

1918

## Further Extension and Revision of the Binet-Simon Scale

F. Kuhlmann

Follow this and additional works at: <https://scholarlycommons.law.northwestern.edu/jclc>

 Part of the [Criminal Law Commons](#), [Criminology Commons](#), and the [Criminology and Criminal Justice Commons](#)

---

### Recommended Citation

F. Kuhlmann, Further Extension and Revision of the Binet-Simon Scale, 8 J. Am. Inst. Crim. L. & Criminology 890 (May 1917 to March 1918)

This Article is brought to you for free and open access by Northwestern University School of Law Scholarly Commons. It has been accepted for inclusion in Journal of Criminal Law and Criminology by an authorized editor of Northwestern University School of Law Scholarly Commons.

# A FURTHER EXTENSION AND REVISION OF THE BINET-SIMON SCALE<sup>1</sup>

By F. KUHLMANN<sup>2</sup>

The extension and revision of the Binet-Simon scale of mental tests with which this paper will deal is the result of seven years of continuous work in the examination of over 3,000 feeble-minded of all ages and grades by the writer, and about 2,000 normal children from birth to eighteen years of age by the writer and assistants. During this time a careful study has been made from every point of view of each individual test, the combination of tests into age-groups, and into an age-scale, and methods of scoring and classification. Over a hundred new tests have been tried out and incorporated, modified, or discarded. The net result up to date has been a greatly extended scale of tests, so as to include all grades and stages of mental development from three months to mental maturity, an increase in the total number of tests from 56 in the original 1908 scale of the authors to 129 in the present scale (counting each test as many times as there are ages for which it has been standardized), a thorough standardization of all the tests, and a working out of the general principles that underlie mental tests and the age-scale. I shall limit myself here to a brief statement of the general principles that were followed in producing the present scale of tests, with a few indications of the main results, and the list of tests incorporated in the scale as it now stands

## REQUIREMENTS OF THE INDIVIDUAL TEST

Certain requirements were made of the individual test. In eliminating tests from the old scale, and in devising new ones, three aims were kept in mind. The first was to eliminate the personal factor of the examiner in the use of the test and in the interpretation of responses to be obtained with it. The second was to secure as great a discriminative capacity for each test as was possible. By this latter is meant tests that would show as large an increase as possible from one age to the next in the percentage of children passing it. The third was to make each test as far as possible independent of training that

<sup>1</sup>Presented at meeting of American Association of Clinical Criminologists, New Orleans, 1917.

<sup>2</sup>Director of Research, Minnesota School for Feeble-Minded.

any individual child might have had. To meet the first requirement the test itself must be of such a nature that the directions as to what the child is to do can be stated in such simple and brief form that it will not require variations or supplementing with different individuals. The test must further be such that the response to it is not capable of more than one interpretation by the examiner. The possible responses must all be readily classifiable as passed or failed. In the present scale the procedure for the use of each test is worked out with as much care as is the test itself. For the upper part of the scale all the responses are scored in terms of time taken to do the test and number of errors made. This eliminates completely the personal factor in scoring responses. The introduction of new tests with time and error scoring only has been carried out as far down into the scale as proved feasible. The importance of this objectivity of the test results needs no special discussion. Many otherwise good tests have been suggested and used by examiners which do not meet this requirement. They yield reliable results in the hands of the proper examiner, but are of no great value to others, just as there are examiners of long experience and training who make reliable diagnoses without the use of standardized tests or other scientific methods that they could give into the hands of another not so experienced.

In regard to the second aim, it is obvious that the test is the better the greater the increase in the percentage of children from one age to the next who pass it. What percentage increase can be accepted as satisfactory is determined not by the individual test directly, but by the combined result of all the tests used. The combined result of a few tests, each of which has a large discriminative capacity, may be more reliable than the combined result of a much larger number of tests, each of which has a smaller discriminative capacity. The reliability of the combined result decides whether the discriminative capacity of the individual test is satisfactory. But inasmuch as the rate of mental development normally decreases with age there is on the whole a bigger difference between the percentage of young children of a given age and the next age that will pass a test than is the case with older children. Consequently, the increase in percentage passing a test from one age to the next that may be accepted as a satisfactory increase itself decreases with the age for which the test is designed. In the present scale this difference is as high as 30% and more for young children, and as low as 5% for tests at the upper end of the scale.

In judging whether or not the individual test result is not unduly affected by special training or the lack of it, the percentage passing from one age to the next is not a sufficient guide. It is entirely possible for a test to be statistically very satisfactory in this respect, and yet be a test more of training than of mental development. This is especially true when the statistics cover a very large number of children from different localities. Some very young children are taught to do certain things, but this number is small. As older children are considered, the number that have been taught a certain task naturally increases. It therefore becomes necessary to take into account the nature of the test from this standpoint. The real task involved in the test should be as little related as possible to every day activities or things likely to be taught, at least for tests at the upper end of the scale. Some of the old tests retained in the scale may still be objectionable on these grounds, but were kept because they were otherwise very satisfactory. The new tests introduced, however, were all selected after a careful consideration of this question.

#### REQUIREMENTS OF THE AGE SCALE

By the age scale is meant a scale that measures in terms of number of years of normal development, and thus scores in terms of mental ages, as does the Binet-Simon system. This method is adhered to in the present scale because it has a number of decided advantages over all other schemes so far presented. The principles underlying the age scale, the requirements that must be met in constructing it, in a word, the theory of the scale, were not worked out by Binet and Simon. These are the contribution chiefly of later workers, and we have still much to learn in regard to them. I shall mention a few general requirements only, on the basis of which the present revision and extension was worked out.

(a) The first is the agreement of age and average mental age as determined by the tests. By this is meant that the average mental age of a large group of normal six-year-olds, for example, must be exactly six years, or at least exactly enough for all practical purposes, and the same for each other year. The original Binet-Simon scale fulfilled this condition fairly well at most points within the age-limits of its real applicability, that is, from the age of about four or five to about ten. It measured nearly a half year too high at its lower end and too low at its upper end. It has proven a relatively easy task to correct the scale in this respect so as to make it accurate in an entirely

satisfactory degree in these average results. But it is not a matter of chief importance.

(b) The chief requirement is reliability of the individual mental age that the scale gives in the examination of the individual case. Average accuracy gives us no clue as to the degree of reliability of the individual mental age, and yet it is in the individual rather than in the group that the interest nearly always lies. The position taken here is that there can be no statistical or mathematical proof of the degree of this reliability. These statistics can give only general indications, but not proofs, because a necessity always involved is that we know beforehand just what the exact mental ages are of the children that we test in order to determine the reliability of the tests. There are two conditions under each of which reliability of the individual mental age increases. These are, first, increase in the discriminative capacity of each individual test, and, second, increase in the number of tests throughout the scale. The efforts in this revision have been directed towards improving these two conditions, and no attempt has been made to find a statistical method whereby to prove the degree of reliability of the individual mental age. No test was included in a given age-group of the scale simply because a satisfactory percentage of normal children of a corresponding chronological age passed it, but only when there was a satisfactory increase in the percentage of children from one age to the next that passed it. The number of tests for each age-group has been increased to eight for age-group III and beyond.

(c) The next consideration concerning the age-scale is the grouping of the tests into age-groups. Two matters only need to be mentioned. The first is in regard to the importance of having each test in its proper age-group, and the second concerns the percentage of normal children of a given age that should pass a test in order to have that test correctly placed in the corresponding age-group. The work on this revision has led me to conclusions quite different from current opinions on both these questions. It has been generally held that the correct placing of the tests in their respective age-groups was of the first importance. I maintain that, theoretically, the grouping of tests into age-groups at all is not a necessity, but largely only a convenience. The convenience, however, is so great that for practical purposes it becomes a necessity, but the fact remains even then that the misplacement of a test by as much as a year or even more is not a serious matter. A given mental age as determined by the age-scale simply means that the case in question can do a certain total number of tests out of

all there are in the system, this number being the same as the average normal child of corresponding chronological age can do. It makes no assumption that the child in question can or cannot do the tests of any particular age-group. The result would be exactly the same if all the tests in the whole scale were mixed up and treated as one group only. But in making an examination it would then become necessary to give all of the 129 tests in the present scale, instead of only four or five age-groups, or 32 to 40 of the 129. With the age-grouping we need only to give enough tests to make sure that none below those given would be failed in, and that none above those given would be passed.

In regard to the percentage of children that should pass a test in order to have that test correctly placed in its age-group my procedure has been first to secure such a grouping empirically as would roughly make the average mental age correspond to the chronological age. All efforts to deduce this percentage theoretically have so far failed. When the scale, however, gives correct average mental ages the question as to the percentage that should pass a test is thereby also empirically decided. The percentage that is found to pass a test in a given age-group when the scale gives a correct corresponding average mental age is the correct percentage for that point of the scale. The revision has gone through a great number of preliminary groupings and regroupings. Old tests were shifted up or down in the scale, and new tests were introduced into an age-group on the basis of what appeared to be the correct percentage that should pass them. These percentages accepted provisionally were revised and corrected again on the basis of the mental ages the provisionally revised scale gave, and so on, until a final adjustment was reached. The scale as it now stands gives only a negligible error in the average mental ages for the different ages. But with this adjustment it was found that the percentage of children of different ages that passed the tests of corresponding age-groups varied from nearly 100% for the very young children to only a little over 50% for the older children approaching mental maturity.

(d) In the question of the number of tests required in successive age-groups there seems to be agreement. This number should be the same for successive age-groups, and not irregular. The revised scale has five tests for each age-group below age-group III, and eight tests for each age-group above age-group II. This number is all that can be used in an examination of a single sitting without extending the latter beyond practical limits.

## METHODS OF SCORING AND CLASSIFYING

The original system of tests sometimes allowed one trial and sometimes several for a given test. In some cases a pass for one trial passed the whole test, in other cases two or more trials had to be passed in order to pass the test. This is a correct principle at least for younger children if based on the right procedure. Assuming the several trials of a test to be all of a similar nature, success in one or two trials proves the child's capacity to perform the kind of task involved. Failure in one or two of the trials means only some disturbing factor, usually poor attention or effort. It would be wrong procedure to construct a scale of tests so that the result for each test would always be affected by this one and the same factor. Poor attention and a varying degree of effort is a prominent trait of young children and should be measured by the tests, but not by every one of them.

In tests in which a time and error score are made the length of time taken for the task set and the number of errors made are not always of equal value in measuring mental development. On the whole, time is much more a measure of effort than is error, and a time score should therefore be used less with younger children than with older, or should not count for as much as the error score, when it is used. While this is a general rule, there are tests of such a nature that an increased effort to work fast readily results in an increase in number of errors that is out of proportion to the gain in time. The rule to be followed in determining the method of scoring the result of the individual test is the same in all cases. That method should be adopted which gives the test the greatest discriminative capacity, if the test by itself is to be made as effective as possible. Sometimes this greatest discriminative capacity is secured by allowing only one failure out of several trials for a test, sometimes two or more failures allowed attains this end. In the case of time and error scoring some simple formula that combines the time and error scores into one score is required. It then becomes easy to so adjust this formula as to allow the proper relative amount for time and error to make the test the best possible. A guide in adjusting this formula is the average time scores and the average error scores for successive ages. If the average time score decreases twice as fast with increase in age as does the average error score, then the time score should on the whole count for twice as much in the formula as the error score. Again, since the average time score in terms of number of seconds is generally many times greater than the average number of errors made in a test,

the formula must reduce the time score or increase the error score or do both if the two are to count for the same in the result. The formula for combining the two scores into one, and in such a way as to give the test its greatest possible discriminative capacity, therefore, takes on various forms. Thus in test VII 6 the formula is  $T + E$ ; in VIII 8 it is  $T/10 + 5E$ ; in XI 6 it is  $3T + 4E$ , etc. With the formula thus determined, the test is placed in that age-group for which the proper percentage of children of corresponding age pass it, when allowing a given score to be attained in order to pass it. Binet and Simon do not discuss in each case why they allow one or two or more failures out of several trials in a test, but it is left to be inferred that it is to make the test of the proper degree of difficulty for the age-group in which they place the test, or in other words, again, to make the proper percentage of children of corresponding age pass it. The present revision makes this a secondary requirement, but adopts the method of scoring the individual test so as to meet the chief requirement of maximum discriminative capacity of the test.

The age grouping of the tests implies that the results are to be scored in terms of mental ages. It is also strongly advised that the grading of different degrees of mental development be made in terms of the intelligence quotient. The advantages of this procedure have been so fully discussed by different writers as to need no further consideration here. One important point, however, remains. This is that a different method of scoring is required for the very upper end of the scale if it is not to measure short at this end. This fault is inherent in the method, for no matter how high the scale is extended, no subject could ever attain the mental age of the highest age-group in the scale. The highest age-group in the scale as it now stands is XV. For the exact determination of any grade of mental development that is at all below normal a change in method of scoring will hardly be required. But the grading of mental development above the mental ages of twelve sometimes has its application, and for this a different procedure is required. The method used here is chosen largely because of its simplicity and ease of adaptation. All the tests in the higher age-groups are scored in terms of time and error. For cases whose mental ages are above eleven years these last and highest ten or twelve tests are used as an independent group, and the average score for the group is determined in examining an individual. The scoring formula for each test is corrected so that each test will count approximately for the same as every other in the group in getting this average. With

this method the maximum mental age any individual can attain is the highest average mental age of normal adults, and is represented by the highest average score for this group of tests that normals of any age from twelve years up attain. This is the limit of any age-scale. If we cannot with mental tests find any further mental development beyond the age of fifteen, for example—that is, if the average score for fifteen-year-olds is just as high as it is for any older normals—then fifteen is the highest possible mental age attainable on a year scale. Theoretically, 50% of fifteen-year-old normals should have a mental development that exceeds the average for fifteen, but this excess cannot be expressed in terms of mental age. Therefore, in the first place, in computing the intelligence quotient of an individual over fifteen years chronologically mental age should never be divided by more than fifteen. In the second place, in order to get the grade of mental development of a person whose score exceeds that of the average normal fifteen-year-old the age scale is in part abandoned. The formula is as follows:

$$\frac{1}{\text{The score he attains with the group of tests}} \div \frac{1}{\text{The average score for his age}}$$

If he is over fifteen chronologically the average score for fifteen is used as the divisor in the second half of the formula. This quotient is in effect the same as the intelligence quotient obtained by dividing mental age by age, and with this procedure the scale of tests becomes unlimited in the height of the score a person who is above the average adult may attain.

#### SELECTION OF CHILDREN FOR NORMS

Theoretically, the age-scale proceeds on the assumption that the different mental ages in terms of which it expresses its results represent the average capacities of non-selected children of corresponding chronological ages. This assumption can never be more than approximately correct. Non-selected groups of children are never available for testing purposes, and how correct this assumption is then depends on the manner of selection of the children used to establish norms, and not on the degree of non-selection in groups as found. This revision and extension of the scale has made use of the following groups of children and adults: 1. Children examined in various baby contests. 2. Children in a state orphan asylum. 3. Kindergarten children in the public schools. 4. Public school children from the first grade to seniors of the high school. 5. Adult employes of an institution for the feeble-

minded. 6. All grades and ages of feeble-minded in the school for feeble-minded, and special classes in the public schools. None of these groups can be said to be entirely non-selected. Space here will not permit discussing the various selective factors that enter with these different groups. My general experience has been that beyond the age of ten years among public school children the effect of selective factors present has become so great as to necessitate a careful further selection on the part of the experimenter in order to offset these factors and secure approximately average capacities in the children for each age. On the whole, pedagogical advancement is the best single basis for selection. For the highest ages, I have selected the pedagogically normal, calling a child pedagogically normal if he was ten years old near the end of his fourth grade in school, and so on. The one unavoidable difficulty that arises in defining the pedagogically normal lies in the fact that the time of the school year must be taken into account. If the ten-year-old is pedagogically normal at the end of his fourth grade in school, he is not so at the beginning or middle of his fourth grade in school, and in practice the work of testing to get norms cannot all be done at any one time of the year, but must extend through the whole year for several years.

Aside from selective factors present that eliminate school children that are below average in mental development from the schools in different degrees at different ages, there are various social factors that have a selective influence in determining the average capacities of the adults, and hence indirectly of the children in different sections of the country and communities. Prominent among these are race and nationality differences, and differences in predominant occupations calling on the whole for certain grades of capacities. The results in using the original tests in different countries give us some measure of assurance that these differences are not large enough to be fatal to the usefulness of one and the same age-scale for all children. Roughly at least, the scale has given very similar results with children of different nationalities. The question as to whether we should attempt to construct a scale for each race is an open one. From a practical standpoint this is not necessary or even desirable. From this standpoint we care less to know whether a given child is equal to the average of his race than we care to know whether he is equal to the average of the nation's children. For it is the nation as a whole that sets the requirements for its citizens, and not the descendants of any particular race in that nation.

## THE LIST OF TESTS IN THE PRESENT SCALE

The following is the list of tests that is used in the present system. They include 37 tests of the 1908 scale, exclude 19 of that scale and add 65 new ones with 92 age-group scorings, or an equivalent of 92 new tests. The latter are entirely new or borrowed from the literature in slightly or very much modified form. In this list they are all marked "B.-S." or "New." If a B.-S. test has been shifted the Roman numeral in parenthesis indicates the age-group from which it was shifted.

*Age Three Months. (All new.)*

1. Carrying an object to mouth. (New.)
2. Reactions to sudden sounds. (New.)
3. Binocular co-ordination. (New.)
4. Turning eyes to object in marginal field of vision. (New.)
5. Winking at an object threatening the eyes. (New.)

*Age Six Months. (All new.)*

1. Balancing head and sitting. (New.)
2. Turning head towards source of a sound. (New.)
3. Opposing thumb in grasping. (New.)
4. Prolonged holding of object placed in hand. (New.)
5. Reaching for seen objects. (New.)

*Age One Year. (All new.)*

1. Sitting and standing. (New.)
2. Speech. (New.)
3. Imitation of movements. (New.)
4. Marking with pencil. (New.)
5. Recognition of objects. (New.)

*Age Eighteen Months. (All new.)*

1. Drinking. (New.)
2. Feeding with spoon or fork. (New.)
3. Speech. (New.)
4. Spitting out solids. (New.)
5. Recognition of objects in picture. (New.)

*Age Two Years. (4 new.)*

1. Pointing out objects in pictures. (New.)
2. Imitations of simple movements. (B.-S.)
3. Obeying simple commands. (New.)
4. Copying a circle. (New.)
5. Removal of wrapping from food before eating. (B.-S.)

*Age Three Years. (2 new.)*

1. Enumeration of objects in a picture. (B.-S.)
2. Pointing out parts of body. (B.-S.)
3. Giving the family name. (B.-S.)
4. Repeating a sentence of six syllables. (B.-S.)
5. Naming of familiar objects. (B.-S. IV.)

6. Repeating two numerals. (B.-S.)
7. Naming pictures from memory. (New.)
8. Tracing a square. (New.)

*Age Four Years. (5 new.)*

1. Giving sex. (B.-S.)
2. Repetition of three numerals. (B.-S.)
3. Comparison of two lines. (B.-S.)
4. Discrimination of forms. (New.)
5. Tracing irregular form. (New.)
6. Recognition of forms. (New.)
7. Comprehension. (New.)
8. Naming pictures from memory. (New.)

*Age Five Years. (2 new.)*

1. Counting four pennies. (B.-S.)
2. Copying a square. (B.-S.)
3. Comparison of weights. (B.-S.)
4. Making rectangle with two triangles. (B.-S.)
5. Repetition of sentences of ten words. (B.-S. 1911.)
6. Definition according to use of object. (B.-S. VI.)
7. Tapping blocks in irregular order. (New.)
8. Naming the primary colors. (New.)

*Age Six Years. (3 new.)*

1. Distinction between right and left. (B.-S.)
2. Aesthetic comparison. (B.-S.)
3. Distinction between morning and afternoon. ((B.-S.)
4. Recognition of mutilation in pictures. (B.-S. VII.)
5. Execution of three simultaneous commands. (B.-S.)
6. Counting irregular series of taps. (New.)
7. Folding a square of paper three times. (New.)
8. Tapping blocks in irregular order. (New.)

*Age Seven Years. (2 new.)*

1. Description of pictures. (B.-S.)
2. Naming the first four coins. (B.-S.)
3. Telling number of fingers. (B.-S.)
4. Repetition of five numerals. (B.-S.)
5. Comparing two objects from memory. (B.-S. VIII.)
6. Giving word opposites. (New.)
7. Repeating three digits backwards. (New.)
8. Copying a diamond. (B.-S.)

*Age Eight Years. (6 new.)*

1. Counting value of stamps. (B.-S.)
2. Size of vocabulary. (New.)
3. Folding square of paper five times. (New.)
4. Comprehension. (New.)
5. Giving word opposites. (New.)
6. Giving similarities. (New.)
7. Counting backwards from 20 to 1. (B.-S.)
8. Counting dots. (New.)

*Age Nine Years. (3 new.)*

1. Giving date. (B.-S.)
2. Arrangement of weights. (B.-S.)
3. Using three words in a sentence. (B.-S. X.)
4. Making change. (B.-S.)
5. Definition better than according to use. (B.-S.)
6. Comprehension. (New.)
7. Repeating four digits backwards. (New.)
8. Counting dots. (New.)

*Age Ten Years. (6 new.)*

1. Drawing designs from memory. (B.-S.)
2. Counting dots. (New.)
3. Spelling familiar words backwards. (New.)
4. Giving word opposites. (New.)
5. Counting irregular series of 9-12 taps. (New.)
6. Giving the associated numbers for the dissected parts of a simple form. (New.)
7. Crossing out q, r, s, t in a pied text. (New.)
8. Detecting absurdities in absurd statements. (B.-S. XI.)

*Age Eleven Years. (5 new.)*

1. Words to put in order to make a sentence. (B.-S.)
2. Repeating one or two sentences with 24 syllables. (B.-S. XII.)
3. Giving definitions of abstract terms. (B.-S.)
4. Crossing out q, r, s, t in pied text. (New.)
5. Giving the associated numbers for the dissected parts of a simple form. (New.)
6. Immediate recall of unfamiliar forms. (New.)
7. Giving word opposites. (New.)
8. Locating sections of a divided square from description. (New.)

*Age Twelve Years. (8 new.)*

1. Crossing out q, r, s, t in a pied text. (New.)
2. Spelling familiar words backwards. (New.)
3. Giving the associated numbers for the dissected parts of a simple form. (New.)
4. Immediate recall of unfamiliar forms. (New.)
5. Giving word opposites. (New.)
6. Following directions in a confusing text. (New.)
7. Locating sections of a divided square from description. (New.)
8. Drawing triangles on squares according to directions. (New.)

*Age Thirteen Years to Maturity. (8 new.)*

1. Immediate recall of unfamiliar forms. (New.)
2. Giving word opposites. (New.)
3. Following directions in a confusing text. (New.)
4. Locating sections of a divided square from description. (New.)
5. Drawing triangles on squares according to directions. (New.)
6. Drawing upright forms in inverted positions. (New.)
7. Making logical inferences. (New.)
8. Simple arithmetical operations. (New.)