

The Binet-Simon Measuring Scale for Intelligence: Some Criticisms and Suggestions

In 1908, the French psychologists Binet and Simon published their second and most famous series of tests for the diagnosis of the grade of intelligence of children. This series, translated and adapted for American use by Goddard and Whipple, has won for itself widespread acceptance and increasing application by those interested in the education of exceptional children.

This widespread practical application has demonstrated that in various minor respects the tests, originally developed for use with French children, are ill adapted to the needs of Americans, and, as a result, many workers are at present engaged in trying out minor variations in the hope of improving the series. In other words, there is general agreement that the measuring instrument is based on correct principles, and constructed on the right plan, but that it needs minor adjustments to make it work smoothly under new conditions.

The object of this article is to present considerations which lead the writer to believe that what we must have is a new instrument rather than the readjustment of the old; a new instrument utilizing what is good in the old, but largely planned on different principles and constructed along new lines.

The Binet-Simon tests consist of a series of fifty-six tasks and questions adapted to the capabilities of normal children of from three to thirteen years of age, and their purpose is to provide a measuring scale whereby the intellectual performance of the child tested may be compared with that of the average normal child of the same age.

This arrangement of the tests on an age scale constitutes the feature of greatest value in the Binet-Simon series, and accounts for the enthusiastic reception accorded them immediately upon their publication. It enables the examiner to test

a seven-year-old child, for example, and to discover that his mental development is two years in arrears if he is able to pass successfully only the tests assigned for children five years of age and younger. This method provides for the first time a definite and universally intelligible scale by which to measure the performance of the subject.

A general idea of the character of the tests for each year may be gained from the following table, which only roughly indicates the nature of the tasks and questions:—

Three Years

- 1 Where is your nose, eyes, mouth?
- 2 Repetition of sentences
- 3 Repetition of numbers
- 4 Describing pictures
- 5 Name of family

Four Years

- 6 Sex of child
- 7 Naming familiar objects
- 8 Repetition of figures
- 9 Comparison of lines

Five Years

- 10 Comparison of weights
- 11 Copying square
- 12 Making rectangle with divided card
- 13 Counting four cents

Six Years

- 14 Indicating right hand, left ear
- 15 Repetition of sentence
- 16 Esthetic comparison
- 17 Definition of objects
- 18 Execution of triple order
- 19 Own age
- 20 Knowing morning and afternoon

Seven Years

- 21 Unfinished picture
- 22 Number of fingers
- 23 Writing from copy
- 24 Copying a diamond
- 25 Repetition of figures
- 26 Description of pictures
- 27 Counting thirteen cents
- 28 Naming four common coins

Eight Years

- 29 Reading and report
- 30 Counting money
- 31 Naming four colors
- 32 Counting backwards
- 33 Writing from dictation
- 34 Comparing objects from memory

Nine Years

- 35 Knowing the date, day, month, day of month, and year
- 36 Reciting days of week
- 37 Making change
- 38 Definition of familiar objects
- 39 Reading and report
- 40 Arrangement of weights

Ten Years

- 41 Reciting months of year
- 42 Naming nine pieces of money
- 43 Sentence building
- 44 Problem questions

Eleven Years

- 45 Detecting absurd statements
- 46 Sentence building
- 47 Naming sixty words in three minutes
- 48 Defining abstract terms
- 49 Sentence building

Twelve Years

- 50 Repetition of figures
- 51 Rhymes
- 52 Repetition of sentence
- 53 Problem questions

Thirteen Years

- 54 Drawing from design cut in paper
- 55 Describing figure made from reversed triangle
- 56 Differences between abstract terms.

These tests are designed to measure native ability, not scholastic attainment. They aim to provide the investigator with an instrument which will enable him to form a trustworthy estimate of the child's capacity for adapting himself to his social environment, and so are designed with special reference to evaluating his judgment, good sense, initiative, and adaptability.

Their value as a measure of this kind of intellectual capacity depends on whether or not they really test the qualities they aim to test and with what degree of accuracy. It is the opinion of the writer that they may be greatly improved in both respects. His criticisms fall under five general heads:—

- I. The tests predominantly reflect the child's ability to use words fluently, and only in small measure his ability to do acts.
- II. Five of them depend on the child's recent environmental experience.
- III. Seven depend on his ability to read or write.
- IV. Too great weight is given to tests of ability to repeat words and numbers.
- V. Too great weight is given to "puzzle tests."
- VI. Unreasonable emphasis is given to tests of ability to define abstract terms.

I. TALKING VERSUS DOING

The first and most serious criticism is that the ability predominantly measured by these tests is the child's ability to use words fluently, and that this gives a warped and partial measure of his real degree of intelligence. The keynote of this criticism is found in the tests themselves, among the questions assigned for ten-year-old children, where the child is asked, "Why should you judge a person by what he does rather than by what he says?"

A correct answer to this question would be that we should judge a person by his acts rather than by his words, because his acts are accurate indicators of what he really is, whereas his words may have only the slightest relation to his real self. This principle is reflected in the proverbs and literature of every age, and of all peoples. It is so axiomatic that Binet and Simon have rightly assumed that it forms a part of the knowledge of every normal ten-year-old child. Nevertheless, by careful count, two-thirds of their tests are tests of the child's ability to use words, and only one-third indicate his ability to act. The

assumption seems to be that native ability to do can be tested by testing the ability to use words about doing.

The fallacy of this assumption is acutely appreciated by the school superintendent who is forced to select principals and teachers on the basis of examinations, and then sadly observes the striking contrast between the way in which the candidates describe how they would teach and administer, and the way in which they actually succeed in teaching and administering. Federal, state and municipal officials, whose field and office forces are selected on the basis of examinations, are equally alive to similar common discrepancies between words and deeds.

The root of the fallacy is the fundamental fact that the motivating stimuli which shape one's actions in coping with a real problem in life are invariably multiple and complex, whereas those which determine his answer to a hypothetical question are simple, few, and different in quality.

An illustration of this may be secured by putting to a number of intelligent adults of demonstrated practical ability the questions assigned to ten-year-old children in the tests under consideration. One of these reads, "What ought one to do before taking part in an important affair?" The writer's experience in putting this question to business men is not encouraging. A few answers have been received, ranging from "Take a bath", and, "Put on your best clothes", to "Take some money from the bank" and "Transfer your property to your wife"; but in general those questioned reply with energetic expressions of short and ugly words and emphatic protestations that the question is unanswerable.

Again, these problem-questions overlook the importance of habit and of the emotions in influencing action. The ten-year-old child or the adult of indifferent mental ability may have a ready answer to the questions "What is the thing to do if you find out that your house is on fire?" "What ought one to do when he has been struck by a playmate who did not do it on purpose?" and, "What would you do if you were punished when you did not deserve it?" But the child or adult who does just the right thing when he has been struck by another, discovers that his house is on fire, or suffers undeserved punishment, thereby demonstrates a quality and degree of native ability to which few indeed among us may hope to attain.

A still further objection is that the tests assume an agreement between verbal equality and real equality which seldom exists. The blood-curdling series of tests put to eleven-year-old children to discover their ability to detect absurdities well illustrates this defect. The statement

“There was found in the park today the body of an unfortunate young girl, frightfully mutilated, and chopped into eighteen pieces. It is thought that she committed suicide.”

may well be a pleasant and entertaining narrative to a normal and somewhat phlegmatic child, but constitutes a serious nervous shock to his more sensitive companion. In the same statement read to both children there is verbal equality; in their psychical import there may be the most serious inequality.

II. RECENT ENVIRONMENTAL EXPERIENCE

Five tests depend in high degree on the child's recent environmental experience. These are the tests relating to time and to money. Some of them are “doing” tests, and some of them “saying” ones. The assumption with regard to the “time” questions is that intelligent children, irrespective of school training, should be able to name the day of the week, the month, the day of the month, the year, etc. Experiment among business and professional men shows that they are frequently unable to supply these data off-hand unless the nature of their business requires constant reference to them. Probably every reader will recall that it requires only three or four days of a camping trip or an ocean voyage to lose track of the days of the week and the days of the month, and that a distinct shock is experienced when some one mentions the fact that “today is Sunday.”

The writer recalls serving as a member of a Federal jury in the West Indies trying smuggling cases, in which the masters of ocean-going trading sloops were the accused. In these cases it was proved beyond any question that these sloop captains were not only illiterate, but that they were absolutely ignorant of the names of the months and did not keep track of the days of the week, with the exception of the Sundays. Nevertheless, these men were distinctly able and intelligent, spoke several languages, navigated dangerous waters, and carried cargoes of considerable value. In the writer's opinion, the ability to name

off-hand the day of the week and of the month is governed almost entirely by daily work and very little by native ability.

A similar objection, but one probably less serious, arises in connection with those questions having to do with money and making change. These again are abilities largely governed by environment. The ability of a child of ten years to recognize and name at sight a quarter, fifty-cent piece, a five dollar bill and a ten-dollar bill, depends not on native ability but rather on whether or not he is accustomed to see, have, handle or spend these pieces of money.

III. ABILITY TO READ AND WRITE

Seven tests depend on the subject's ability to read and write, which commonly depends on the amount and kind of school experience he has had, and may be only slightly related to his native ability.

IV. REPETITION OF WORDS AND NUMBERS

The repetition of words and numbers has an even more remote relation to the ability to cope with the problems of life, and yet one-seventh of the tests are of this sort. The simpler of them can be successfully passed by a gifted parrot; the more difficult ones recently proved beyond the ability of a university professor tested by the writer.

V. PUZZLE TESTS

Several of the tests seem best designated as "puzzle tests" and appear to have strikingly little relation to anything the normal person has to do in the ordinary day's work. Such a one is the demand that the eight-year-old child count backwards from twenty. Counting backwards is one of the rarest things most people are called upon to do and yet the proposal has recently been seriously made that these present tests be "improved" by requiring the subject to recite the names of the months from December back to January instead of forward from January to December. To teach children to recite backwards lists of words that have a normal fixed order is educationally vicious. To include such a requirement in tests of intellectual ability is at least questionable.

Another "puzzle test" for thirteen-year-old children seems

as foreign to every day experience as the foregoing. It uses for material a visiting card cut along the diagonal and asks the child to describe the resulting shape if one of the triangles were turned about and placed so that its short leg was on the other hypotenuse and its right angle at the smaller of the two acute angles. So far the writer has failed to find any one able to describe the resulting shape.

VI. ABSTRACT TERMS

Definitions of abstract terms and expressions of the difference between abstract words of similar sound but different meaning constitute the last class of tests to be here considered. The first objection to these is that philosophers are almost the only people who think in abstractions, and the second objection is that words of peculiarly difficult character have been chosen. Let the reader himself try the eleven-year-old test which demands definitions of *Charity*, *Justice*, and *Goodness*. To pass he must give two good definitions. Then let him try the thirteen-year-old tests and tell the difference between

Pleasure and Honor
 Evolution and Revolution
 Event and Advent
 Poverty and Misery
 Pride and Pretension.

The third of these pairs is a good one to try on your friends. If not satisfied with their explanations recourse may be had to the Standard Dictionary where one will be rewarded by finding that an *advent* is the coming of an *event*, but just what the difference between them is remains undiscovered.

To sum up the case to this point: Two-thirds of the Binet-Simon tests are tests of the child's ability to use words, and only one-third tests of his ability to do acts. Among the reasons why certain of the tests fall short of providing satisfactory criteria for the judging of native ability are the following:

1. They overlook the fundamental difference between the multiple and complex stimuli which contribute to the motivating impulse in coping with real problems and the few and simple ones entering as factors in answering questions or obeying commands.
2. The importance of the emotions and habit in influencing action is disregarded.
3. Real equality is attributed to verbal equality.

4. Ability to answer many of the questions depends on the child's daily environmental experiences which differ radically among different children.

5. Ability to meet the requirements of several of the tests depends directly on the excellence of the child's schooling.

6. Several tests depend on the mere ability to repeat words and numbers.

7. Counting backwards and solving puzzles constitute several tests.

8. Several tests turn on the ability to express in words comprehension of difficult, abstract terms.

There are two important sets of evidence in favor of the tests, and they are both good in the sense that they constitute "pragmatic" arguments showing that the tests "work" successfully when applied. In the first place these tests have won rapid and widespread use and endorsement among hundreds of practical teachers and workers with children, whereas all previous tests of intelligence have been practically restricted in use to workers in psychological laboratories.

In the writer's opinion the reason for this has been pointed out by Terman of Leland Stanford, Jr., University, who calls attention to the fact that here for the first time we have a set of tests arranged with reference to steps on a scale that is constant and universally understood. Everyone has a fairly accurate concept of what is meant when one says that a given child shows intelligence equal to that of a ten-year-old normal child. We have had graded tests before but no one knew what the steps on the scale meant in terms of anything else, or where the lower end began or how far the upper end reached. This application of the tests to a definite, universally understood scale constitutes the great contribution of Binet and Simon and it is so important a contribution that its excellence outweighs the shortcomings of the tests themselves.

The second set of evidence consists of the records of applying the tests to large numbers of normal school children with the result that the distribution of the children into retarded, normal, and advanced groups corresponds fairly well with what is termed in statistics the normal frequency distribution. Such studies have been made by Binet and Simon in France with 203 normal children and by Goddard in this country with 1547 children. In both cases the results showed about such a distribution of re-

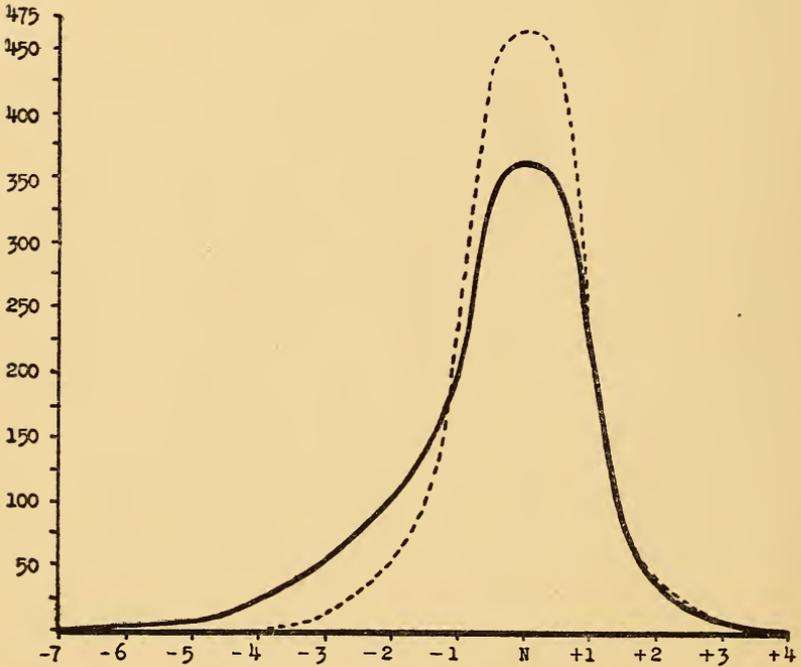
tarded, normal, and advanced children as the theory of normal frequency distribution tells us that we should find, and in both cases the results have been widely cited as constituting a scientific demonstration of the correctness of the tests in so far as their degree of difficulty is concerned.

Unfortunately this conclusion is hardly justified by the results of the investigations as made public, for the reason that we have only the mass figures for the entire group tested and not the figures showing the results for children of each age. This process hides the details from view and if, as many workers report, the tests for the youngest children are too easy and those for the oldest ones too hard, these important facts are concealed by putting all the results for all the ages together.

How this works is illustrated by comparing the results obtained by Goddard in his application of the Binet-Simon tests to 1547 normal school children with data recently gathered by the writer showing the progress of children in the public elementary schools of twenty-eight cities. In the accompanying diagram the solid line represents the distribution of the children tested by Goddard according to the tests showing them to be normal, one year behind, one year ahead, two years behind, two years ahead, etc. The dotted line is based on data showing how long it has taken 14,762 children in twenty-eight American cities to complete the work of seven grades. Those who have done so in seven years are rated as normal; those who have taken six years, as one year ahead in progress; those taking eight years, as one year behind, etc. To secure a proper basis of comparison both the Goddard data and the writer's have been reduced to relative figures and are presented on the basis of 1000 cases.

The significant feature of the diagram is that the curves are closely similar. If the solid curve constitutes a scientific demonstration of the correctness of the Binet tests, then it may fairly be argued that the dotted one constitutes a scientific demonstration that the public school systems and courses of study of these cities are correctly adjusted to the abilities of their pupils; neither too hard nor too easy, but almost exactly right. If this were true there would be far less need for securing a measuring scale of intelligence than there undoubtedly is, for in our public school system we should have just such a scale, scientifically correct and already at hand.

The fact is, however, that the progress figures from the twenty-eight cities referred to show great variations between different grades and localities, and it is only by combining the figures for all the cities that the nearly normal curve shown is secured. A similar comparison could easily be made with the figures showing the results of the tests made by Binet and Simon in France. Indeed, it so happens that this curve almost exactly coincides with that showing the progress of the children in the eight grades of Bayonne, New Jersey, and here again the almost normal curve disappears when the data are presented by separate grades.



Distribution curves showing variations from normal of 1547 children tested by the Binet-Simon scale (solid line) and 14,762 children in 28 cities rated by their progress through seven grades (dotted line). Curves based on relative figures showing distribution of 1000 cases of each kind.

In presenting the foregoing considerations the writer does not wish to appear as an antagonist of the Binet-Simon Measuring Scale for Intelligence, for he is not. He does wish to sound a note of warning against accepting it in its present form as final and satisfactory. What is here set down is the result of

his own attempts to discover ways in which they may be improved, together with ideas secured through lengthy discussions of their application with Mrs. Louise Stevens Bryant of the Psychological Clinic of the University of Pennsylvania.

Binet and Simon have done a great and lasting service for the cause of childhood in basing their tests on a definite scale. The present situation offers a splendid opportunity to psychologists, teachers, and mothers to observe, discover, and record things which normal children do and know at each age. Work of this sort conducted by a large number of observers and co-ordinated by some central agency or agencies would soon give us a series of tests retaining all the good of the present series and replacing present tests wherever experiment and observation show better ones can be found.

Above all let us steadfastly bear in mind that all measuring instruments must be judged for two qualities; first, what they measure, and, secondly, how accurately they measure it. The fact that the Binet-Simon tests are more or less accurately adjusted to the normal capabilities of the children of each age is only one, and the less important, criterion. The problem of paramount importance is whether or not they really measure native ability, and if they fall short, how we may develop a series of tests that will measure it.

Some Publications of the Department of Child Hygiene, Russell Sage Foundation

Medical Inspection of Schools. Luther H. Gulick, M.D., and Leonard P. Ayres, Ph.D. 276 pp. Price, postpaid, \$1.00. Charities Publication Committee, 105 East 22nd Street, New York City.

Gives in convenient form information about the historical, educational, administrative, and legal phases of the work. The most extensive bibliography on the subject yet compiled is added. Third edition.

Laggards in our Schools. Leonard P. Ayres, Ph.D. 236 pp. Price, postpaid, \$1.50. Charities Publication Committee, 105 East 22nd Street, New York City.

A study of the over-age child, the child who repeats grades, and the falling out of school of pupils before the completion of the course in American city school systems. It is illustrated with some forty charts and diagrams, and its inter-city comparisons are set forth in nearly one hundred tables. Third edition.

Wider Use of the School Plant. Clarence Arthur Perry. Introduction by Luther H. Gulick, M.D. 423 pp. Price, postpaid, \$1.25. Charities Publication Committee, 105 East 22nd Street, New York City.

All the activities, from evening schools to social centers, now carried on after school hours in school buildings are described in this book. It shows them in actual operation, outlines the various forms of administration and gives pertinent details as to cost, development, and the social amelioration which they are effecting. Fully illustrated and helpfully indexed. Second edition.

Seven Great Foundations. Leonard P. Ayres, Ph.D. 79 pp. Price, postpaid, 35 cents. Department of Child Hygiene, Russell Sage Foundation, 400 Metropolitan Tower, New York City.

Information concerning the origin, purposes, activities and history of the following foundations: Peabody Education Fund, John F. Slater Fund for the Education of Freedman, Carnegie Institution of Washington, The General Education Board, Carnegie Foundation for the Advancement of Teaching, Russell Sage Foundation, and the Anna T. Jeanes Foundation.

Open Air Schools. Leonard P. Ayres, Ph.D. 171 pp. Price \$1.20 (postage 12 cents). Doubleday, Page & Co., New York.

This volume gives the important and significant American and foreign material with respect to outdoor schools. It describes the English, German and American types, gives the results and furnishes definite information with respect to clothing, food, cost, administration, etc. It has more than 70 pages of illustrations and diagrams. Bibliography.

