

A TENTATIVE REVISION AND EXTENSION OF THE BINET-SIMON MEASURING SCALE OF INTELLIGENCE.

PART III. SUMMARY AND CRITICISMS.

LEWIS M. TERMAN AND H. G. CHILDS.

Leland Stanford University.

It is evident from the results of our investigation that the Binet scale requires a radical revision to make it at all suitable to conditions in this country. We wish to submit as a basis for further work the following series of tests, which is based chiefly upon the tables of results presented in the preceding articles of this series, but partly upon the results secured by Bobertag and by Binet in 1911. Although we realize that no series of tests can claim more than a rough approximation to accuracy until it has been given a painstaking try-out with many hundreds of children under widely-varying conditions of culture and training, we believe, nevertheless, that the series here offered will prove a more serviceable basis for further exploration than either the 1908 or 1911 series of Binet. *At the same time, the frankly tentative nature of our revision should be remembered.*

REVISED SCALE.

N. B.—Examiner should check all items separately.

YEAR III (unit value .1666)

1. a. "What's the thing to do when you feel sleepy?"
b. "What's the thing to do when you feel cold?"
(Both must be passed.)
2. Weights:—3-12 6-15 3-12. (Two out of three.)
3. Naming objects: key closed knife penny
(no failure).
4. Knows sex.
5. Repeats three digits: 641 352 783 (Two successes.)
6. Pictures (enumerates) 1 2 3

YEAR IV (unit value .2).

1. Copies square.
2. Counts four pennies
3. Divided rectangle.
4. Chooses prettier: 1 . . . 2 . . . 3 . . . (must be no failure).
5. Defines (use): a table . . . a chair . . . a horse . . . a mamma . . . (three out of four).

YEAR V (unit value .2).

1. Morning or afternoon.
2. Names colors: red . . . yellow . . . blue . . . green . . . (no failure).
3. Puts key on chair, brings box, shuts door.
4. Vocabulary index (score?). (See third article of this series.)
5. Repeats: a. "We expect to have a great time down at the seashore."
b. "When the train passes, the engineer will blow the whistle."
c. "Why should anyone want to harm a beautiful bird?"
(One success out of three.)

YEAR VI (unit value .1666).

1. Right hand, left ear.
2. Number of fingers.
3. Counts thirteen pennies.
4. Repeats: 4739 . . . 2854 . . . 7261 . . . (Two out of three.)
5. a. "What's the thing to do if it's raining when you start to school?"
b. "What's the thing to do if you have missed a train?"
c. "What's the thing to do if you find that your house is on fire?" (Two out of three.)
6. Vocabulary index 12 per cent. (2160).

YEAR VII (unit value .1666).

1. Copies diamond.
2. Omissions from pictures: 1 . . . 2 . . . 3 . . . 4 . . . (Three out of four.)
3. Names: penny . . . nickel . . . dime . . . quarter . . . (No failure.)
4. Pictures (describes): 1 . . . 2 . . . 3 . . . (Two out of three.)
5. Vocabulary index 14 per cent. (2520).
6. Repeats sentences 14 to 16 syllables:
 - a. "We will go out for a long walk. Please give me that pretty straw hat."
 - b. "We are having a fine time. We found a mouse in the trap."
 - c. "Brother Frank had a fine time on his vacation. He went fishing every day." (One out of three.)

YEAR VIII (unit value .1666).

1. Writes from dictation: "The pretty little girl."
2. Compares: butterfly—fly; wood—glass; paper—cloth. (Two out of three.)
3. Counts 20 to 0 (twenty seconds).
4. Ball and field (score 2). (See third article in this series.)
5. Repeats: 31759 . . . 42385 . . . 98176 . . . (Two out of three.)
6. Vocabulary index 18 per cent (3240).

YEAR IX (unit value .125).

1. Value of stamps: 111222.
2. Names date.
3. Reading for four memories. Time
4. Weights, 3—6—9—12—15. Trial 1 2 3
5. Fifty words in two minutes.
6. a. "What's the thing to do when you have been struck by a playmate who did not do it purposely?"
b. "What's the thing to do when you have broken something which does not belong to you?"
c. "What's the thing to do when you have been detained so that you are in danger of being late for school?" (Two successes out of three.)
7. Vocabulary index 23 per cent. (4140).
8. Completion test (score . . .). (See third article in this series)

YEAR X (unit value .125).

1. Copies designs: 1 2 (See Binet 1911 series)
2. Three words in one sentence.
3. Six digits: 374859 825746 762953 (Two out of three correct.)
4. Bull and field (score 3).
5. Vocabulary index 26 per cent. (4680).
6. Fables 1 2 3 4 (See second article in this series)
7. Completion test (score 20).
8. Makes change (25c., 4c.).

YEAR XI (unit value .1666).

1. Arithmetical reasoning (score ?). (Using the following problems taken from Bonser's "Reasoning Ability of Children in the Fourth, Fifth and Sixth Grades," page 2. The child is given the problems in the following form, and is asked to write the answer after each problem, making no other figures:)
(a) If three-quarters of a gallon of oil costs 9 cents, what will 7 gallons cost?
(b) At 15 cents a yard, how much will 7 feet of cloth cost?
(c) A man whose salary is \$20 a week spends \$14 a week. In how many weeks can he save \$300?
(d) How many pencils can you buy for 50 cents at the rate of 2 for 5 cents?
(e) A man spent two-thirds of his money and had \$8 left. How much had he at first?
2. a. "What ought one to do before taking part in an important affair?"
b. "What ought you to say if someone asks your opinion about a person you only know a little?"
c. "Why ought we to judge a person more by his acts than by his words?"
d. "Why do we excuse a wrong act committed in anger more readily than a wrong act committed without anger?" (Three out of four correct.)
3. Vocabulary index 30 per cent. (5400).
4. Fables: 1 2 3 4
5. Completion test (score 25).
6. Sees the point in following samples of wit and humor. E reads each passage and asks S "What is the point of that joke?" "What is funny about that?" etc.

- (a) A man called at the postoffice to inquire if there was a letter for him. "What is your name?" said the postmaster? "Sure," said the man. "you'll find my name *on the back of the letter.*"
- (b) A woman was once told of a man who had twice had small-pox and had died of it. "Did he die the first time or the second?" the woman asked.
- (c) A young fellow who wanted to be witty once said to a barber. "Did you ever shave a monkey?" "Why, no, sir," said the barber; "but if you will please sit down, I will try."
- (d) A religious old lady used to say that God was very good to make the greatest rivers flow past the largest cities.
- (e) A peddler in his cart overtook another peddler on the road and thus addressed him, "Hallo, what do you carry?" "Drugs and medicines," the other replied. "Go ahead, then," said the first; "I carry gravestones."
(Three out of five satisfactory.)

YEAR XII (unit value .1666).

1. Absurdities, as follows:
 - (a) An unfortunate bicycle rider broke his head from a fall and died instantly. He was picked up and carried to a hospital, and they do not think he will recover.
 - (b) I have three brothers—Paul, Ernest and myself.
 - (c) There was a railroad accident yesterday, but it was not serious. The number of dead is only 48.
 - (d) Yesterday the police found the body of a young girl cut into 18 pieces. They believe that she killed herself.
 - (e) The engineer said that the more cars he had on his train the faster he could go. (Four out of five correct.)
2. Disarranged sentences: 1 2 3
3. Reading for seven memories (Time)
4. Suggestion. (See Binet, 1911.)
5. Vocabulary index 36 per cent. (6480).
6. Repeats 26 syllables (one out of three correct).
 - a. "My little children, you must work very hard for a living. You must go every morning to your school."
 - b. "The other day I saw in the street a pretty yellow dog. Little Bessie has spots on her new apron."
 - c. "Ernest is often punished for his bad conduct. I bought at the store a pretty doll for my little sister."

YEAR XIII (unit value .1666).

1. Seven digits: 2183439 9728475 3247196
(Two out of three correct.)
2. Vocabulary index 42 per cent. (7560).
3. Fables: 1 2 3 4
4. Arithmetical reasoning (score?)
5. Completion test (score 36).
6. Problems of fact, as follows (two out of three):
 - (a) My neighbor has been having queer visitors—first a doctor, then a lawyer, then a priest. What's happening at my neighbor's?
 - (b) An Indian coming to town for the first time watched a white man riding along the street. As the white man rode by, the Indian said: "White man lazy; him walk sitting down." What was the white man sitting on?

- (c) A man who was walking in the woods near a city stopped suddenly, very much frightened, and ran to the nearest policeman, saying that he had just seen, hanging from the limb of a tree, a ———.

YEAR XV (unit value 1666).

1. Fables (score ? . . .).
2. Changes hands of clock (4 minutes to 3 o'clock). (Must be able to recognize the slight discrepancy).
3. Pictures (interprets) 1 . . . 2 . . . 3 . . . (Two out of three.)
4. Vocabulary index (score ? . . .).
5. Completion test (score ? . . .).
6. Uses code (See Goddard's article, "The Training School," May, 1911).

ADULT.

1. Reversed triangle.
2. President and King. (See Goddard's article cited above.)
3. Ball and field (score 4, with conditions stated).
4. Completion test (score ? . . .).
5. Gives sense of selection (see Binet, 1911).

"One hears very different judgments about the value of life. Some say it is good; others say it is bad. It would be more correct to say that it is mediocre, because, on the one hand, it brings us less happiness than we want, while, on the other hand, the misfortunes it brings are less than others wish for us. It is the mediocrity of life that makes it endurable, or, still more, that keeps it from being positively unjust."

Some of the Binet tests which we found were especially dependent on training, and which came at years already crowded with more suitable tests were omitted. Reading for two memories, 8th year, was omitted because but 65 per cent. passed it with no time limit; furthermore, the difficulty here was not one of memory, but of the mechanics of reading, which is purely a matter of training. Definitions of abstract terms in the eleventh year and differences in meaning of abstract terms in the thirteenth year did not give satisfactory results at any age, and furthermore, the vocabulary test is of much greater value and can replace them. Questions of comprehension in Binet's fifth year we would place before the fourth year. This test probably belongs in the third year.

In general, we have considered that two-thirds of the children of any age ought to pass a test to make the test characteristic of that age. Where the number passing at a given year is just at or below two-thirds and the following year there was a growth in ability, we have ordinarily fixed on the latter year as the more characteristic. Where for two or three years

there is but little change in ability we have generally chosen the earlier of these years as characteristic, even though not quite two-thirds passed, especially when there had been a sharp rise in ability from the year before. In two or three instances we have placed a test where less than 60 per cent. passed; once where it seemed our method of applying the test had been at fault, and in two other cases where our data were rather limited and we believed that the particular group in question had suffered some slight adverse selection.

The above tentative revision has made the lower end of the scale more difficult by setting back many of the tests of Binet's higher years, and the upper range has been supplemented, rearranged, and some of the tests even discarded. As no three-year children were included in our tests, we cannot offer our third and fourth-year tests with assurance that they are all correctly placed; it is possible that three-year-olds would pass some of the fourth-year tests. We are sure that normal four-year-olds can pass them, and that they do not belong to later years, where many of them were placed by Binet.

Believing that tests of memory, vocabulary, observation, reasoning and reaction to a complex social or moral situation bring out fundamental characteristics of mental ability, we have given a well-scattered range of tests on memory, questions of comprehension, reasoning, tests involving observation, linguistic invention and association (such as the completion test and rearranging a sentence of mixed words, vocabulary, etc.), so that a child of any age will be tested on a number of these important questions. Especially satisfactory also are such tests as description of pictures, omissions from pictures, aesthetic judgment, comparison of things from memory and the naming of words. They depend relatively little on formal training, can be arranged to show definite progress and require responses to complex situations that bring out what we really desire to get at. Tests of comprehension, vocabulary, practical judgment and ability to generalize seem especially desirable. The naming of words is interesting in that it brings out the difference between the type of mind that thinks only in terms of concrete, present objects and the mind that can deal with things not present. One S names objects which he sees

lying about the room; another names things from his past experience and according to logical sequence. This test also throws light upon the S's mental self-control. One child plunges in so precipitously that his stock of thoughts is exhausted in a half minute, and he is unable to collect himself again; another starts in calmly, maintains his composure all through the test and makes a very creditable record.

A detailed criticism of the individual tests is reserved until data shall have been secured from the application of the revised scale upon about 1000 non-selected children. This work, made possible by the kind co-operation of several psychologists in various parts of the country, is now in progress.

In this connection it may be said that the tests in general are significant not so much for displaying the child's intelligence in its entirety as for detecting that type of mind which is not capable of profiting from its social environment. A failure on any one of these tests has but little significance, but if a child of 12 or 14, or even younger (as we have found in several cases), is unable to name the months, make change, remember five digits, define better than by use or read a simple selection and give two memories, these facts collectively show that something is lacking in his mental make-up, and often give clues as to the source of trouble. This is illustrated by Fig. II, in which the central line represents age norms on the scale and each of the various crosses to the left the starting point for a serially-graded test. The figure pictures the results secured by applying the test to a 12-year-old who is two years retarded in every respect, the dotted lines extending to the age norm which ought to be reached, the unbroken line to the age norm which is actually attained. The unbroken lines will not usually cross the central line at the same point, but when they show similar divergent tendencies from their dotted companions, the presumption is very strong that we are dealing with a case of greater or less mental deficiency. If thorough investigation of the pupil's home life, health record and educational advantages discloses no sufficient cause for the retardation, the presumption becomes very strong, indeed. There is, however, still the possibility that the condition may be one of temporary retardation rather than of permanent

deficiency. The pupil may be only 10 years old physiologically, though 12 chronologically, and if so, the prognosis is not necessarily unfavorable. In such a case the pupil may finally attain a normal, if late, maturity.

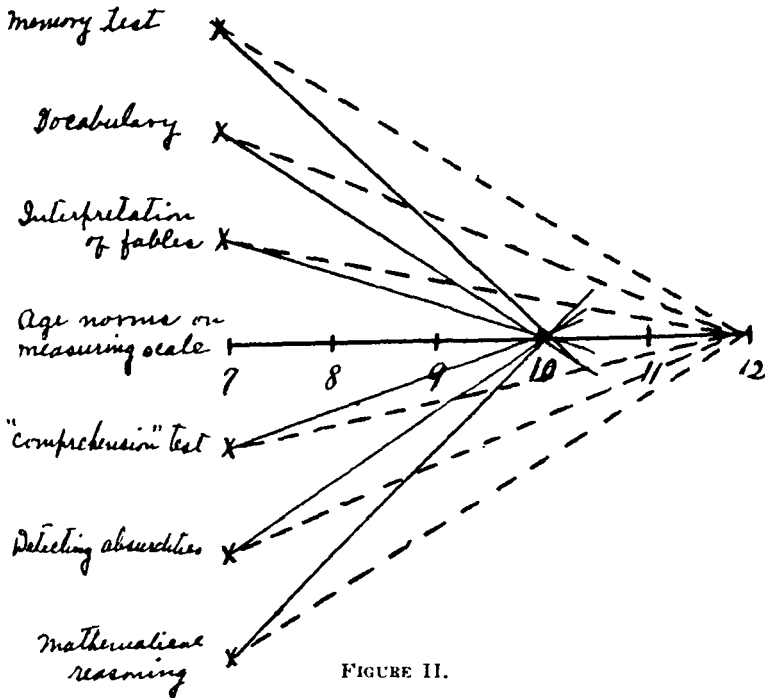


FIGURE II.

In order to make the tests of the greatest practical use, it may prove desirable to apply along with the measuring scale of intelligence the newly-developed tests of physiological age. Unless this is done, it is conceivable that some children will be misjudged as to native ability, a mistake which in individual cases might have very serious consequences. In justice to the child the normality of whose intelligence has been questioned, tests for physiological age should always be applied before institutional treatment is recommended, and perhaps occasionally after commitment.

Another important means of rendering the diagnosis more certain would be to apply tests of ability to learn. If 100 14-year-old boys are running the mile race, their relative posi-

tions at the end of two minutes are not a reliable indication of their individual differences in native ability to run. Some may be crippled, some fatigued from previous exertion, some in training, some raw. If all were put through a standard course of training of sufficient length and under controlled conditions, the differences still remaining would represent, approximately, their difference in endowment in this capacity. We have no standardized norms of ability to learn, except the rough ones offered by a graded school curriculum. Failure to meet the school standard satisfactorily creates a suspicion of sub-normality, and usually not unjustly. But experience proves that this test is not wholly reliable. Some children fail to progress in school for other reasons than real inability to learn. We are in urgent need of supplementary tests of learning ability of such a nature that they will not fail to enlist the child's greatest effort.

ASSIGNING CREDIT.

Binet's original method of assigning credit was faulty in not giving equal value to the different year groups. The method he proposes in the 1911 article avoids this difficulty by equalizing the number of tests for the different years. This arbitrary limitation of the number of tests per year is, however, an unnecessary stricture. We submit another plan, which avoids the difficulty above mentioned; which gives credit for every test passed, discarding no results because they are only a fractional part of the year or half-year unit; which gives partial credit where it seems to be deserved, and which gives a final result more convenient for statistical treatment.

Inasmuch as the number of tests in each year is not the same, no one number has any special value over any other as a unit of value for a year group. Whatever number may be chosen as the year unit, it should be divided by the number of tests in any year to determine the unit value of each test of that year. For sake of convenience, we suggest that 1 be taken as the value of any year group. Accordingly, dividing 1 by the number of tests in any year gives us the unit value of each test in that year. In reckoning an S's total performance add

in, at its unit value, each test passed and add .5 or a half year. (The necessity for adding a half year's credit is due to the method of age grouping which has been employed, and has been explained in full on page 68 of the February number.) In reckoning we count from the first year instead of from the third, disregarding the fact that tests for one and two years have not been agreed upon. To illustrate the procedure, let us suppose that an S has passed tests the sum total of whose value is 8.2. Add the value of half a year and we have 8.7, or a test age for this S of 8.7 years. Half credit may be allowed where the performance seems to make it desirable.

An S should be required to begin as low in the scale as necessary to insure that he is not given credit for what he cannot do, but, of course, should be given full credit below the point where he is started, counting the first and second years. The tests should always continue until there is not even a remote possibility of the S winning further credit. Where these rules regarding desirable range of testing are not observed we can never be sure of the reliability of the norms so established. To make such a thorough-going try-out requires much time and patience. We would repeat that wholesale application of the tests, using but a few minutes of time and a range of but two or three year-groups for each child, can never give us the standard norms we need, and will only serve in the end to bring discredit upon all efforts in this important field.

In spite of the numerous imperfections and inadequacies of the Binet scale above mentioned, our experience in the work of applying it has given us a most favorable opinion of the practical value of such a series when it has been refined and extended, an impression which grows upon us as the work continues. By its use it is possible for the clinical psychologist to submit, after a 40-minute diagnostication, a more reliable estimate of a child's intelligence in relation to normal children of his age than most teachers can offer after a year of daily contact. Since all human estimations are relative to some standard, the teacher has no means of discovering whether her class on the whole is above or below the normal for the corresponding age. Her standard may be too high or too low, vague, mechanical or fragmentary. We venture

the prediction that before many years it will become a matter of course to apply serial mental tests in the public schools to all pupils who are retarded, or about to become retarded, or who give indication of unusual ability. In New Jersey a recent State law makes the application of tests of mentality mandatory in the case of all school children retarded as much as three years.

The scientific management of special classes for atypical children in the public schools will be impossible until similar series of tests are multiplied indefinitely. To ascertain the full extent and qualitative nature of a case of mental retardation, together with its causes, is extremely difficult. In fact, with existent methods it is hardly possible at all. Our treatment of such cases barely transcends guesswork. May we not hope that 10 or 15 test scales will soon be devised, each more perfect than the Binet series, which, collectively, will enable the clinical psychologist in a few days to lay bare the "natural history" of a child's development; tests which will explore every line of efficiency, intellectual, volitional, motor, personal, pedagogical, social, linguistic, etc., and which will relate all these at every point to individual peculiarities of instincts and interests and to all significant incidents of his experience? Now that the individual treatment of pupils in the schools has begun, there is no stopping short of this ideal. Tests must be developed which will enable us to differentiate all degrees of intellectual ability and all kinds of intellectual unevenness; others which will mark for special educational effort the pupil whose emotional equipment and volitional tendencies threaten the onset of criminality, insanity, hysteria, neurasthenia, or other of the neuroses. With the accumulation of positive experimental data on so many aspects of individual mentality, it would seem that John Stuart Mill's suggestion looking toward a science of human character, *ethology*, as he called it, is on the verge of realization. The modern studies in the psychology of testimony, suggestibility, measurements of mental growth, super- and sub-normality, submerged complexes, emotivity, psycho-analysis and association reactions offer in addition many other fascinating fields for the genetic study of personality.

A word may be added respecting criticisms which have been offered to the effect that mental tests, being only measures of a product, can have no value for psychology. This, of course, is the extreme point of view of the uncompromising structural psychologist who limits by definition the legitimate scope of the science to the study of the content of consciousness. We prefer, however, to conceive of psychology in the broadest sense as the science of behavior. In so far as the structural elements shed light upon function they are significant, but we believe they can no more usurp the entire domain of psychology than the chemistry of the bones and muscles can swallow up or displace the science of physiology.

SUGGESTIONS FOR FURTHER DEVELOPMENT.

We have considered our study only a brief preliminary step in surveying the intellectual capacities of normal children. If our tentative revision of Binet's scale has sufficient value, it should be subjected to a thorough trial upon thousands of representative children. At the same time, numerous supplementary tests should be subjected to experiment, for it cannot be supposed that the tests thus far hit upon even approach perfection. There are also wide gaps in the series where tests need to be supplied. It is especially desirable that age norms be secured for tests of the general nature of some of those used by Healy and Fernald in the Chicago Juvenile Court.

If possible, norms for many tests should be established beyond the thirteenth or fourteenth years, though the difficulties here are very serious. At this time sex differences, and probably also physiological age differences in the same sex, become much more prominent. Perhaps from the twelfth or thirteenth year it will be found necessary for practical purposes to divide the scale into two branches, one for girls, the other for boys. Even this bifurcated scale would have little value unless correlated with scales of physiological age. Again, in this range, emotional disturbances come in to invalidate our norms of intellectual performance. The sex interest, emotional control, religious upheavals, vocational

yearnings, all these and many other adolescent phenomena are sure to complicate the activities of the intellectual functions. Still another difficulty in obtaining norms above 13 years is the fact that the public school, the only place where extensive experiments of this nature can be carried on, does not hold all the pupils beyond the fourteenth birthday. Where compulsory school laws actually keep all pupils in school till the fifteenth birthday, norms can be secured for 14 years, provided care is taken to get representative 14-year-olds. This cannot be done if only 14-year-old children who have not progressed to the high school are included.