

THE PSYCHOLOGICAL BULLETIN

GENERAL REVIEWS AND SUMMARIES

MENTAL TESTS¹

BY FRANK N. FREEMAN

University of Chicago

THE ARMY TESTS

The widespread experimentation with the army tests, particularly Scale Alpha, and their influence on the development of other tests, makes it appropriate to give them the place of emphasis in this year's review. These tests, the publicity which has been given them and the large number of men who gained experience in giving tests by using them as psychological examiners in the army, are doing very much to popularize intelligence tests, and to make familiar the outstanding facts concerning human abilities which tests reveal. Much of the experimentation, to be sure, is the merely routine repetition of what others have done, but occasionally there is an investigator who attacks the problems from a slightly new angle.

Of considerable interest and value to those who are interested in the army tests is the publication of the virtually official manual under the authorship of Yoakum and Yerkes (48). The book contains a reproduction and description of the different scales which were in use in the psychological service, including the group tests, Alpha and Beta, and the individual tests, both language and performance. In addition are given directions for giving and scoring, scoring keys, the scheme for rating and for interpreting

¹On account of the closing of one library and the moving of another at the time this review was being prepared some of the articles were not available. Those of especial importance will be included in the next article.

the rating on each scale in terms of the others. Besides this account of the tests themselves is a sketchy statement of the methods of designing the tests and a brief account of typical results. This account is evidently a popularized anticipation of the fuller and more technical report which is to follow. For the technical reader the present presentation will raise more questions than it will answer.

During the period covered by this survey there is record of the application of army test Alpha, the general group scale, to students in nine higher institutions and three high schools. The institutions and the authors of the studies follow: Dickinson College, M. G. Filler (6); Hamline University, G. D. Walcott (43); Southern Methodist University, H. T. Hunter (10); Purdue University, G. L. Roberts and G. C. Brandenburg (36); Oberlin College, E. S. Jones (11); University of Minnesota, M. J. Van Wagenen (42); Ohio State University, E. L. Noble and G. F. Arps (25); University of Oklahoma, A. M. Jordan (12); the high schools of Madison, Rock Island and Sioux City, I. N. Madsen and R. H. Sylvester (19, 20 and 21). From these studies data of interest can be gathered with reference to such matters as institution medians, class medians, differences between school groups and sex groups within institutions and correlation with marks.

While the comparison between institutions is somewhat vitiated by some uncertainty concerning the class and sex tested and the method of timing, it is apparent that there are rather wide variations. The highest freshman median recorded is that for the students at Yale, 159.5. The score for Oberlin freshmen is 153, and for the Ohio freshmen, 130. Other scores probably include women, for which an allowance must be made, as will appear shortly. The median at Dickinson is 141, at The Southern Methodist University, 127, and at The University of Oklahoma, 119. Making due allowance for the above-mentioned factors it appears that there is a difference in the general ability of typical students in various educational institutions, due either to selection or training. The explanation involves the general problem of the interpretation of test scores, but whether the differences be ascribed to training or capacity they present administrative problems, such as that of accrediting students and accepting degrees.

Almost uniform sex differences are found in favor of the men in the army test scores. That these differences are specialized rather than general appears when the scores in the individual tests

are compared separately. It appears that the superiority of the men, which amounts in total to from four to eleven points, resides almost entirely in their higher scores in tests 2 and 8, the arithmetic and information tests. The information test is clearly better adapted to men than to women. That the same difference appears between the sexes in the grammar school was found by Uhrbrock (41).

College and university students, and as well the high school students who were tested, are a very highly selected group. The poorest pupils in high school get a rating of C, and in some colleges very few, in others none, are rated as low as this. Undoubtedly a part, but certainly not all, of this superiority is due to training.

The practical usefulness of the tests is indicated to some extent by degree to which the scores correlate with school and college marks. The correlation in college varies from .305 to .52, and in high school from .20 to .38. Of the nine first honor men at Yale, out of 373 freshmen, two were below the median in the test. If the tests are to be used for prognosis a very wide margin of error must be allowed for. There is evidence that other tests are better adapted to the differentiation of high-grade ability.

In view of the usually moderate attainment of members of fraternities as measured by marks, it is interesting that at Dickinson fraternity men made 11.8 points more in the tests than others, and fraternity women 8.7 more. Different schools at Purdue varied from 124.1 (pharmacy) to 137.7 points (chemical engineering) median score, and at Ohio from 112 (veterinary) to 157 (graduate). A comparison of private school and high school alumni at Yale brought out no significant difference. The medians of successive classes indicate a steady and marked progression up to the freshman year of the college, and a less marked progression throughout the college. How much of this is due to elimination of the poor students and how much to maturity the evidence does not show.

Doll (4 and 5) argues from the reported results of the army tests that mental growth in the "average" individual ceases at 13 years, and bases his conclusion on the fact that the average "mental age" of recruits was found to be 13.5 years. He further proposes that 13 be used instead of 16 as the age to correspond to adult mentality in reckoning the IQ of adults. But the evidence is so strong that mental growth continues even beyond the age of 16 that a contrary result should lead to a search for imperfection in the test. Refuge cannot be taken in a fine-drawn distinction be-

tween pure growth and the result of experience, because nobody has succeeded in clearly separating the two. One clue is to be sought in the limitation in the difficulty of the test, which fails to give opportunity to the higher grades of ability, or to the type of ability represented in the older individuals.

TECHNICAL PROBLEMS IN ORGANIZATION AND EVALUATION

One of the problems which is prominent in the design of the modern point scale is the determination of the method of scoring the individual tests and of combining these scores; and a phase of this problem is the weighting of the individual tests. Two elaborate schemes of weighting are presented in the articles by Arthur and Woodrow, and by Herring (2 and 9). The tests used in the first-mentioned study are mostly familiar,—memory span, opposites, substitution, word-building, language completion, cancellation and comprehension (Kuhlmann's). They were given to children from six to thirteen years of age. The point value of each tests was calculated on the basis of its discriminative value, that is, the extent to which the scores from it were differentiated in successive years. The formula used was

$$D.V. = \frac{Av.1 - Av.2}{\frac{1}{2}(\sigma_1 + \sigma_2)},$$

in which *Av. 1* and *Av. 2* represent the average scores in two successive ages. The values obtained are used as point scale values, and the value of the score of a given age is found by adding the values up to that point. The value for each age of the entire scale is found by adding those for the individual tests. This, the authors hold, constitutes an absolute scale of mental growth; but it should be noted that the form of the growth curve still depends on the adaptation of the tests to the ability of the pupils at successive ages. Herring used successive grades instead of ages and found the differences in the median performance of successive grades in terms of the P.E. of the distribution in each of his 33 tests. Unlike Arthur and Woodrow, Herring relates his scale values to zero.

The technique of the interpretation of the scores in tests is discussed in three articles. Kelley (13) points out the important fact that the amount of overlapping in the scores of successive ages or grades represents the true overlapping in ability only when the test is perfectly reliable, and that the unreliability of tests has

resulted in gross overestimation of the true amount of overlapping. Thorndike (38) names and calls attention to a factor, which those who have dealt with estimates of ability must have noticed, through which the correlation between the estimates of various traits by the same judges is unduly raised. This factor, which Thorndike names appropriately the "error of the halo," is the effect on the judges estimate of an individual's capacity in a particular trait of his general opinion of the individual's ability. This factor undoubtedly has a large influence on school marks. Thorndike suggests its avoidance by having one person produce the data and another judge each trait independently on the basis of the data. Myers (23) points out what he calls a fallacy in correlation of test scores, which rests on the fact that if all unselected individuals—presumably by ages—are put into a group for the calculation of correlation, the coefficient will be higher than if the more homogeneous group within a given school grade are used—presumably again by ages. This by no means makes the first practice a fallacy. In fact, it is more useful for the administrator to know what the facts are, independent of the artificial grade grouping.

A series of brief studies on a number of tests, particularly to discover means of making them non-coachable, is reported by F. L. Wells and C. M. Kelley (47). The differential reaction of persons over fifty years of age to the tests of the Yerkes Point Scale is reported by Foster and Taylor (7). They find marked peculiarities in their scores in naming words, combining three words in a sentence, drawing from memory and rearranging dissected sentences, and give norms with these omitted.

MISCELLANEOUS CORRELATIONS

The possibility of finding specific tests in ability in aviation and of predicting success or failure in extreme cases is shown by Henmon (8). The composite series or team of tests selected as a result of the correlation study of each one gives a correlation with judgments of flying ability of .70. Parsons and Segar (26) found no correlation between the Bárány chair test and flying ability.

Practically identical results with the Pearson and the army rating scales used by instructors in rating college students is reported by Kitson (14). This suggests the need of investigating the error of the halo in the use of rating scales.

Several studies, in addition to those with the army tests, have

been made between tests and other measures of ability. Caldwell (3) reports a correlation of .44 between the adult Stanford tests and college marks in the case of students at the Randolph-Macon Woman's College, and .47 with estimated intelligence. Sunne (37) gave the Yerkes and Stanford adult tests to high school and college students and obtained correlations ranging from .43 to .74 in the point scale, and from .30 to .65 in the Stanford tests. The correlation between reading ability as measured by the Monroe Silent Reading test and general ability as measured by Alpha, is reported by Webb (46), also the correlation of Alpha and Thurstone's tests A and B tests with marks. The coefficients are, Alpha with comprehension .68, Alpha with marks .57, Thurstone A with marks .41, Thurstone B with marks .49. Ratings of college students by Scott's Rating Scale and college marks were found by Kohs and Irle (15) to have such low predictive value for promotion in the army as to be of no practical value.

DIFFERENCES BETWEEN RACIAL AND ENVIRONMENTAL GROUPS

Important differences between racial, occupational and regional groups are reported to have been found in the results of the army tests, but the complete returns have not yet been published. In the meantime less extensive comparisons are being made from time to time. Pressey and Teter report a comparison of 187 colored children with white children in Indiana by the Pressey group scale (34). In conformity with other studies they find the colored children about two years behind. Their inferiority is less in the simpler rote memory tests than in those which require thinking, and they are poorer in the later tests of the scale, which suggests fatigue. A similar difference was found by Partlow and Haines (27) in Alabama. A comparison of Chinese with American students made by Walcott (44) by means of the Stanford revision gave rather inconclusive results.

A comparison of children in a good and a poor country district in Indiana is made by S. L. Pressey and Thomas (35), and a comparison of country and city children and of four occupational groups in the city by L. W. Pressey (29). Another comparison of occupational groups is reported by S. L. Pressey and Ralston (32). In both studies in which country children were compared with city children the former were found to be markedly behind. Only from 20 per cent to 36 per cent exceeded the median of the city group. A considerable portion of this inferiority in the tests is

ascribed to incidental causes, but that some of it represents inherent difference in ability is strongly suggested by the fact that 16 per cent more of the children in the good country district exceeded the median of the city children than of the poor district. Within the city the occupations were grouped under the heads, professional, executive, artisan and laborer. The children stood in this order in the average scores of the groups, and in each case the average of the professional group was a little over double that of the laboring group. There was, of course, considerable overlapping.

Delinquents and dependents in four industrial schools in Alabama were found by Partlow and Haines (27) to be very inferior in a test patterned after the army test. Of the white boys and girls nearly half are below the five percentile score of normal children. The colored boy delinquents, however, differed much less from a random sampling of colored children. The group tests place in the lowest two percentile nearly all of those who are diagnosed by individual tests to be feeble-minded.

NEW TESTS OR NEW FORMS OF OLD TESTS

New group tests are appearing very rapidly. In addition to those which are described in the literature and which are referred to in the list of references below a few others which have come to the attention of the writer may be named. Space does not permit further description.

Author or Name	Grades	Type	Publisher or Distributor
Chicago Intelligence Test	7 to 12	Verbal	Univ. of Chicago bookstore
Dearborn, W. F.		Non-verbal	Lippincott
Haggerty, M. E.			
Delta 1.	1 to 4	Non-verbal	World Book Co.
Delta 2.	4 to 9	Verbal	World Book Co.
Kingsbury, F. A.	1 to 4	Non-verbal	Univ. of Ill. Bureau of Educ. Research
Illinois Examination.	{ 3 to 5	Verbal }	Univ. of Ill. Bureau of Educ. Research
{ 6 to 8	Verbal }		
National Intelligence Test			
Scales A and B	3 to 9	Chiefly verbal	World Book Co.
Omaha Group Test.	5 to 8 (?)	Verbal	P. R. Stevenson, Univ. of Illinois
Terman Group ..	7 to 12	Verbal	World Book Co.
Trabue, M. E. and Stockbridge, F. P.	?	Mixed	Doubleday, Page & Co.
Whipple, G. M.	5 to 8	Chiefly verbal	Public School Pub. Co.

Besides those listed above a number of group tests have recently appeared, and are reported in articles. A group scale for primary

grades consisting chiefly in the adaptation of certain of the tests of the Binet Scale is described by Miss Lowell (17) and 18). Five tests each are places at ages five to nine. Scoring is on the age level principle. A group, non-verbal point scale of four tests for primary grades is described by C. and G. Myers (24). Pintner (28) is the author of a non-verbal group scale of six tests which gives a reasonably normal curve of distribution for children of the upper grades. A brief scale of crossout tests, that is, tests which can be passed by crossing out one element, has been devised by the Presseys (31). One of these is a test for moral judgments, and further suggestions are made for moral tests. These suggestions are elaborated in an interesting manner in the article by Pressey and Chambers (30). This is perhaps the most novel contribution to the field of tests during the year. The five tests are designed to measure, (1) emotional spread and displacement, (2) emotional distractibility, (3) moral discrimination and experience, (4) free association (with, however, limited choice), and (5) emotional memory.

Three adaptations of the Binet Scale are described by Pressey and Shively (33), who substitute terms involving practical information for these in the vocabulary test, by Lincoln and Cowdery (16), and by Washburne (45), who seeks to make an analytical diagnosis by classifying the tests according to mental function.

Three groups of tests having the purpose of vocational diagnosis are reported by Thurstone (39) and 40) and Murray (22). Thurstone's are general tests adapted in content to the experience and interests of the groups to whom they are given. Miss Murray attempted to distinguish four general types of ability among college students,—intelligence, accuracy, practical ability and social ability, but was not able to find convincing objective verification of her hypothesis.

REFERENCES

1. ANDERSON, J. E. Intelligence Tests of Yale Freshmen. *School & Soc.*, 1920, 11, 417-420.
2. ARTHUR, G. & WOODROW, H. An Absolute Intelligence Scale: a Study in Method. *J. of Appl. Psychol.*, 1919, 3, 118-137.
3. CALDWELL, H. H. Adult Tests of the Stanford Revision applied to College Students. *J. of Educ. Psychol.*, 1919, 10, 477-488.
4. DOLL, E. A. The Average Mental Age of Adults. *J. of Appl. Psychol.*, 1919, 3, 317-328.
5. DOLL, E. A. The Growth of Intelligence. *J. of Educ. Psychol.*, 1919, 10, 524-525.
6. FILLER, M. G. A Psychological Test. *School & Soc.*, 1919, 10, 208-209.

7. FOSTER, J. C., & TAYLOR, G. A. The Applicability of Mental Tests to Persons over Fifty Years of Age. *J. of Appl. Psychol.*, 1920, 4, 39-58.
8. HENMON, V. A. C. Air Service Tests of Aptitude for Flying. *J. of Appl. Psychol.*, 1919, 3, 103-109.
9. HERRING, J. P. The Derivation of a Scale to Measure Ability in Scientific Thinking. *J. of Educ. Psychol.*, 1919, 10, 417-432.
10. HUNTER, H. T. Intelligence Tests at Southern Methodist University. *School & Soc.*, 1919, 10, 437-440.
11. JONES, E. S. The Army Tests and Oberlin College Freshmen. *School & Soc.*, 1920, 11, 389-390.
12. JORDAN, A. M. Some Results and Correlations of Army Alpha Tests. *School Soc.*, 1920, 11, 354-358.
13. KELLEY, T. L. The Measure of Overlapping. *J. of Educ. Psychol.*, 1919, 10, 458-461.
14. KITSON, H. D. Comparison between Two Scales for the Estimation of Intelligence. *J. of Appl. Psychol.*, 1919, 3, 310-316.
15. KOHS, S. C. & IRLE, K. W. Prophesying Army Promotion. *J. of Appl. Psychol.*, 1920, 4, 73-87.
16. LINCLON, E. A. & COWDERY, K. M. An Abbreviated Mental Age Scale for Adults. *School & Soc.*, 1919, 10, 707-709.
17. LOWELL, F. A Preliminary Report of Some Tests of General Intelligence. *J. of Educ. Psychol.*, 1919, 10, 323-344.
18. LOWELL, F. A Group Intelligence Scale for Primary Grades. *J. of Appl. Psychol.*, 1919, 3, 215-247.
19. MADSEN, I. N. & SYLVESTER, R. H. High-school Students' Intelligence Ratings according to the Army Alpha Test. *School & Soc.*, 1919, 10, 407-410.
20. MADSEN, I. N. High-school Students' Intelligence Ratings according to the Army Alpha Test. *School & Soc.*, 1920, 11, 298-300.
21. MADSEN, I. N. The Army Intelligence Test as a Means of Prognosis in High School. *School & Soc.*, 1920, 11, 625-628.
22. MURRAY, E. Psychological Tests as Diagnosis of Vocational Aptitudes in College Women. *J. of Appl. Psychol.*, 1920, 4, 30-38.
23. MYERS, G. C. A Grave Fallacy in Intelligence Test Correlations. *School & Soc.*, 1920, 11, 528-529.
24. MYERS, C. & MYERS, G. A Group Intelligence Test. *School & Soc.*, 1919, 10, 355-360.
25. NOBLE, E. L. & ARPS, G. F. University Students' Intelligence Ratings according to Army Alpha Test. *School & Soc.*, 1920, 11, 233-237.
26. PARSONS, R. P. & SEGAR, L. H. Bárány Chair Tests and Flying Ability. *J. Amer. Med. Ass.*, 1919, 70, 1064-1065.
27. PARTLOW, W. D., & HAINES, T. H. Mental Rating of Juvenal Dependents and Delinquents in Alabama. *J. of App. Psychol.*, 1919, 3, 291-309.
28. PINTER, R. A Non-language Group Intelligence Test. *J. of Appl. Psychol.*, 1919, 3, 199-214.
29. PRESSEY, L. W. The Influence of (a) Inadequate Schooling and (b) Poor Environment upon Results with Tests of Intelligence. *J. of Appl. Psychol.*, 1920, 4, 91-96.
30. PRESSEY, S. L. & CHAMBERS, C. R. A Group Scale for the Emotions. *J. of Appl. Psychol.*, 1920, 4, 97-104.

31. PRESSEY, S. L. & PRESSEY, L. W. Cross-out Tests, with Suggestions as to a Group Scale of the Emotions. *J. of Appl. Psychol.*, 1919, 3, 138-150.
32. PRESSEY, S. L. & RALSTON, R. The Relation of General Intelligence of School Children to the Occupation of their Fathers. *J. of Appl. Psychol.*, 1919, 3, 366-373.
33. PRESSEY, S. & SHIVELY, I. M. A Practical Information Test for Use with Delinquents and Illiterate Adults. *J. of Appl. Psychol.*, 1919, 3, 374-380.
34. PRESSEY, S. I. & TETER, G. F. A Comparison of Colored and White Children by Means of a Group Scale of Intelligence. *J. of Appl. Psychol.*, 1919, 3, 277-282.
35. PRESSEY, S. L. & THOMAS, J. B. A Study of Country Children in (1) a Good and (2) a Poor Farming District by Means of a Group Scale of Intelligence. *J. of Appl. Psychol.*, 1919, 3, 283-296.
36. ROBERTS, G. L. & BRANDENBURG, G. C. The Army Intelligence Tests at Purdue University. *School & Soc.*, 1919, 10, 776-778.
37. SUNNE, D. The Relation of Scholarships to the Yerkes and Terman Adult Tests. *J. of Educ. Psychol.*, 1919, 10, 520-524.
38. THORNDIKE, E. L. A Constant Error in Psychological Ratings. *J. of Appl. Psychol.*, 1920, 4, 25-29.
39. THURSTONE, L. L. A Standardized Test for Office Clerks. *J. of Appl. Psychol.*, 1919, 3, 248-251.
40. THURSTONE, L. L. Mental Tests for Prospective Telegraphers, etc. *J. of Appl. Psychol.*, 1919, 3, 110-117.
41. UHRBOCK, R. S. Sex Differences as Noted in Grammar-school children. *School & Soc.*, 1920, 11, 507-509.
42. VAN WAGENEN, M. J. Some Results and Inferences derived from the Use of the Army Tests at the University of Minnesota. *J. of Appl. Psychol.*, 1920, 4, 59-72.
43. WALCOTT, G. D. Mental Testing at Hamline University. *School & Soc.*, 1919, 10, 57-60.
44. WALCOTT, G. D. The Intelligence of Chinese Students. *School & Soc.*, 1920, 11, 474-480.
45. WASHBURNE, C. W. A Classified Scale for Measuring Intelligence. *J. of Educ. Psychol.*, 1919, 10, 309-322.
46. WEBB, L. W. Ability in Mental Tests in Relation to Reading Ability. *School & Soc.*, 1920, 11, 567-570.
47. WELLS, F. L. & KELLEY, C. M. Briefer Studies from the Psychological Laboratory of McLean Hospital. *J. of Appl. Psychol.*, 1919, 3, 172-193.
48. YOAKUM, C. S. & YERKES, R. M. *Army mental tests*. New York: Henry Holt & Co., 1920. Pp. 303.