

DEPARTMENT OF THE INTERIOR  
BUREAU OF EDUCATION

BULLETIN, 1923, No. 57

# EDUCATIONAL TESTS

BY  
STEPHEN S. COLVIN

[Advance Sheets from the Biennial Survey of Education  
in the United States, 1920-1922]



WASHINGTON  
GOVERNMENT PRINTING OFFICE  
1924

**ADDITIONAL COPIES  
OF THIS PUBLICATION MAY BE PROCURED FROM  
THE SUPERINTENDENT OF DOCUMENTS  
GOVERNMENT PRINTING OFFICE  
WASHINGTON, D. C.  
AT  
5 CENTS PER COPY**

## EDUCATIONAL TESTS.<sup>1</sup>

CONTENTS.—The development of mental testing—The Binet tests—Army Alpha tests—Army Beta tests—Recent group tests.

A decade ago intelligence testing was in its beginnings in the United States. There were no standardized tests available except those of the Binet-Simon scale. These tests had been used but little, and chiefly for the detection and classification of the backward and the feeble-minded. Goddard had just begun pioneer work in this field, while Kuhlmann and Huey had added to the scant literature that was just beginning to appear in this country. Altogether the important articles numbered scarce a score. The point scale of Yerkes had not yet been published, Terman's important revision was still in the process of making, while group tests to measure the mentality of numbers of persons simultaneously had not been thought of.

To-day all is changed. The programs of the meetings of psychologists devote more space to the discussion of intelligence tests than to any other single topic. When schoolmen and teachers meet, this question claims their attention and interest, and in the schools themselves throughout the country extensive "testing programs" are conducted and the results used for improvement in teaching and administration.

The development and standardization of intelligence tests have resulted in four main types—two of which are for administration to individuals and two to groups. These are:

1. The Binet tests and their revisions and additions.
2. The performance tests for individual administration, including the various form-board tests, puzzle tests, picture tests, etc.
3. The group-intelligence tests of the Army Alpha type—paper and pencil tests for the most part of a linguistic nature.
4. The group-intelligence tests of the Army Beta type, performance tests reduced to paper and pencil form, for use particularly, though not exclusively, in the examination of little children, illiterates, and non-English-speaking groups.

### I.—THE DEVELOPMENT OF MENTAL TESTING.

Definite and systematic attempts to measure mentality began 18 years ago, when Binet and Simon in 1905 published in *L'Année Psychologique* of that year, their collection of tests designed to de-

<sup>1</sup>At the time of his death Dr. Stephen S. Colvin was preparing this manuscript for publication. Apparently not fully completed, it is published in substantially the form in which he left it.

termine the native intelligence of school children. These tests were framed for the purpose of segregating defective children in the schools of Paris from those of normal mentality, with the aim of providing these unfortunate pupils with the instruction best suited to their limited intellectual capacities. Thus, like many other significant and far-reaching movements in psychology and education, mental testing began in an attempt to help the subnormal and defective, and has since spread until it finds its largest and most useful field in the realm of normal psychology.

These first mental tests were not merely the outcome of a happy guess or of a flash of genius. Binet's many years of expert psychological observation and experimentation achieved their most significant results in the construction of these tests. The first series of tests were preliminary and tentative in their nature. From applying them to children of various ages Binet found, for example, what was the average attainment in the various tests of normal children at various ages. This resulted in the construction of the intelligence scale of 1908. In this scale, groups of tests were arranged for children beginning with the third year and continuing through the thirteenth year. The tests were arranged "according to the ages at which the majority of children succeed in them." The number of tests for the various ages varied from eight for 7 years to three for 13 years, the most frequent number being five. The scale was revised by Binet in 1911, the year of its author's death. In the 1911 scale there were tests for all ages from 3 to 12, inclusive, a test for the fifteenth year, and a final test for adults. Under each age there were five subtests, with the single exception of the fourth year, for which only four subtests were given.

Although the most extensive and significant work in mental testing has been done in America, the Binet tests were slow in making their appearance in this country. Goddard, then psychologist at the school for feeble-minded in Vineland, N. J., first learned of these tests in 1908, and in December of that year published a brief account of the tests of 1905. In January, 1910, Goddard published an abstract of the scale of 1908. Goddard was at first extremely critical of the value of this scale. He says, "It seemed impossible to grade intelligence in that way. It was too easy, too simple." However, when the abstract of the scale was used he found that it "was a surprise and gratification." The classification of 400 feeble-minded children at Vineland by the Binet method during this year "agrees with the institution's experience," and Goddard became an enthusiastic proponent of Binet's scale.<sup>2</sup> He followed his survey of the children at Vineland by applying the tests to 2,000 normal children.<sup>3</sup>

<sup>1</sup> Fed. Sem., Vol. XVII, No. 3, Sept., 1910.

<sup>2</sup> Op. cit., Vol. XVIII, No. 2, June, 1911.

Since the introduction of the Binet scale to America several important revisions and adaptations have been made. In 1915 appeared the point scale by Yerkes, Bridges, and Hardwick, and a year later the Stanford revision by Terman. Yerkes's revision is particularly notable because of its method of scoring and the order of the presentation of the tests. Among other revisions may be mentioned that by Kuhlmann and the recent emendations by Herring.

As has been pointed out already, the Binet tests were found insufficient and inadequate in certain particulars. Perhaps the chief criticism to be brought against them, however, is that they are individual tests, and that they can not be satisfactorily given unless the person who administers the test has adequate training and considerable practical experience. The fact that these tests are individual means that they must be given to each child separately. In a room of 40 children this would require a total time for testing of about 20 hours. It would further require a skilled person to give the children the tests, and therefore any attempt to measure the intelligence of a group of school children, such as would be found in a town or city even of moderate size, would necessitate an amount of time and an expenditure of money that would seem in many cases prohibitive. When a few children are to be tested for specific reasons, the Binet tests may be advantageously employed, but when the intelligence of all the children in a school system is to be surveyed, then another instrument for measuring intelligence must be found if possible. Such an instrument now exists in the various group tests to-day available for testing the intelligence of all children in our schools from the kindergarten through the high school, and indeed for mature students, in colleges and professional schools.

While the development of intelligence tests, based on the pioneer work of Binet, was going on, psychologists were employing various tests to discover how individuals differed in certain physical and mental capacities. The study of individual differences, begun by Galton in England, was made known in this country in the last decade of the nineteenth century, largely through the work of James McKeen Cattell, then professor of psychology in the University of Pennsylvania and later of Columbia University. Cattell gave psychological tests of the sensory and motor type to students at the University of Pennsylvania and later at Columbia University. The article by Cattell and Farrand appearing in the *Psychological Review* more than a quarter of a century ago gives an interesting statement of this work and the results achieved. An examination of Cattell's tests shows that they concern themselves largely with sensory discrimination and rapidity of reaction. Likewise immediate memory (memory span) is tested by finding the number of letters a

subject remembers at one hearing. Ability to estimate space is determined by a test requiring the bisection of a line of 50 centimeters; ability to estimate time is tested by estimating a 10-second interval. A judgment of least noticeable differences in weight is also included. In a later article by Cattell and Farrand we find a description of the further extension of the work of mental testing as employed with students of Columbia University as subjects. The tests used included handwriting, visual acuity and color vision, auditory acuity and perception of pitch, sensitivity of the skin, perception of weight, sensitivity to pain, accuracy and steadiness of movement, reaction time, cancellation of A's, perception of time and space, memory span, memory of length of a line previously drawn, after-images, and mental imagery. In regard to the tests Cattell says: "Our experience with these tests leads us to recommend that they be made a part of the work of every psychological laboratory."

These tests were given to individuals of normal mentality. At the same time other psychological tests of a somewhat different type were being developed through efforts to train the feeble-minded. Here the work of Sequin can not be overlooked.

In his work with children of low-grade intelligence Sequin found the form board of value. The form-board test has passed through various adaptations, but its essential character has been kept. It consists in fitting wooden blocks of various shapes into forms cut out to receive them. The board may be very simple, or it may be made as complex as desired, not only as to the shape and number of forms used, but also in regard to the blocks to be fitted, since each block may be a single solid piece or composed of a number of pieces, in which case the pieces must themselves be fitted together as well as placed in the proper form. A variation of this test consists of a puzzle in which various parts of a figure or shape are required to be fitted together, as, for example, in the Healy manikin puzzle. Picture-puzzle tests have been largely used in recent years as performance tests. In this type of test the various parts of a picture are to be arranged in their proper order. In some instances a picture with parts omitted is given the subject, and he is required to complete the picture by filling in the gaps with the proper blocks. Another type of picture test consists in arranging a series of pictures in such an order that they tell a complete story. A form of the performance test that is now frequently used is the "maze test." This test was used extensively 20 years ago, in the earlier days of animal psychology when the intelligence of an animal, such as a white rat, was studied by finding how easily and surely the animal could learn to go through the passages of a maze and get to the center where the food was placed. The Porteus maze test for detecting

feeble-mindedness is the best adaptation of this test. The maze test when used with human beings is a paper and pencil test of the performance type. The maze is printed on a sheet of paper, and the person tested is required to trace with a pencil the correct way of going through the maze. The form-board test and the various picture-puzzle tests have also been adapted to paper and pencil use, but nevertheless retain essential characteristics as performance tests.

The performance tests, like the Binet tests and the tests employed by Cattell, were originally of the individual type and required a large amount of time in the aggregate for their administration as well as expert skill on the part of the person giving the test. The great advance in the technique of administering intelligence tests came when these tests were so constructed as to enable them to be given to groups of individuals rather than to persons one at a time. Group testing, although practiced to some extent before 1917, owes its chief impetus to the formulation of the so-called Army tests that were employed on an extensive scale after America had entered the World War.

At this time the various attempts at intelligence testing were brought to a head in the construction of the Army Alpha mental examinations for the literates and the Army Beta examinations for the illiterates. The Army tests were the first instances of paper and pencil tests that were applied to groups of individuals on an extensive scale. The Alpha tests trace their origin more or less directly to the various psychological tests employed to determine individual differences, and to the Binet tests and their various revisions. The Beta tests were more closely allied to the performance tests previously mentioned, but were adapted to the paper and pencil form, and were designed, like the Alpha tests, for measuring the mentality of those tested in groups rather than through individual examinations.

The results of the Army tests were so satisfactory that on the conclusion of the armistice they were made public, and the Army Alpha was given widely in schools, colleges, and universities during the year 1919. Prior to this time, in the fall of 1918, the so-called Brown University test was given to the men in the Students Army Training Corps and the naval unit at Brown. It has since been continued with all entering classes at that institution and is now in its fifth year of trial. This fact is mentioned because it would seem to be the first group intelligence test to be consistently and continuously used and standardized. It precedes somewhat, in point of time, the various group tests that have been devised and published during the last three years, in all numbering nearly 40 separate tests.

## II.—THE BINET TESTS.

The Binet tests as they exist in their most careful and comprehensive revision and extension are known as the Stanford-Binet, and in their present form are the work of Lewis M. Terman. The Binet scale as perfected by its author is composed of 54 tests. The Stanford revision consists of 74 main tests and 16 alternates—90 in all. There are 36 tests not found in the original Binet scale. The present character of intelligence tests owes so much to the Binet test and its revisions that it will be of assistance to describe in some detail the nature of the Stanford-Binet.

This scale like its prototype has a series of tests arranged for various years. Terman speaks of them as "stunts, or problems, success in which demands the exercise of intelligence." This description while fairly good is not entirely accurate, since many of the tests included clearly involve no problem in the generally accepted sense of the term, being based on recognition of familiar objects, on simple associations, and on acts of skill. Still others require concentrated attention and memory for immediate impressions. In all roughly less than half involve comparison, judgment, and reasoning.

In the Stanford-Binet, there are six main tests for each year from 3 through 10. There are eight tests for the twelfth year and six for the fourteenth year. An average adult and a superior-adult level are included; each with six main tests.

The Stanford-Binet, like its original, begins with simple tests to determine the child's knowledge of common objects in his environment. The child is asked to point to his nose, eyes, mouth, hair; to name familiar objects, to enumerate objects in a picture, to give his sex, and to tell his last name. In later years he is asked to name colors, to distinguish between his right and left hand, to discover what is missing in a mutilated picture, to recognize familiar coins, to tell whether it is morning or afternoon, to name the days of the week, to tell the month, the day of the month, and the year, to give the meaning of words in a series of vocabularies progressing in difficulty for the various ages tested, and the like. Tests of this character clearly do not involve a mental ability beyond that of the perceptive and associative processes of the human mind, and they are definitely dependent on the experiences of the child in school and out. No mental ingenuity of a high degree of complexity, no rational processes of a developed sort are here demanded. These tests seek primarily to discover the child's range of information.

In the tests of the third year is found a simple *memory-span test*. The child is required to repeat such a sentence as, "In summer the sun is hot," or an alternative test to repeat three digits as, "3, 5, 2." The memory-span test appears in various forms at other levels.

The child of 4 is required to repeat 4 digits, or as an alternative test 12 to 13 syllables. At 6, he is required to repeat 16 to 18 syllables; at 7, 5 digits, and as an alternative test to repeat 3 digits backwards. At 8, an alternative test requires the child to write from dictation the sentence, "See the little boy." At 9, one of the tests is to repeat 4 digits backwards. At 10, there are two alternative tests of the memory-span type, one to repeat 6 digits and a second to repeat from 20 to 22 syllables. A test for the twelfth year requires the repetition of 5 digits backwards, while an alternative for the fourteenth year requires the repetition of 7 digits in the order of original presentation.

The average-adult test has the requirement of repeating 6 digits backwards, or as an alternative 28 syllables; while the superior-adult test requires the repetition of 8 digits in the order of presentation and 7 digits backwards. Tests of this type have been used by psychologists for many years, the general conclusion being that there is a fairly close relation between immediate-memory ability and school attainment. A memory test of a somewhat different type is included under the tenth year. A brief report of a fire is read and the child is required to give back the essential ideas. A more complicated test of this type is placed as the superior-adult level.

In the tests for the fourth year there appears the attempt to measure the ability of a *simple act of skill*, the counting of 4 pennies; at the six-year level the requirement is to count 13 pennies, while at the eight-year level the child is required to count from 20 back to 1. Such tests as these require a degree of attention, but little of a high intellectual type.

At the fourth-year level the child is required to copy a square; at the seventh, a diamond; and at the tenth, to draw designs from memory. These latter may be considered not only as a test in comprehension of details but a test in execution as well. The Binet scale is weak in tests of this type. Indeed, it has been criticised because it is very largely verbal and includes very few *performance tests*. A thoroughgoing performance test is included in the sixth year, when the child is required to tie a bow knot after being shown a model. One of the alternative tests for the tenth year is the Healy and Fernald form board. In the eighth and twelfth years is found the ball-and-field test (original with Terman). Here the child is required to indicate by drawing lines how he would attempt to locate a ball lost amid the high grass of a circular field. This test includes not only performance but the perfecting of a plan, and may be considered as one of a variety of tests demanding *ingenuity and judgment*.

In the fourth year, first appear tests demanding *discrimination and comparison*. The child is required to compare two lines and tell which is longer. In later years (fifth and ninth) he is asked to compare weights. At the fifth-year level his aesthetic judgment is tested by presenting to him paired pictures and asking him which is the prettier. In the fourth-year test, also, the child is required to identify a form shown him by indicating another like it. Tests requiring the detection of *similarities and differences* on higher levels are frequently found in the Stanford-Binet, accompanied by requirements demanding knowledge of the meaning of words and verbal comprehension. Instances of tests to determine similarities are found at the eighth-year level, when the child is asked to indicate resemblances between wood and coal, etc., and at 12, between snake, cow, and sparrow. Differences in meaning are emphasized in the test for the seventh year, when the child is asked to tell the difference between a fly and a butterfly, and again in the test for the fourteenth year, when the child is asked to point out the respects in which a president and a king differ, and finally for the average adult when differences in meaning between abstract terms are required. These *same-opposites tests*, as they have been recently termed, appear in modified forms in most intelligence tests, and they belong in a way to the large group of *controlled association tests* that have long been material for psychological experimentation. The opposites test in particular dates back a score of years, and has been found by many investigators to show rather high agreement with other evidences of intelligence.

*Linguistic knowledge* and discrimination are further evidenced by the *vocabulary tests* already spoken of and by a number of tests involving to a degree abstract and logical abilities. These qualities are called into play by the *definitions tests* beginning with the fourth year. At this level of development the child is expected to define such simple words as chair, horse, and fork, in terms of their use, e. g., "Fork is to eat with." Here, little more than simple association is demanded. However, in later years, definitions in terms of qualities of objects, such as balloon and tiger (eighth year) and pity and revenge (twelfth year), are required. In this latter test, particularly, the higher verbal and mental abilities are explored.

*Verbal fluency* on the basis of the simple associative processes is again tested in the *uncontrolled association test* measuring the "flow of ideas." This test was mentioned in the psychological literature more than 30 years ago and has since been extensively employed by investigators. The subject is asked to start with some word at a given signal and to keep on saying words as fast as he can until he has thought of, for example, a hundred, or has occupied a certain number of minutes. The test is used in the Stanford-Binet at the

tenth-year level when the child is required to think of 60 words in three minutes to pass the test. A *rhyming test*, likewise involving verbal fluency, is found in the ninth year.

Verbal ingenuity coupled with comprehension of meanings is found in the so-called *dissected-sentences test*, termed in more recent tests "disarranged sentences," given in the Stanford-Binet as one of the tests for the twelfth year. A sentence with the words mixed up, as for example, "A defends dog good his bravely master." is shown to the child, and he is asked to put the words in a sensible order. A test involving a higher degree of linguistic invention, a *sentence construction test*, is placed at the ninth-year level. Three words are given, such as desert, rivers, lakes, and the child is asked to put these into a sentence. This is a test that antedates by a number of years the Binet scale of 1905.

There are a number of tests involving *comprehension* and *rational ability* in the Stanford-Binet not so closely identified with linguistic knowledge and fluency as those above described. Among these may be mentioned the interpretation of the meaning of fables. This Ternant calls a "test of generalization." He is chiefly responsible for its development, though it was first suggested somewhat earlier. A fable is read to the child, who is asked to tell its meaning. This test is for the twelfth year. It is also an item in the average-adult test. Another rational test which may be considered to measure the *practical judgment* is called in the Stanford-Binet a *comprehension test*. It is given in varying degrees of complexity at the fourth, sixth, eighth, and tenth-year levels. At the fourth-year level such a question as the following is asked, "What must you do when you are sleepy?" At the tenth-year level, one of the questions reads, "Why should we judge a person more by his actions than by his words?"

Another test somewhat similar to the foregoing, since it relates to practical affairs and involves an element of judgment, is the *absurdities test* employed for the tenth-year level. A sentence is read to the child, who is asked to point out the absurdity contained in it, as for example, "Yesterday the police found the body of a girl cut into 18 pieces. They believe that she killed herself." A further test requiring the use of reason in a similar degree is the fact-problem question for the fourteenth year. This test is as follows: "My neighbor has been having queer visitors. First a doctor came to his house, then a lawyer, then a minister. What do you think happened there?"

Tests requiring the use of the reasoning abilities in a still higher measure are the *induction test* also for the fourteenth year and the *ingenuity test* for the superior adult.

The induction test consists in arriving at a rule from the presentation of concrete instances. The ingenuity test is as follows:

A mother sent her boy to the river and told him to bring back exactly 7 pints of water. She gave him a 3-pint vessel and a 5-pint vessel. Show me how the boy can measure out exactly 7 pints of water using nothing but these two vessels and not guessing at the amount.

Terman comments on this test:

This requires practical judgment and a certain amount of inventive ingenuity.

\* \* \* Formal education influences the test little or not at all, the unschooled business men making a somewhat better showing than high school students.

Two further tests involving rational ability should be mentioned, one the *comprehension of physical relationships* and the other the *problem of the enclosed boxes*. Both are for the average-adult level. The first named is an alternative test. The subject is asked to draw the path of a cannon ball pointed horizontally and fired across a perfectly level field.

The subject is also asked what a bucket filled with water and tipping the scales at 45 pounds will weigh when a 5-pound fish is added. Quite frequently the person tested will reply 50 pounds (the correct answer), but when the administrator of the test asks quite seriously, "How can this be correct, since the water itself holds up the fish?" The subject may become confused and reply that he is uncertain or admit that he is in error.

As a third part of the physical relations test, the subject is asked if it is harder to hit a mark of a certain size at 100 yards (providing that the gun carries for that distance) than at 50 yards? The subject must not only reply in the affirmative, but must give a satisfactory reason for his answer.

The three tests described above are interesting not only because they involve rational procedure of a fairly high degree of complexity, but also because they involve a knowledge of physical facts that are or should be a part of everyday experience, as well as of the more abstract tuition of the schools. In the present-day intelligence tests very few of these practical problems are emphasized, though there are some tests that demand *knowledge of scientific and mechanical facts*<sup>4</sup> in a limited degree, a knowledge that most educated persons seem to lack in a large degree. The second part of this test is interesting also because it measures not only knowledge of physical facts of principles, but to a degree the *ability* of the individual *to resist suggestion*,<sup>5</sup> and confidence in his own knowledge. This aspect of the test at once suggests that it is in part a will and character test as well as a test of knowledge and intelligence.<sup>6</sup>

<sup>4</sup> See particularly Thorndike's Examination for High School Seniors and College Freshmen.

<sup>5</sup> Ability to resist distractions is one of the elements included in the Downey will-temperament test, that seeks to measure certain fundamental character qualities.

The second test of the problem type, mentioned above, that of the enclosed boxes, is as follows: The person administering the tests shows the subject a cardboard box and then says, "This box has two smaller boxes inside, and each of the smaller boxes contains one tiny box. How many altogether?" The problem is further complicated in subsequent questions. This test is not only a test of rational ability, but appeals to the visualizing element as well. In the Stanford-Binet a number of visualizing tests are found. These are discussed in a later paragraph.

In recent intelligence tests *arithmetical problems* have often been included. In the Stanford-Binet such a test is placed at the fourteenth-year level. Four simple problems are given, and one minute is allowed for the solution of each problem. Intelligence is manifested in school largely as learning ability. In one sense of the word the Binet tests all measure learning ability, since they are based on materials acquired through experience, and not a few through the definite instruction of the school, as for example, tests involving the meaning of words and the distinction between words, counting, making change, arithmetical reasoning. There is, however, one test that directly involves the ability to learn something essentially novel, and it tests this *learning in operation*, so to speak. This test is one of the six main tests devised for the average adult. The subject of the test is shown a code consisting of the letters of the alphabet arranged in such a way that each is represented by a part of a geometrical figure, forming angles of various sorts. The subject is given a brief instruction in the code, and then is required to write the words, "Come quickly," when the code is removed. This test has a certain similarity to the so-called *substitution tests* (to be described later in discussing group intelligence tests), but it also involves the factor of memory, not required in the substitution tests as ordinarily employed. The factor of visual imagery also enters into the code test.

As already stated, a number of the Stanford-Binet tests are wholly or in part measures of the *ability to visualize*, a rather specialized capacity and not necessarily a mark of high mentality. The form-board test, previously mentioned, is dependent to some extent on visual ability. The clock test for the fourteenth-year level demands a considerable degree of visual imagination. In this the subject is asked such a question as the following:

Suppose it is 6.22 o'clock, that is, 22 minutes after 6; can you see in your mind where the large hand would be, and where the small hand would be? Now, suppose the two hands of the clock were to trade places. . . . what time would it then be?

A still more complicated kind of visualization is demanded in the paper-cutting test for the superior-adult level. A piece of paper about 6 inches square is folded first once in the middle and then again in the middle, at right angles to the first fold. Then a small

notch is cut in the middle of the side that presents but one edge and the subject of the test is asked to draw how the sheet of paper will look when it is unfolded. It is possible to solve this test with slight use of visual imagination, by the aid of abstraction and rationalization. However, it is primarily a test of visualization. A simpler form of a visualization test is found as early as the fifth year in the "patience, or divided rectangle test." In this test two rectangular cards are used, one divided into two triangles by cutting it along one of its diagonals. The instructions read:

Place the uncut card on the table with one of its longer sides to the child. By the side of this card \* \* \* lay the two halves of the divided rectangle with their hypotenuses turned from each other \* \* \*. Then say to the child: "I want you to take these two pieces (touching the two triangles) and put them together so they will look exactly like this" (pointing to the uncut card).

This test is interesting not only because of its appeal to visualization, but because it has in it the transition stage from what may be considered a pure performance test of the form-board type to *performance tests of the paper and pencil type*, in which positions of geometrical figures are indicated by drawing lines on paper. This paper and pencil adaption of the form-board test is used extensively in group intelligence tests (to be described later) for small children and illiterates, of which the Army Beta test is the prototype.

A description of the Stanford-Binet should not conclude without a description of the *directions test*. At the fifth-year level the child is asked to execute three commissions, as follows:

Here's a key. I want you to put it on the chair over there; then I want you to shut (or open) that door, and then bring me the box that you see over there. Do you understand? Be sure to get it right. First put the key on the chair, then shut (open) the door, then bring me the box. Go ahead.

This test involves attention to the instructions given, a comprehension of them, and sufficient immediate memory to execute them. It is far less dependent on instruction in the schools than the majority of the tests we have discussed. The directions test, as will be pointed out later, has in various forms been long employed by psychologists. It occupies an important place in the Army Alpha and Beta, and is found in many of the group intelligence tests now commonly in use.

A summary of the nature and scope of the Stanford-Binet includes the following important facts:

1. The types of tests used, arranged approximately in order of their frequency, are:

- (a) Range of information tests, designed to measure the child's familiarity with objects of common experience, this knowledge to a considerable extent gained outside of specific school instructions. These tests include, roughly, one-quarter of the total number.

(b) Tests involving rational abilities, ranging from the most simple to the more complex. These include comparison of objects and words, the noting of similarities and differences in these, ingenuity, ability to generalize, and ability to solve problems. In these tests are included the ability to form practical judgments and to comprehend and interpret physical relationships. In all, nearly a quarter of the tests are to be classified under this second head. These tests are predominantly verbal and abstract in their nature and closely related to schooling.

(c) Memory-span tests and tests for immediate memory of ideas. Nearly a fifth of the tests are of the memory-span type, and not closely related to school progress.

(d) Tests primarily depending on verbal fluency and ingenuity. These included a free-associations test, a dissected sentence test, sentence construction test, a rhyming test, and several definitions tests. A considerable number of vocabulary tests (primarily classified as range of information tests) might also be included here. Further, the whole series of Binet tests is pronouncedly verbal in its nature, and in this particular the result is definitely dependent on schooling.

(e) Tests involving knowledge of numbers and their relationships. Here are included tests in counting, making change, etc., as well as the single test in arithmetical problem solving. These tests are largely dependent on the formation of habits of simple skill most commonly taught in the schools.

(f) Tests involving concrete visualization and eye-imagery. These tests detect a rather special kind of ability that is not generally emphasized in school training. However, the schools do place emphasis on that kind of visual imagery involved in the spelling and recognition of words and symbols. This latter is, however, of the abstract sort, and is not definitely tested in the Binet scale.

(g) Performance tests are found in a few instances. In these the individual is required to do something, as distinguished from knowing something. In a sense all kinds of simple and complicated skills are evidences of performance. Here, however, the term is applied to something definitely related to manual dexterity, such as tying a bow-knot, finding a missing ball, and the execution of a form-board test. As has already been said, such tests as these are infrequent in the Binet scale. A test closely allied to this type is the directions test (executing three commissions), discussed above.

(h) Other tests, occurring but once and not as easy to classify under the preceding analysis, are:

Detecting parts in a mutilated picture (on the whole a range of information test).

Interpreting the meaning of a picture (possibly to be classified as a reasoning test).

Writing two words, according to a code previously studied (a test for learning and memory).

Detecting absurdities in a statement (involving knowledge, comprehension, and at times a sense of the significance of words).

2. The dependence of the tests on the child's experience and particularly his schooling is to be noted. This has previously been pointed out. It is quite clear that the validity of these tests is based on the assumption that children tested have all had a common experience, and hence that the differences are not due to differences in training, but to differences in innate mentality. This important point will be referred to and developed later, since it is essential in the whole theory of intelligence testing.

3. It is an important fact to be noted particularly in a later discussion of the nature of native intelligence that the Stanford-Binet includes in its scale so many tests that are really the measure of acquired ability, rather than an indication of abilities now in operation. They measure the results of mentality in acquisition of knowledge and skill, much more than they measure mentality in its immediate operation, so to speak. That is, a vocabulary test shows the product of previous learning, not learning in progress. The same is true of a counting test, and the many tests that appeal primarily to perceptions already formed. On the other hand, for example, the memory-span tests, ingenuity tests, and the like involve an active and alert mind, and build on experiences already acquired. Nevertheless, all of these tests rest on a definite basis of acquired experiences, and only to a limited degree test experiences in the making. This very important fact is characteristic not only of tests of the Binet type, but of all intelligence tests so far devised. This fact will be discussed at length in a later section.

### III.—ARMY ALPHA TESTS.

Since the great impetus in group intelligence testing was brought about by the use of the Army tests, and since the paper and pencil tests that are now in use are more or less directly constructed on the basis of the Army tests, a brief analysis of these tests will be helpful in an understanding of the tests more recently devised.

First, in the construction of these tests the following important points<sup>6</sup> were kept in mind, considerations that should apply to the formulation of all tests of this general type:

In the first place, the tests were so constructed that large groups could be examined rapidly. Further considerations emphasized the necessity of devising tests that should measure as far as possible native intelligence rather than acquired knowledge and skill (i. e., mental ability rather than schooling.) In the third place, the tests attempted to measure intelligence of low as well as of high levels.

<sup>6</sup> Army Mental Tests, Yerkes and Yerkes, New York, 1920, pp. 2 to 7.

They were so constructed that those of poorest mentality could make a score of some sort, while those of the highest intelligence would still be enabled to obtain a perfect score. It is clear that if on the one hand any considerable proportion of a group tested received zero scores that such a test would be an unsatisfactory measure of the abilities of the group. It is equally clear that if a large proportion received perfect or very high scores the test would be equally undesirable. In order that both of these extremes should be avoided great care was taken in the construction of these tests not only to include materials sufficiently easy for those of the lowest mentality likely to be measured, and sufficiently difficult for those of the highest mentality, but also to make the time for administering the test sufficiently long for the majority tested to have opportunity to achieve scores up to their capacity, and sufficiently short to require mental alertness for completion of the tests within the limits necessitated by the practical conditions of the examination. A time limit of 50 minutes was finally fixed.

The tests were arranged so that the scoring could be objective. Personal opinion in regard to the correctness of the answers given was practically limited. For each test and for each subtest one answer and only one answer was correct. Stencils (keys) to the right responses were provided, so that any intelligent and careful clerk, although he might not be able to obtain a high score himself, could grade with entire accuracy any individual test, and would be in entire agreement with any other similar person grading the same test. When the fact of the enormous variability in teachers' grades of the ordinary school examinations is remembered, this emphasis on objective scoring is seen to be absolutely necessary if reliable results are to be obtained. Further, not only was objective scoring insisted on, but complete uniformity of administration of the tests was also required, a very important consideration in the use of all tests in which various groups of individuals and various individuals within a group are to be compared.

Among other considerations kept in mind in the construction of the tests, the most important were that alternative examinations, tests of the same character, but different in content, should be provided to prevent the possibility of coaching; that there should be required only a minimum amount of writing on the part of the individual tested; and that the materials should stimulate the interests of those examined.

With these purposes in mind the committee constructed a preliminary series of tests, using in this work their knowledge of mental tests that had already been employed, and securing the judgment of competent psychologists as to the relative values of these prelimi-

nary materials. The next step consisted in actually trying out the preliminary series and studying the scores obtained for purposes of further revision. When two subtests were found that agreed very closely in the results secured this was an indication that one was superfluous, since both apparently measured the same thing. When, on the other hand, a very slight relationship between a subtest and the total test score was found, this was taken as an indication that the test in question was of little value in measuring the mental abilities of those examined.

Another important consideration kept in mind by the committee in studying the results obtained in these preliminary tests was their so-called reliability, namely, Do the test results obtained on a first trial agree with results obtained on subsequent trials?—a very important matter to have in mind in judging the value of tests of this type.

After various preliminary try outs and revisions according to these findings, a general intelligence test was formulated that was tried out in the fall of 1917 on about 80,000 men in the United States Army. As a check on these results about 7,000 students in colleges, high schools, and elementary schools were given the same examination. Then more than two months were spent on the careful study of these results.

The above somewhat detailed statement of how the Army tests were prepared is given in order to emphasize the fact that valid mental tests are not the outcome of individual opinion or clever guessing, but are the result of careful and painstaking study and statistical treatment. Only in this way has mental testing advanced to the position that it now holds, and only in this way can it hope to develop further and become a more perfect instrument for the guidance of teachers and school administrators.

Test No. 1 in the Army Alpha, the *directions test*, has been used in various forms in a number of the later group tests. In the Army Alpha the test is administered orally, as has been pointed out in the above discussion. In the more recent tests the directions are printed and the subject follows these instructions. A further difference is found in the fact that when administered in the latter manner the instructions need not be carried in mind since they can be referred to from time to time. Thus a less degree of mental alertness and a minimum of memory span is required.

In the Otis group intelligence scale advanced examination, Form A, the tests take the following form in connection with the letters of the alphabet printed at the top of the test sheet:

Direction 8. Write the letter which follows the third letter of the alphabet.

Direction 19. Find the letter which in this sentence appears a second time nearest the beginning. Write it, using a capital.

In the Otis self-administering tests of mental ability a variation appears as follows:

If all the odd-numbered letters in the alphabet were crossed out, which would be the tenth letter not crossed out? Print it.

For measuring intelligence among little children the directions test is used generally in connection with pictures,\* as, for example:

Look at the cat and the rat. Draw a line from the cat's mouth to the rat's tail, or

Draw a ring around the mouse.

In the Smith College tests the directions test is closely modeled after the Woodworth and Wells hard directions test, an example of which has been given above. Roughly a dozen of the most commonly used intelligence examinations employ in some form the directions test.

Test No. 2 in the Army Alpha is an exercise in simple *arithmetical problems*. As has been pointed out, a test practically identical with this is found in the Stanford-Binet. In recently devised group tests arithmetical problems occupy an important place, being included in nearly all the scales that attempt to measure intelligence in the upper grammar grades, high school, and college. Clearly this test measures intelligence through a specific skillful activity largely dependent on schooling. Even to a greater extent tests based on exercises in fundamentals in arithmetic, not included in the Army tests, not in the Stanford-Binet, but found in a number of group intelligence tests for school and college, measure a specific attainment in learning acquired in the elementary school.

Test No. 3 in the Army Alpha is a common-sense or *best-reasons test*. It requires practical judgment of a very simple sort, but can be complicated to include rational abilities of a fairly high order. It is as follows:

Below are 16 questions. Three answers are given to each question. You are to look at the answers carefully; then make a cross in the square before the best answer, as in the sample:

Sample:	Why do we use stoves? Because— <input type="checkbox"/> they look well, <input checked="" type="checkbox"/> they keep us warm, <input type="checkbox"/> they are black.
---------	--

The best reasons test is found in about a third of the important group intelligence examinations now in common use. Although it is not definitely copied from the Stanford-Binet it relates itself clearly to the practical judgment tests in this scale.

\* See:

Haggerty, Delta I, Grades 1-3.  
 Detroit Kindergarten and Detroit First Grade.  
 Otis Primary, Form A.  
 Dearborn, Grades 1-3.  
 Myers' Mental Measure.

Test No. 4 in the Army Alpha is a *same-opposite test*. Two words are paired and after each are the words "same-opposite." The subject of the test is to indicate the relationship of the paired words by underlining either *same* or *opposite*, as for example:

good—bad\_\_\_\_\_same—opposite.  
 little—small\_\_\_\_\_same—opposite.

This test depends very largely on knowledge of the meaning of words and thus is dependent to no small extent on schooling, as is clearly evidenced by the more difficult parts of the test; for example:

largest—donation\_\_\_\_\_same—opposite.  
 encomium—eulogy\_\_\_\_\_same—opposite.

This test finds no exact counterpart in the Stanford-Binet. As has been pointed out, there are, however, various tests in this scale that have to do with likenesses and differences.

The same-opposite test in its various forms is one of the tests most frequently employed in the group tests now in common use. At times it appears as a same-different test in which the paired words or objects may be related through similarity or difference (not necessarily opposition). For example, numbers, proper names, symbols, etc., are printed in a column or on an entire page. In another column, or on a separate page, those numbers, names, and symbols that correspond with those in the first column are scattered among others that do not so correspond. These former are to be picked out and identified. In examinations framed for little children another variation is found. Pictures are to be compared for the purpose of discovering similarities of simple words; are to be identified, and differences are to be detected in words orally presented.

Still other variations of this test are as follows:

Select from the parentheses opposite the first word the word that has nearly the same meaning and underline it:<sup>9</sup>

Prolix\_\_\_\_\_ (terse, profound, drastic, verbose).

Underline the word in the parentheses that is the best opposite of the word at the left:<sup>9</sup>

1. East\_\_\_\_\_ (north, west, south, pole, equator).

25. Unless\_\_\_\_\_ (and, therefore, however, also, if.).

An early form of the opposite test, still preserved in a number of the present group tests, consists of a series of words after each of which an opposite is to be written.<sup>10</sup> The objection to this form

<sup>9</sup> Smith College tests, Form 2.

<sup>9</sup> Compare:

Chicago, Freeman & Rugg, Form B.

Smith, Univ. of Montana, High School and College.

Otis, Self-Administering, Higher Examination, Form A.

Brown University Intelligence Examination.

<sup>10</sup> See:

Mentimeter 2.

Pressey—Indiana Mental Survey, Scale 1.

of the test is in the difficulty of scoring it objectively since the individual judgment of the person correcting the test determines whether a correct opposite has been given or not. When, however, several words following the stimulus word are printed in parentheses, one and only one of these words can be correctly underlined, and thus the scoring is independent of the personal equation of the individual correcting the test, a very important matter in the administration of mental tests. In the form in which the test is given in the Army Alpha the scoring is quite objective, but the opportunity for guessing is great, since without actual knowledge of the relationship of the words paired, the individual taking the test may (and on the whole will tend to) get half of the answers right by mere chance. This has made it necessary in this test, and indeed in all tests where only two replies are possible, one right and the other wrong, to score such a test by subtracting the number wrong from the number right. However, the justice of this procedure has recently been criticised severely in a certain discussion.

The safest way probably is to provide in all tests a number of alternative replies, thus making negligible the factor of mere chance in the response.

Test No. 5 in the Army Alpha is the *disarranged sentence*. It is as follows:

Below are 24 mixed-up sentences. \* \* \* Think what each would say if the words were straightened out, but don't write them yourself. Then if what it would say is true, draw a line under the word "true," but if what it would say is false, draw a line under the word "false." If you can not be sure, guess:

Samples: { a eats cow grass ----- true false  
           : horses feathers have all ----- true false

It is to be noted that this test consists essentially in the correct rearrangement of the mixed-up words and not in the judgment as to whether the statements resulting are true or false. The underlining of one of these words is merely to indicate whether the mental operation of rearrangement required has been correctly performed. The device saves time in response and leads to ease and objectivity in scoring. However, like the test previously described, the chances of guessing correctly are one out of two.

This test, as has previously been pointed out, is found in the Stanford-Binet scale at the twelfth-year level in practically the same form as it appears in the Army Alpha.

The disarranged-sentence test is used in about a quarter of the most common group examinations. In some of its variations it avoids the factor of guessing found in its army prototype. For example, the Otis self-administering examination has the following:

If the words below were arranged to make the best sentence, with what letter would the last word of the sentence end? Print the letter as a capital:  
 sincerity, traits, courtesy, character, of, desirable, and, are \_\_\_\_\_ ( )

The Pressey Indiana Mental Survey, Scale 1, uses still another device. A sentence of which the following is an example is given: "asked me way I girl the to show the." Then comes the instruction—"Put one line under the first word of the sentence; two under the last."

The Pressey Cross-out tests place among the words in the disarranged sentence one not belonging there. The response required is to cross out this misplaced word.

Still another variation in the form of a directions test is in the Miller mental ability test for grades 7 to 12 and for college freshmen. Examples are:

Don't rewrite the sentences. Do what they tell you to do.

This underline sentence.

Comma the a place after word longest this in.

Test No. 6 in the Army Alpha series is a *number completion test*. Six numbers bearing a certain relation are arranged in order. It is required that two numbers be added having this same relationship. Examples are:

3 4 5 6 7 8 — —  
3 6 8 16 18 36 — —

No test of this exact type is in the Stanford-Binet scale. Roughly a third of the important group intelligence examinations have this test in some form. One of its variations as found in the Chicago, Freeman and Rugg tests, Form B, is as follows:

A. 1 2 3 4 5 — 7 8 9 — 11 12 13  
J. 84 — — — — 54 — — — 30 — — — 6

In the Déarborn tests, revised series 2, grades 4-12, this test appears as follows:

Write numbers in the empty spaces so that the sums will be the same if added by column up and down, or across by rows.

Example:     5  
              5   5  
              4   3   7

In place of dots supply numbers to give answers as printed.

Sample:     9   3   1  
              7   .  
              —  
              1 4 9 '2

Test No. 7 in the Alpha series is a so-called *analogies or mixed-relation test*. It is as follows:

In each of the lines below, the first words are related to each other in some way. \* \* \* See what the relationship is between the first two words and underline the word in the heavy type which is related in the same way to the third word.

1. gun—shoots—knife—run cuts hat, bird
40. cold—ice—heat—lightning warm steam coat

This test has no counterpart in the Stanford-Binet. It has, however, been known to psychologists for some time, being first used by Binet in England in 1911.

A large number, at least two-thirds, of the important group intelligence examinations use in some form a mixed-relations test. Although a considerable proportion of these follow the Army type, the test appears in a number of variations, examples of which are as follows:

Egg: Bird:: (?) : plant. . . . seed, shell, leaf, root.

(Otis Group Intelligence scale advanced examinations).

Woman: Girl:: man: ?

(Pressey Indiana Mental Survey, Scale 1.) (Smith College Tests.)

Underline the two words that have the same logical relation to each other as locomotive and train.

Station horse hub baggage buggy

(Thurstone's College Test.)

A variation in Thorndike's examinations for high school seniors and college freshmen (also in Smith's University of Montana high school and college tests) consists of three geometrical figures to which a fourth is to be added that bears the same relation to the third as the second bears to the first.

The last of the Army Alpha tests is a *range of information test*. It is as follows:

In each of the sentences below you have four choices for the last word. Only one of them is correct. In each sentence draw a line under one of the four words which makes the truest sentence. If you can not be sure, guess.

1. America was discovered by Drake Hudson Columbus Balboa.

40. Scrooge appears in Vanity Fair The Christmas Carol Romola Henry IV.

The Stanford-Binet has no test of this exact type, but as has been pointed out about one-fourth of the scale is composed of tests of the informational type. The range of information test appears in some form in nearly all intelligence tests now in use. It is based entirely on what has been learned in school and out and involves little of ingenuity and judgment. In addition to the form in which it appears in the Army Alpha we find these variations:

Put a line under the thing that is the largest and a cross after the thing that is the smallest.

Man  
Cow  
Hen  
Dog  
Cat

(Pressey Indiana mental survey, scale 1)

In what sports and occupations are the following objects used? Answer with a single word.

Scenario	-----	} (Smith College tests.)
Halyard	-----	
Casserole	-----	
Last	-----	
Zither	-----	

1. Do dogs run? ----- Yes, No.

40. Are judicial decisions ever enforced? ----- Yes, No. "

Put a plus (+) sign before each statement that is true and a minus (-) sign before each statement that is false.

1. ( ) A lake is bigger than a pond.

25. ( ) A pint of cream weighs more than a pint of milk.

In this last test cited is some chance for judgment and inference in item 25. In the Miller Mental Ability Test this factor is still more in evidence, as shown by the following:

The word you underline must be a NOUN that may be an EFFECT of the first word: EDUCATION (high school, books, knowledge, study).

In the Detroit first grade and kindergarten tests, in which pictures are employed in place of words, the range of information test approximates very closely the Binet originals. Examples are as follows:

Mark two things we cut with.

Mark five things we dig out of the ground.

When pictures of the crescent moon, a cup, and a sled are shown, the examiner says to the child, "Show me the thing that is up in the sky."

Again four pictures are shown and the child is asked to point to the two that show that it is summer. Here inference as well as familiarity is involved in the response.

By way of summary it may be said that the eight tests of the Army Alpha, namely, (1) *directions test*; (2) *arithmetical problems*; (3) *best reasons test*; (4) *same-opposites test*; (5) *disarranged sentences*; (6) *number completion*; (7) *mixed relations*; (8) *range of information*, have been extensively employed in various forms in subsequent group tests. Not one of these eight originated in principle in the Army Alpha. Tests Nos. 2 and 5 are copied almost exactly from the Stanford-Binet; tests Nos. 1, 3, and 8, are closely related to similar tests found in the Stanford-Binet. Of the remaining three tests No. 7 was devised about seven years before it was incorporated into the Army Alpha, while test No. 4 belongs to a still earlier period. The number completion test alone possessed an element of novelty. The contribution that the Army Alpha made to the development of mental testing was not in the originality of the tests themselves, but in the fact that a battery of tests were arranged and standardized, capable of being given to a large group of indi-

<sup>22</sup> Haggerty intelligence examination, Delta 2; National intelligence tests, Scale B, Form 1.

viduals at the same time, and of being rapidly and accurately scored by individuals who need have no knowledge of the tests themselves, nor ability greater than that of reasonably intelligent and conscientious clerks. Further, these tests are notable because they were given to a very large group of individuals, and demonstrated ultimately their intrinsic worth as an instrument for discovering the mental ability of those tested and for classifying such individuals in such a way as to be of practical service in the organization of the United States Army.

#### IV.—ARMY BETA TESTS.

As has been previously stated the Army Alpha test was found to be applicable to only about 7 men out of 10 tested, since nearly 30 per cent of the draft Army could not "read and understand newspapers and write letters home." For these illiterates as well as for foreign-born men who knew little English, a new type of test was necessary. This was designated as the Army Beta. This latter consisted of seven tests as finally perfected, although originally eight were included to correspond to Alpha.

The first of the Beta tests requires the subject of the examination to trace with a pencil the correct path through five increasingly complex mazes represented on paper. The *maze test* was first used nearly a quarter of a century ago to measure the learning ability of white rats and other brutes. Later it was reduced to the paper and pencil form and used in experiments in human learning. To-day it is found in a number of group intelligence examinations, following in this respect the form used in the Army Beta<sup>12</sup> or a variation containing the essential principle. This test requires foresight and planning and can be varied from a simple problem-solving test to one of great complexity.

The second test in Beta is styled *cube-analysis*. Sixteen pictures of cubes arranged in piles are shown, and the subject of the test is required to indicate the number in each pile, varying from 2 to 50. In the first four piles, all the separate blocks can be seen on one or more of their faces. As the number increases, the blocks entirely hidden become more numerous. Thus this becomes a problem in visualization of an increasing degree of complexity. When visualization fails, inference and reasoning are necessary. This test is closely related in principle to a number of the tests found in the Stanford-Binet. Visualization tests in various forms are employed in recently devised intelligence tests. Dearborn, Series 2, grades 4-9, employs this test in the Beta form. In his primary ex-

<sup>12</sup> See :

Mentimeter 2; Dearborn, series 2, grades 4 to 9.  
Otis, primary 1, Form A.



symbol test. In all these tests, however, the fundamental principle is the same.

Test five in the Beta is a *number checking test*. A paired series of 50 numbers are presented. In half the series the first and second part of the pairs are identical, in the other half there are differences. The subject is required to check the identical pairs. This test is similar to test 4 in the Alpha, which, as already pointed out, is related to the various likeness and difference tests of the Stanford-Binet scale, and to the many forms of the same and opposites tests that have later developed in group testing. Many of these, as pointed out above in the discussion of the Alpha same-opposites test, have closely followed the Beta model, requiring the identification of names, symbols, and numbers. The number checking test has also been used in trade tests measuring the probable abilities of clerical workers.

The sixth element of the Beta series is a *picture-completion test* of 20 items. This type of test has already been discussed in our consideration of the Stanford-Binet examination, and as has been pointed out it is finding extensive use to-day in group tests devised particularly for children in the lower school grades.

The seventh element of Beta is called "*geometrical construction*." It is an adaptation of the form-board test reduced to pencil and paper. There are 10 separate items, each item consisting of a blank square at the left, and various smaller geometrical figures at the right. These smaller figures fit exactly into the square if properly placed. The subject of the test is to indicate by drawing lines in the square just what this arrangement should be. This test is found also in Mentimeter 2, the instructions reading: "Draw lines in the large figure at the right in such a way as to make of it the small figures at the left." A similar test is found in Dearborn, Series 2, grades 4-9.

#### V.—RECENT GROUP TESTS.

In the Stanford-Binet and in the Army group tests are to be found in some form most of the elements employed in the group tests now in use. Some of these parts have been taken over without change, while others have been modified in various respects. However, certain elements in recent group tests are sufficiently unlike those found in the earlier tests to warrant description and comment.

One of the most important of these is the *sentence completion test*. This test was devised by Ebbinghaus about 1905 in connection with his investigations of fatigue among school children in the city of Breslau, Germany. As used by its originator the test consisted of a paragraph in which syllables were elided. The child

was required to supply these omitted syllables. Terman later used this test but omitted words in the place of syllables. He did not, however, include it in the Stanford-Binet. This test subsequently appeared in the form of a series of sentences in which the omissions became more and more difficult to supply. In this form it was standardized by Trabue as a language test, and again worked over and restandardized by Kelley. In its present form it appears to be one of the best measures of intellectual ability that have so far been devised.

A typical form of this test appears in Miller's Wisconsin examination. Twenty sentences are given, ranging from easy to difficult, as, for example:

1. The dog ——— black.
20. ——— should prevail in churches and libraries.

It is found in this same form in nearly a dozen other examinations.<sup>14</sup>

This test is used in its original form as devised by Ebbinghaus (word completion) in the Presscy Indiana Mental Survey, Scale 1. It is combined with an opposite test by Dearborn,<sup>15</sup> as follows:

Supply words meaning the opposite of underlined words:

1. We had a hot summer and a ——— winter.
24. The ignorant believed; the ——— ———.

He also uses it in connection with disarranged proverbs in the following way:

Rearrange and supply the missing words:

1. In time saves stitch ——— ———.
10. Does make a swallow not one ———.

One of the objections in the use of this test is encountered in the fact that it is difficult to score on an absolutely objective standard, since it is frequently possible to make completions reasonably sensible by employing any one of a number of words. Take for example the following sentence: "A ——— man ——— not invest ——— in ——— securities."

It may be filled in thus: "A *wise* man *will* not invest *money* in *wildcat* securities." Or thus: "A *foolish* man *does* not invest *money* in *sound* securities." Or thus: "A *business* man *should* not invest *much* in *speculative* securities." Or thus: "A *poor* man *can* not invest *earnings* in *expensive* securities."

These variations may be multiplied almost indefinitely. Evidently some of the substitutions that make sense are better than

<sup>14</sup> See particularly: National, Scale A, Form 1; Whipple's Grammar Grades; Smith's University of Montana, High School and College Test; Smith College Intelligence Tests; Thurstone College Tests; Thorndike High School Seniors and College Freshmen Tests; Brown University Psychological Examinations.

<sup>15</sup> Series 2, grades 4-9.

others, but how much better? It is here that the opinion of the person correcting the test plays an important part, and introduces an element of uncertainty into the scoring. In order to avoid this uncertainty the following variation of the test is used in Otis Advanced Examination, Form A:

Once upon a ——— there was a ——— (time, place, man.)

The person taking the test is instructed to underline the word in the parentheses that belongs in the blank space. A similar device is used in the Smith College Test, Series 2. While this method makes the scoring of the test entirely objective it materially changes its character and greatly reduces its difficulty.

Another important test that has been added to recent group intelligence examinations is taken directly from one of the school attainment tests. It is a *reading test* and appears in a number of forms. In the Thorndike High School Senior and College Freshmen Examination, a paragraph of rather difficult prose is followed by a number of questions based on the paragraph. This same form is used in the Batson South Dakota High School Tests. In Menti-meter 2 the test is as follows:

If the oscillation of the pendulum were not facilitated by any other force than gravity, what would be the effect upon their amplitude? It would gradually be flagellated, augmented, swaged, diminished.

Check the best reply.

Although this is classed as a reading test it is at the same time a range of information test, a vocabulary test, and a directions test. However, reading is such a complex activity that it is difficult to employ a test that measures it in any simple and direct way.

In the various forms of the Binet test and in the Army tests some attempt has been made to measure what may be termed, for want of a better name, *rational ability*. However, as has already been pointed out above, many items in these tests are not concerned in any direct way with complex mental processes. In recent group tests a number of elements are found that are specifically intended to measure these abilities. Some have a place in the earlier tests already discussed, while others are novel in form if not in principle.

Among the tests that seek specifically to get at the element of thinking in a response to a mental examination, those which are relatively independent of specific past experience should be given an important place. Many tests that involve reasoning ability are so tied down to detailed knowledge that it is difficult to determine how much is actually due to acquisition of a definite sort and how much is due to the factors of analysis, selection, judgment, and so on.

For example, the *essential attributes* test, quite frequently employed in various forms<sup>10</sup> demands abilities that go beyond mere memory and knowledge of facts and of words and their significance. However, it is essentially conditioned on such knowledge. An example of this test is as follows, taken from National, Scale A, Form 1:

In each row draw a line under each of the two words that tell what the thing always has:

1. Elephant (circus, ears, hay, keeper, trunk).
24. Fiction (falsehood, hero, imagination, impossibility, invention).

A variation (Otis Self-Administering, Higher Examinations, Form A) instructs the person taking the test to cross out among a number of words the one which does not belong in the series, as follows:

A bird does not always have (?)

1. wings, 2. eyes, 3. feathers, 4. a nest, 5. a bill

The essential attributes test is the same in principle as the various *classifications tests*, previously discussed. In the majority of these, verbal knowledge plays a very important part.

---

<sup>10</sup> See: National, Scale A, Form 1: Pressy Indiana Mental Survey, Scale 1; Whipple's Grammar Grades; Otis Self-Administering, Higher Examinations, Form A.