, F 9	..	..	..	..	..	..	..	..	..	37	71	48	68	37	33	36	48	34
,, G 10	..	..	..	..	..	..	..	..	..	..	30	52	34	40	29	35	43	31
,, H 11	..	..	..	..	..	..	..	..	..	..	..	33	56	29	24	32	34	20
,, K 12	..	..	..	..	..	..	..	..	..	..	..	..	49	40	49	45	55	17
,, L 13	..	..	..	..	..	..	..	..	..	..	..	..	..	51	47	50	52	35
,, M 14	..	..	..	..	..	..	..	..	..	..	..	..	..	..	72	65	64	46
,, N 15	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	64	62	46
,, O 16	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	71	44
,, P 17	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	53
Rail test 18	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
First factor loading	72	66	−12	62	80	71	56	32	75	51	64	64	81	65	62	67	77	49
Communality (h^2)	52	44	−02	38	64	51	31	10	56	26	41	42	66	42	38	45	60	24

Calculations were made to 3 decimal places, and have been rounded off to 2 decimal places.

all except the Binet Vocabulary correlate positively and have, in consequence, significant positive loadings on the first common factor extracted from the battery, scores on this test failed to correlate significantly with *any* of the other cognitive tests. This result was not expected by the investigators, since, as is well known, Vocabulary scores are commonly considered to have high *g* loadings, and, indeed, have been recommended as " the best single test that can be given " to cases suspected of mental deficiency (Curran and Guttmann, 1946).

(iii) *Our Test Results Compared with those of Certifying Officers.*

A third finding which emerged very clearly from the investigation was that both our test results, and those of the psychiatrists in the institution in which the testing took place, indicated that many of the patients were much more intelligent than one would have been led to believe from reading their certification orders. No evidence on this point will be presented in this paper, since many certification orders do not indicate how the diagnosis was arrived at, or if they give I.Q.'s, do not state the test from which the I.Q. was derived. But many examples could be given of patients whose behaviour, and whose test findings alike, indicated a degree of cognitive efficiency apparently undetected by the medical officers signing their certification papers.

The number of cases with I.Q.'s above 75 on one or more of the tests, together with the numbers for each test considered separately, is presented in Table III.

TABLE III.—*Number of Patients with One or More I.Q.'s Above 75 on Five Intelligence Tests.*

	Total No. in each category.	No. in each category for each test.				
		Kohs.	Matrices.	Binet Voc.	Porteus I.Q.	Cattell.
No. of S's with I.Q.'s above 75 on 5 intelligence tests . . .	5	5	5	5	5	5
No. of S's with I.Q.'s above 75 on 4 intelligence tests. . .	9	9	9	5	9	4
No. of S's with I.Q.'s above 75 on 3 intelligence tests. . .	24	21	17	10	23	1
No. of S's with I.Q.'s above 75 on 2 intelligence tests. . .	26	16	3	9	24	0
No. of S's with I.Q.'s above 75 on 1 intelligence test . . .	19	1	2	7	9	0
Total number of cases with I.Q. on one or more tests above 75 .	83	..	..	..	..	..
No. of I.Q.'s greater than 75 on each test.	..	52	36	36	70	10

(b) *Other Test Findings.*

Tests Given.

As mentioned in Section I, much less attention has been paid by psychologists and psychiatrists to the devising and standardizing of tests of abilities which are not considered to be closely associated with general intelligence,

than to the construction and standardization of intelligence tests. Consequently, in attempting to find tests by means of which the distribution of these abilities could be compared with that of a normal population, we were handicapped by the comparative absence of reliable and adequately validated tests. Many tests, such as the Oseretsky (Doll, 1936), Van der Lugt (1948), and Yarmolenko (1933) scales of motor proficiency, had to be abandoned, either because the material was unavailable, or because the standardization was felt to be too inadequate to make possible a comparison between our results and those of other workers dealing with normal adults, or with children. The battery finally decided upon was the United States Employment Service General Aptitude Test Battery, which has been soundly standardized on American industrial workers (27), and Heath's Rail-Walking test of locomotor co-ordination, which was standardized on American Army recruits (Heath, 1943, 1944), and has been used for testing defectives at Vineland Training School, New Jersey (Heath, 1942, 1946). A brief description of the tests included in the General Aptitude Test Battery, and of the method of administration and the standardization of the Rail-Walking test, is included in the Appendix. All it is necessary to say here is that eight of the twelve subtests of the General Aptitude Test Battery were paper-and-pencil tests, while the other four were apparatus tests (peg-board and pin-board tests). Scores on the paper-and-pencil tests are combined so as to give a weighted score with a mean of 100 and a standard deviation of 20 (for a normal adult population) on each of four aptitudes which the tests purport to measure. These aptitudes are: Spatial Aptitude, Form Perception, Aiming or Hand-eye Co-ordination, and Motor Speed.

The apparatus test scores are likewise combined to give scores purporting to measure Finger Dexterity and Manual Dexterity.

Results on Other Cognitive Tests.

A frequency distribution of the aptitude scores on each of the aptitudes which the tests purport to measure is presented in Table IV. The scores on the Rail tests are raw scores. The mean score on Heath's army population was 130, and the standard deviation 20 points. The scores for normal subjects on this test are markedly skewed, the maximum score obtainable being 153 points.

What emerges from the tabular presentation of scores is the following: (i) On tests of Spatial Aptitude the median score for this sample is just over one standard deviation below the population mean for normal adults; (ii) the median score on the tests of Form Perception is considerably lower, falling about $1\frac{1}{2}$ standard deviations below the population mean, as does the score for Aiming, or Hand-eye Co-ordination; (iii) the median scores for the sample on tests of Finger and Manual Dexterity fall about 2 standard deviations below the population mean; (iv) the median score on the tests of Motor Speed falls nearly $2\frac{1}{2}$ standard deviations below the population mean.

It would seem from this evidence that the mentally defective subjects had higher median scores on the more abstract tests of Spatial Aptitude and

Form Perception than on the tests demanding precision of movement, manual and finger dexterity, and, especially, motor speed.

TABLE IV.—*Frequency Distributions of Aptitude Scores on the General Aptitude Test Battery, and of Scores on Heath's Rail-Walking Test of Locomotor Co-ordination.*

Aptitude	S.	P.	A.	T.	F.	M.	Rail.
Median	79	69	67	54	62	61	84
Aptitude score.							
141+	..	0	0	0	0	1	4
121—140	1	0	1	1	0	0	9
101—120	10	7	10	4	2	3	13
81—100	38	34	14	9	19	19	32
61—80	44	25	37	24	33	29	21
41—60	11	29	27	37	32	26	17
21—40	0	7	12	24	17	16	6
0—20	0	2	3	5	1	10	2

(Mean score for an American industrial population on all G.A.T.B. tests = 100; S.D. = 20. On the Rail test the mean score for 1,013 soldiers was 130 points; median 140; S.D. 20)

Aptitudes mentioned in the table are described in the U.S.E.S. Guide (28) as follows:

S. *Spatial Aptitude.*—Ability to comprehend forms in space and to understand relationships of plane and solid objects.

P. *Form Perception.*—Ability to perceive pertinent detail in objects or in pictorial or graphic material. Ability to make visual comparisons.

A. *Aiming or Eye-hand Co-ordination.*—Ability to co-ordinate eyes and hands or fingers accurately so as to make precise movements with speed.

T. *Motor Speed.*—Ability to make hand movements, such as tapping rapidly. Ability to make a movement response swiftly.

F. *Finger Dexterity.*—Ability to move the fingers, and manipulate small objects with the fingers, rapidly or accurately.

M. *Manual Dexterity.*—Ability to move the hands easily and skilfully. Ability to work with the hands in placing and turning motions.

(c) *Summary of Empirical Findings.*

The following points, discussion of which will be taken up in the next section, emerged from this analysis of scores on 18 cognitive tests given to a sample of 104 young adult, high-grade, institutionalized male defectives. (i) There were discrepancies of up to 20 points between the mean and median I.Q.'s got on one intelligence test (Cattell's) and those got on another (Porteus Maze), while those on three other tests lay in between. (ii) Intercorrelations between all of the cognitive tests were positive, and almost without exception significant, except for those between the Binet Vocabulary and other tests, none of which were significant. Sixteen of these were in fact negative. (iii) Our test findings, like those of the psychiatrists in the institution in which the testing took place, indicated that many of the defectives were more able than would have appeared from a reading of the reports made by some certifying officers. (iv) The General Aptitude Test Battery scores suggested that the defectives as a whole were relatively less retarded in Form Perception and in Spatial Aptitude than they were in Hand-eye Co-ordination, Manual Dexterity, Finger Dexterity, and Motor Speed.

IV. Discussion of Test Findings.

(a) *Factors Influencing the Size of Intelligence Quotients.*

As mentioned in Section I, common psychiatric opinion inclines to the view that an I.Q. of 70 to 75 marks the upper limit of mentally defective intelligence. Our own results do not support this belief. From Table II it can be seen that the number of patients with I.Q.'s above 75 varies enormously from test to test. Moreover, even on the most difficult test some 10 per cent. of the sample scored over 75 points of I.Q., even when these I.Q.'s remain uncorrected for size of standard deviation.*

In order to throw light on these findings something will be said about the nature of I.Q.'s, since considerable confusion on this point exists in mental deficiency practice.

(i) The value of an I.Q. depends, among other things, on the standard deviation of I.Q.'s given by the test. In the early Goddard, Terman, and Burt Revisions of the Binet test, from the use of which the concept of I.Q. levels first gained currency, the standard deviation was believed to be about 12 or 13 points. Hence, an I.Q. of 70 was believed to fall about 2½ standard deviations below the mean, and so cut off only about 1 per cent. of the population who were regarded, partly on other grounds, as being mentally defective.

On some more recently devised tests, however, such as Cattell's test, which was included in our battery, the standard deviation is 20, and possibly 25 points (Cattell (1948) gives both figures). Hence, an I.Q. of 70 falls not much over one standard deviation below the mean, and so cuts off about 10 per cent. of the population. Clearly differences in standard deviation are of great importance.

It should be noted that the size of the standard deviation is of especial importance as far as the I.Q.'s of defectives are concerned.

Unfortunately differences in and uncertainties regarding the size of standard deviation are found in even the most elaborately standardized tests, such as the 1937 Revision of the Stanford Binet test. It is significant that though the 1937 Revision is said to be especially useful in diagnosing cases of mental defect, the standard deviation varies from 13 points at age 6 to 20 points at age 12, so that an I.Q. of 74 at age 6 is equivalent to one of 60 at age 12. How little attention has been paid to this can be gathered from the fact that though Terman and Merrill note these discrepancies, they do not make corrections for them in the Tables of I.Q.'s given at the back of their Manual. (These corrections have, however, been published by McNemar (1942) in his statistical discussion of the test.)

One consequence of this is that standards of certification based on this, or similarly standardized tests, are bound to differ from one age to another—for purely statistical reasons—unless tables of corrected I.Q.'s are used.

A further consequence is that any attempt to assess the incidence of mental

* When the I.Q.'s are corrected to a standard deviation of 16 points, nearly 50 per cent. had I.Q.'s above 75 on this test.

deficiency in psychometric terms, using this test, is likely to give results which are seriously misleading.*

(ii) If we could be sure that the standardization sampling adequately represented individuals of all levels of ability, and that the distribution of scores was normal, corrections in the size of the standard deviation could easily be made. In fact, however, it is rare that the assumption of representative sampling and normality of distribution can be made in psychological testing. It is worth pointing out in this respect that the 1937 Stanford Binet test, despite the presence of "Average Adult" and "Superior Adult" items, has never been standardized as an adult intelligence test. In the standardization sample only 50 boys and 50 girls for each of the ages 15 to 18 were included. Of these, moreover, nearly all were still at school. From the table presented by Terman and Merrill (1937) it can be seen that of "subjects who did not complete high school," only 4 aged 16, 12, aged 17 and 11 aged 18 were included in the standardizing sample. Though subjects who did not complete high school are likely to include the vast bulk of the dull and defective, for whom the test is supposed to be especially useful, it appears that the standardization sample included only 27 such subjects, spread out over a 3-year age range. It is, of course, impossible to correct for such sampling inadequacies in the manner adopted by Terman and Merrill.

This means that the standard deviation of I.Q.'s on this test for older children and adolescents is almost certainly greater than the 18–20 points given by the Stanford investigators. It means in consequence that one is bound to *under*-estimate the intelligence of older children and of adults and adolescents who deviate markedly on the mean for this test. Ten per cent., possibly more, of the adult population would have I.Q.'s under 70 if tested by the Binet test.

(iii) Similar criticisms could be levelled against other tests for children which provide tables of norms for adults, since their standardization in most cases is weak. The fact is that there is no intelligence test which has been standardized so as to make possible an accurate estimation of the I.Q. of subnormal English adults. The Wechsler Bellevue test has been standardized for an American population, but not for an English one; and the face validity of some of the items in the Performance Scale, as far as an English population is concerned, seems to suggest the need for restandardization.

(iv) Owing, then, first to differences in the reported standard deviations of intelligence tests, and secondly to the likelihood that the estimations of the

* The statistical character of the I.Q. was pointed out by Burt (1921) in setting his thresholds for defective and normal intelligence. Penrose (1949) also deals briefly with this point, but fails to note the differences in the standard deviation of the I.Q.'s of the 1937 Terman Merrill Revision for different chronological ages, or to suggest that the standard deviation of the I.Q. for adults is probably at least 20 points. Penrose also continues to advocate the use of the concept of "mental age," despite the destructive criticism of this concept given by Wechsler (1944).

That the standard deviation of I.Q.'s on the 1937 Revision differs from one age to another for statistical and not for psychological reasons, is explained by McNemar (1942) in his discussion of the standardization of the test. McNemar does not, however, stress the inadequacies of the standardization sampling which makes the norms inapplicable to a more heterogeneous sample of the population. This gives to I.Q.'s more than one standard deviation below the mean an appearance of accuracy which is quite spurious, owing to the small number of cases on which such norms are based.

I.Q. of cases which differ markedly from the mean have been based on very few cases, the selection of which has been haphazard or systematically biased, it is virtually impossible to make a strict comparison between I.Q.'s got on different tests. The possibility of doing so is still further decreased if we abandon the assumption that intelligence tests are testing a single function of inborn, all-round mental ability. For once we abandon a single-factor theory of intelligence there is no reason why mean scores of defectives on different tests should not differ considerably, just as the mean scores on different Aptitude tests in the General Aptitude Test Battery differed considerably. We might then go on to explain such differences in psychological terms.

(v) Faced with results such as ours, some investigators have, in fact, proceeded to explain them in psychological terms, without considering statistical and selection factors which make their conclusions invalid. For example, it is sometimes stated that defectives tend to get much poorer scores on group tests than they do on individual tests, where they are able to be given personal supervision and appropriate encouragement. Alternatively, one may postulate that many so-called defectives are deficient only in verbal ability, or educational ability, but may have high or low practical ability, which, according to this theory, is only moderately correlated with verbal ability. Such a hypothesis has been put forward by Duncan (1942), and has gained the support of W. P. Alexander (1946) following an examination of the test scores of E.S.N. school-children on the Binet and Alexander tests.

One or both of these hypothesis may well be true, and we had hoped, at one time, to be able to make a rigorous examination of them. So little information was available as to the standardization of the tests, and the distribution of extreme scores, that comparisons between one test and another would have been at best of doubtful value.

(vi) If, however, we assume that Cattell's published figure of 25 points represents the "true" standard deviation of I.Q.'s for an adult population on this test, we can correct this to 16 points so as to make Cattell's scores comparable with those achieved on other tests. If we make this correction, the results on this test now support the findings of Binet Vocabulary, Kohs, and Matrices. The mean and median I.Q.'s on all tests are above 70 I.Q.

Hence, though the evidence is not conclusive, it suggests that, according to our results, between a quarter and a half of our subjects regarded as feeble-minded could more properly be classed as dull or subnormal rather than cognitively feeble-minded. Both psychological and statistical factors make discussion of this finding difficult.

(b) *Vocabulary Scores.*

As noted in Section III(a) above, the correlations between Binet Vocabulary scores and other cognitive test scores were effectively zero, although all other tests included in the battery correlated positively, and had, in consequence, significant loadings on the first common factor. In view of the great importance which attaches to Vocabulary tests in mental deficiency practice, it was felt that some explanation of this finding was demanded.

One possible explanation was that since the remainder of the tests were non-verbal tests, the general factor extracted could not be identified with Spearman's *g*, but was, instead, a complex non-verbal, practical factor unrelated to verbal intelligence, of which Binet Vocabulary is considered to be a good test. This explanation undoubtedly takes us part of the way towards explaining our results. However, lack of statistical information about the distribution of test scores in an unselected sample, and their intercorrelation, makes us hesitate to explain our results by means of an untested psychological theory.

An additional difficulty comes from the selection process through which mental defectives pass before being committed to institutions. One of the most frequently used tests is the Binet Vocabulary test, or some equivalent. Inability to define common words is commonly thought to be a significant indicator of mental deficiency.

Since this is so it follows that individuals with low Vocabulary scores are more likely to be considered defective than are individuals with high Vocabulary scores, irrespective of their other abilities. If, as appears often to be the case, few if any other standardized tests are given, Binet Vocabulary becomes practically the sole psychometric arbiter of intelligence. In consequence, Binet Vocabulary scores are effectively partialled out, or would be, were it not for the fact that individuals with high Vocabulary scores are sometimes admitted to mental deficiency institutions. Such individuals, however, tend to be so grossly incompetent in other respects that their high verbal ability is not considered a sufficient reason for excluding them from M.D. institutions.

In so far as this argument is valid, it helps to explain why it was that scores between Vocabulary and other intelligence tests correlated negatively. Plotting Vocabulary and other intelligence test scores showed that a few individuals with high Vocabulary scores did, in fact, tend to score below the mean on other tests, and thus to account for the slight negative association between Vocabulary and other cognitive test scores.

We cannot, however, from our results, draw any conclusions as to the relationship between Vocabulary and other cognitive test scores of an unselected sample, even if composed of individuals of a restricted range of ability. For, as R. L. Thorndike (1949) has pointed out in a recent book on personnel selection, " When selection is based, as it often is, on a clinical judgment, which combines in an unspecified and inconstant fashion various types of data about the applicant, and when this judgment is not expressed in any type of quantitative score, one is at a loss to know how to estimate the extent to which the validity coefficient for any test procedure has been affected by that screening."

None the less, we believe that more research into the association between low verbal ability and other cognitive ability needs to be carried out. We feel, too, that Binet Vocabulary, as an index of cognitive deficiency, should be treated with some caution.

(c) *Our Test Results Compared with those of Certifying Officers.*

The discrepancies between our test results on the one hand, and on the other hand, the test scores recorded by certifying medical officers, has been remarked on

in Section III. These discrepancies are believed to be related to the fact that owing to increases in the standard deviation of the 1937 Terman Merrill Revision of the Stanford Binet test, the I.Q.'s of subnormal adults when tested by this test are likely to be lower than I.Q.'s got from earlier tests upon which I.Q. thresholds are based.

It is believed, too, that the extensive and sometimes exclusive use of the Binet Vocabulary test contributes to an underestimation of the mental ability of some defectives. Very possibly other factors, such as emotional disturbance at the time of testing, actual changes of I.Q. owing to growth, and, of course, the part which factors other than cognitive ones play in certification, all contribute to cause a writing down of the I.Q. on the part of certifying officers. Whatever the reason, the fact remains that certifying officers probably certify as defective more dull or subnormal adults than they record on their certification orders.

This would not matter greatly if the dull and often unstable adults, who are at present certified as mentally defective, were able to be given the care, supervision and control they need, in institutions and homes designed to deal with just this type of patient. Unfortunately this is by no means the case to-day. There are few institutions in the country which do not include in their population both low-grade custodial cases, and those high-grade subnormal adults whose only need is usually for temporary protection against society, and for a certain amount of guidance and control.

At present, in large institutions, such patients are usually placed in separate wards and are given a different type of occupation and treatment, so far as this can be arranged. It is not possible always to segregate these patients completely from low-grade cases. The effect, both on the patients themselves and on their relatives, who are sometimes shocked to find their quite "normal" children in an institution in which many low-grade imbeciles and idiots can be seen, is likely to be traumatic.*

A desirable experiment in the field of mental deficiency practice would be the setting up of more institutions in which only high-grade cases were placed. Here training programmes designed explicitly for high-grade adults could be put into operation. This need is widely recognized to-day; but it has yet to be fulfilled through appropriate administrative and legislative action.

(d) *Discussion of Results on the General Aptitude Tests.*

(i) The results of the General Aptitude Test Battery aptitude scores suggest that in the cognitive field tasks demanding motor co-ordination and precision of movement are more difficult for defectives than are tasks demanding form and space perception.

This finding is, however, put forward with some hesitation. We have little information regarding the distribution of the aptitude scores for the United States standardizing sample. It is possible that for scores which

* Dr. Crawford emphasizes that some feeble-minded or subnormal women in mixed wards show a sense of responsibility towards and interest in low-grade patients which gives them a feeling of confidence and an outlet for affection. In such cases it is the effect on the relatives rather than the patients which is likely to be unfortunate.

fall more than about one standard deviation below the mean, the distribution is no longer normal. The median performance of our sample might, in consequence, be relatively no worse on the Sensori-motor and Speed tests than on the Space and Form Perception tests. The field is clearly one in which more research needs to be carried out.*

(ii) *Significance of Results for Vocational Guidance.*

The United States Employment Service has put out its General Aptitude Test Battery for vocational guidance purposes. If we accept the norms given by them it would seem that, largely because of the comparative excellence of the Form Perception scores, 34 per cent. of the sample could be regarded as suitable for jobs falling into one or more of the occupational categories for which these tests are designed to select workers. These occupational categories do not include simple manual and labouring work, but comprise semi-skilled and skilled jobs.

The writer would be very happy with this finding were it not for the fact that in the majority of the cases the type of work for which they would qualify would be that designated by the United States Department of Labour as "Close Visual Inspection" or "Simple Visual Inspection." It may be doubted whether, on the basis of these test results, defectives should be drafted into work of this type, rather than into work demanding simple motor movements, for which, to judge from these findings, they are relatively less well equipped. Clearly more research needs to be done before the results of such tests as these can be used with confidence for vocational guidance purposes.

V. Conclusions and Summary.

1. Literature bearing on the intelligence of mentally defective adults is reviewed. It is concluded that psychiatric practice still accepts a Binet I.Q. of between 70 and 75 as marking the upper limit of mentally defective intelligence, despite the fact that the versions of the test upon which these standards were originally based are no longer in use. The 1937 Stanford Revision of the Binet test, which has replaced them, has greater standard deviations, so that about ten times as many adults would have I.Q.'s as low as 70 on this as had I.Q.'s of 70 on older versions of the test. The use of the Binet test as an adult intelligence test is criticized, and it is pointed out that no test at present in use can be considered satisfactory; the Wechsler Bellevue test, which is the best individual adult intelligence test available, has been standardized for an American, but not an English population.

* It is of some interest that Eysenck (1947) found that Speed tests and Sensori-motor tests were among the tests which gave the most significant differences between scores of normals and neurotics. In other words, these discrepancies we found in our sample might be due to selective sampling on both intelligence and neuroticism, with the stress on the latter. An attempt to check this hypothesis has been made by O'Connor (16, 17), who correlated these and other test scores with a psychiatric rating of the defectives comprising this sample.

Correlations of 0·295, 0·237, 0·292, and 0·233 were found between the Rail-Walking test, the Peg Board, the Pin Board, and progressive Matrices on the one hand, and a four-point rating of Stability.

The concept of "the" intelligence quotient is criticized, and difficulties in the way of deciding whether defectives are relatively more retarded in some respects (e.g. verbal intelligence) than in others (e.g. practical ability) are discussed.

2. The inadequacies of our present knowledge regarding cognitive abilities of defectives which are believed to be not highly correlated with intelligence are pointed out.

3. Results of a battery of 18 cognitive tests given to a sample of 104 high-grade adolescent and adult male defectives, resident in an M.D. institution, are presented and discussed. The following results were found:

(*a*) The mean and median I.Q.'s in five intelligence tests were, when corrected for differences in standard deviation, all over 70 points.

(*b*) The correlations between the tests were positive and significant, except for those between the Binet Vocabulary and other tests, which were not significant.

(*c*) Our research gave I.Q.'s which were in many cases considerably higher than one would have expected to find from reading reports made by certifying officers about the cognitive abilities of the patients. (No evidence for this statement was presented.)

(*d*) The test scores on the United States Employment Service General Aptitude Test Battery which was given to the patients showed them to be less retarded in Form Perception and Spatial Aptitude than in Hand-eye co-ordination, Manual Dexterity, Finger Dexterity, and Motor Speed.

4. Difficulties in the interpretation of these results are discussed. The use of the General Aptitude Test Battery for vocational guidance purposes is questioned.

5. There are probably more dull or subnormal (rather than cognitively *defective*) adults in M.D. institutions than appears to be generally realized to-day.

6. The separation of high-grade from low-grade cases, by placing them in institutions designed to deal explicitly with just this type of patient, is urged.

Acknowledgments.

Our thanks are due to Dr. J. K. Collier Laing, Medical Superintendent of Darenth Park, for his co-operation and for the facilities which he and his staff made available at Darenth Park.

Our thanks are due, too, to the United States Employment Service for a gift of 200 copies of the paper-and-pencil tests of the General Aptitude Test Battery for research purposes, and to Professor A. Lewis, Dr. S. Crown and Dr. H. J. Eysenck for criticisms of this paper.

Appendix.

(a) *Sample.*

The sample consisted of 104 consecutive male admissions, over 16 years of age, to Darenth Park mental defective colony, over the preceding 3 years. Not included in the sample were (i) old men and those with a mental age of

less than 7 years (I.Q. less than 50); (ii) those who had been transferred, discharged, or sent on licence before being tested; (iii) those suffering from gross complicating physical disabilities. The selection of patients was made by Dr. J. M. Crawford, Deputy Superintendent of Darenth Park: 104 patients took all the tests. The mean chronological age was 20·9 years, standard deviation 4·6 years. The mean I.Q.'s and standard deviations of I.Q.'s on five intelligence tests are given in Table I.

(b) *Administration.*

Tests were administered by N. O'Connor and the writer between January and June, 1949. Eighteen cognitive tests were given. Testing was carried out in the institution. Group tests were given to small groups of from 4 to 8 patients. Each patient spent between 4 and 5 hours taking the tests, which included objective tests of personality not discussed here.

(c) *Tests Given.*

Tests given are listed below. Tests (1) to (5) are standard intelligence tests widely used in this country. Tests (6) to (17) are subtests of the United States Employment Service General Aptitude Test Battery, copies of which were given to us by the United States Government for research purposes. Test (18) is a Rail-Walking test which has been used in the United States both with defectives and with soldiers.

The battery of cognitive tests thus consisted of:

(1) *Kohs Blocks, Alexander Modification.*—Norms derived by Vernon, and at present in use in the Maudsley Hospital were used in deriving I.Q.'s.

(2) *Progressive Matrices*, 1938.—This was given as an untimed test to small groups. Norms for the untimed version were used in estimating I.Q.'s. About half the patients finished the test within 20 minutes, finding it "too easy" to spend more time over. Some took considerably longer, however, one subject (Matrices I.Q. 108) spending 55 minutes over the test.

(3) *Binet Vocabulary Test* from 1937 Terman Merrill Revision, Form L.—Norms taken from R. L. Thorndike (1942) were used in estimating I.Q.'s.

(4) *Porteus Maze, Vineland Revision* (Porteus, 1933).

(5) *Cattell Non-Verbal Intelligence test, Form IB.*—Form IA was not available at the time of testing. In calculating I.Q.'s the norms for Form IA were used, since those for Form IB, which is an equivalent test designed for re-testing purposes, make allowance for practice effects following testing on Form IA. In general, I.Q.'s for any raw score on Form IB are about 4 points lower than those for the same score on Form IA.

(6) *Tool Matching* (G.A.T.B., Part A).—Consists of a series of exercises containing a stimulus drawing and four black-and-white drawings of simple shop tools. The examinee indicates which of the four black-and-white drawings is the same as the stimulus drawing. Variations exist only in the distribution of black and white in each drawing.

(7) *H Markings* (G.A.T.B., Part C).—Consists of a series of large capital H's. The examinee draws a short vertical line through the bar of each H with-

out touching the sides, working rapidly to draw as many lines as possible during the time allowed.

(8) *Computation* (G.A.T.B., Part D).—Consists of a number of arithmetic exercises requiring the addition, subtraction, multiplication, or division of whole numbers.

(9) *Two-dimensional Space* (G.A.T.B., Part F).—Consists of a series of exercises containing a stimulus figure and five geometrical figures (two dimensional line drawing). The examinee indicates which one of five geometrical figures is made by a rearrangement of the parts of the stimulus figure.

(10) *Speed* (G.A.T.B., Part G).—Consists of a series of large rectangles. The examinee taps with pencil to make three dots in each of the rectangles, working as rapidly as possible during the time allowed.

(11) *Three-dimensional Space* (G.A.T.B., Part H).—Consists of a series of exercises containing a stimulus figure and four drawings of three-dimensional objects. The stimulus figure is pictured as a flat piece of metal which is to be either bent, or rolled, or both. Lines indicate where the stimulus figure is to be bent. The examinee indicates which of the four drawings corresponds to the stimulus figure.

(12) *Mark Making* (G.A.T.B., Part K).—Consists of a series of squares in which the examinee is to make three pencil marks, working as rapidly as possible. The marks to be made are short lines, two vertical, and the third a horizontal line beneath them.

(13) *Form Matching* (G.A.T.B., Part L).—Consists of two groups of variously shaped line drawings. The examinee indicates which figure in the second group is exactly the same size and shape as each figure in the first or stimulus group.

(14) *Place* (G.A.T.B., Part M).—The equipment used for this test, and for Part N, consists of a rectangular wooden board (Peg board) divided into two sections, each section containing 48 holes. The upper section contains 48 cylindrical wooden pegs. The examinee removes the wooden pegs from the holes in the upper part of the board and inserts them in the corresponding holes in the lower part of the board, moving two pegs simultaneously, one in each hand. This performance is repeated two more times, with the examinee working rapidly to move as many of the pegs as possible during the time allowed for each performance.

(15) *Turn* (G.A.T.B., Part N).—The equipment described under Part M is used for this test. In this case the lower section contains the 48 cylindrical pegs. The examinee removes a wooden peg from a hole, using one hand, turns the peg over with the same hand so that the opposite end is up, and returns the peg to the hole from which it was taken. The examinee works rapidly to turn and replace as many of the 48 cylindrical pegs as possible during the time allowed. This performance is repeated two more times.

(16) *Assemble* (G.A.T.B., Part O).—The equipment used for this test consists of a small rectangular board (Finger Dexterity Board) containing 50 holes, and a supply of small metal rivets and washers. The examinee takes a small metal rivet from a hole in the upper part of the board and at the same time removes a small metal washer from a vertical rod with the other hand;

examinee puts the washer on the rivet, and inserts the assembled piece into the corresponding hole in the lower part of the board, using only one hand. The examinee works rapidly to move and replace as many rivets and washers as possible during the time allowed.

(17) *Dis-assemble* (G.A.T.B., Part P).—The equipment used for this test is the same as that described for Part O. Examinee removes the small metal rivet of the assembly from a hole in the lower part of the board; slides the washer to the bottom of the board; puts the washer on the rod with one hand and the rivet into the corresponding hole in the upper part of the board with the other hand. The examinee works rapidly to move and replace as many rivets and washers as possible during the time allowed.

(18) *Rail-Walking test.*—This simple test, devised by Heath (1942), requires a subject to walk heel to toe along three rails lying an inch or two off the ground. Of the three rails used, the first is 9 ft. long by 4 in. wide, the second 9 ft. by 2 in., the third 6 ft. by 1 in. Three trials are given for each rail. Score is a weighted sum of the number of feet walked. Full details, together with norms, have been published by Heath in his papers on the test.

Intercorrelations and Factor Analysis.

Intercorrelations between pairs of scores on all 18 of the cognitive tests included in the battery were computed. A table of correlations, together with mean scores and standard deviations, is presented in Table II. Raw scores were used in all cases in calculating correlations, and the means and standard deviations presented in Table II are those of raw scores. (Porteus Maze I.Q.'s were regarded as raw scores. They were derived from Porteus's (1933) norms.)

A factor analysis of the correlation matrix was made, using the centroid method. Only one factor was extracted, since significant residuals after extraction of the first factor were too few to warrant continuing the analysis. Inspection of the residuals showed that there was evidence in favour of postulating a second factor common to the Manual and Finger Dexterity tests (Tests 14 to 17), and a third factor common to the Space tests (tests 9 and 11).

Though the standard deviations of the U.S.E.S. subtests are somewhat less than those obtained by the U.S.E.S. in their standardizing samples (private communication from the U.S.E.S.), the correlations between the G.A.T.B. tests (Tests 6 to 17) are higher than those reported by the U.S.E.S.

Bibliography.

(1) Alexander, W. P. (1946), *A Performance Scale for the Measurement of Practical Ability*, p. 60. London: Nelson & Sons.
(2) Burt, C. (1921), *Mental and Scholastic Tests*, p. 465. London: Staples Press. (2nd ed., 1947.)
(3) Cattell, R. B. (1948), *A Guide to Mental Testing*, p. 411. 2nd ed. London: U.L.P.
(4) Currann, D., and Guttmann, E. (1946), *Psychological Medicine*, p. 246. 2nd ed. Edinburgh: E. & S. Livingstone, Ltd.
(5) Doll, E. A. (1936), *The Oseretsky Tests of Motor Proficiency*, p. 48. Minnesota Educ. Test Bur.
(6) Duncan, J. (1942), *The Education of the Ordinary Child*, p. 240. London: T. Nelson & Sons.

(7) DVORAK, B. J. (1947), " The new U.S.E.S. General Aptitude Test Battery," *J. appl. Psychol.*, **31**, 372–376.
(8) EYSENCK, H. J. (1947), *Dimensions of Personality*, p. 308. London : Kegan Paul.
(9) GODDARD, H. H. (1914), *Feeble-mindedness, its Causes and Consequences*, p. 599. New York : Macmillan.
(10) HEATH, S. ROY, jun. (1942), " Rail-walking Performance as Related to Mental Age and Etiological Type among the Mentally Retarded," *Am. J. Psychol.*, **45**, 240–247.
(11) *Idem* (1943), " The Military Use of the Rail-walking Test as an Index of Locomotor Co-ordination," *Psychol. Bull.*, **40**, 282–284.
(12) *Idem* (1944), " Clinical Significance of Motor Defect with Military Implications." *Am. J. Psychol.*, **57**, 482–499.
(13) *Idem* (1946), " Mental Pattern found in Motor Defectives," *J. abnorm. soc. Psychol.*, **41**, 223–225.
(14) HENDERSON, D., and GILLESPIE, R. D. (1950), *A Text-book of Psychiatry*, p. 740. 7th ed. London : Oxford University Press.
(15) MCNEMAR, Q. (1942), *The Revision of the Stanford-Binet Scale*, p. 185. Boston : Houghton Mifflin Co.
(16) O'CONNOR, N., " The Prediction of Psychological Stability and Anxiety Aggressiveness from a Battery of Tests Administered to a Group of High-grade Male Mental Defectives." To appear in *J. Gen. Psychol.*
(17) *Idem*, " Neuroticism and Emotional Instability in High-grade Mental Defectives." (To be published.)
(18) PENROSE, L. S. (1949), *The Biology of Mental Defect*, p. 285. London : Sidgwick & Jackson.
(19) PORTEUS, S. D. (1933), *The Maze Test and Mental Differences*, p. 219. Vineland, N.J. : Smith Pub. Co.
(20) *Idem* (1941), *The Practice of Clinical Psychology*, p. 579. New York : American Book Co.
(21) TERMAN, L. M. (1916), *The Measurement of Intelligence*, p. 362. Boston : Houghton Mifflin Co.
(22) *Idem* and MERRILL, M. A. (1937), *Measuring Intelligence*, p. 460. London : Harrap & Co.
(23) THORNDIKE, R. L. (1942), " Two Screening Tests of Verbal Intelligence," *J. appl. Psychol.*, **26**, 128–135.
(24) *Idem* (1949), *Personnel Selection ; Test and Measurement Techniques*, p. 358. New York : Wiley.
(25) TIZARD, J., and O'CONNOR, N. (1950), " The Employability of High-grade Mental Defectives (Parts I and II)," *Am. J. ment. Def.*, **54**.
(26) TREDGOLD, A. F. (1947), *A Text-book of Mental Deficiency*, p. 534. 7th ed. Baltimore : Williams & Wilkins.
(27) U.S. EMPLOYMENT SERVICE (1946), " One Year of Employment Counselling in the U.S. Employment Service," *Empl. Serv. Rev.*, (4) 3–20, 24.
(28) *Idem* (1947), *Guide to the Use of General Aptitude Test Battery*. Parts I and II. Washington : U.S. Dept. of Labor.
(29) VAN DER LUGT, M. J. A. (1948), " V.D.L. Adult Psycho-motor Series for the Measurement of Manual Ability," *J. Consult. Psychol.*, **12**, 440.
(30) WECHSLER, D. (1944), *The Measurement of Adult Intelligence*, p. 258. 3rd ed. Baltimore : Williams & Wilkins.
(31) YARMOLENKO, A. (1933), " The Motor Sphere of School Age Children," *J. Genet. Psychol.*, **42**, 298–316.