THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

A CRITICISM OF THE YERKES-BRIDGES POINT SCALE, WITH ALTERNATIVE SUGGESTIONS¹

ARTHUR S. OTIS²

Stanford University

INTRODUCTION

In the Yerkes-Bridges Point Scale the "all-or-none" method of scoring intelligence is replaced by a method which gives partial credit for partial answers presumably in proportion to merit. The avowed purpose of the Point Scale as stated on page 9, is that "It brings out the full value of the test material without increasing the expenditure of time and energy."

The purpose is undoubtedly a laudable one. Certainly, much valuable data are unused by the all-or-none method of scoring. If one child puts the form board together in ten seconds and another takes three minutes to perform the same task, it would seem that there is a difference of intelligence between the two which would be well worth taking account of, and that it is a needless disregard for accuracy to give the same score to each. It is all too difficult to get sufficiently accurate data as it is. It is not urged that more accurate data be necessarily gotten, but that all that is gotten be used if convenient.

In the opinion of the writer, however, the method used in scoring with the Point Scale fails entirely to accomplish the purpose sought. The criticisms that are to be made are of such a nature that it has seemed impractical to attempt to make them without first giving a rather full outline of the requirements of a valid point scale, as conceived by the writer, in the light of

¹R. M. YERKES, J. W. BRIDGES AND ROSE S. HARDWICK. *A Point Scale for Measuring Mental Ability*. Warwick and York, 1915.

²The writer is indebted to Dr. Lewis M. Terman of Stanford University for many helpful suggestions in the preparation of this paper.

which it is believed the criticisms can be made more clearly and easily.³

CONSIDERATIONS REGARDING THE CHOICE OF UNITS FOR A SCALE OF INTELLIGENCE

Although the word "intelligence" as employed by Binet was originally used to denote those mental qualities, differences of which differentiated the "feeble-minded" from the "normal," the general assumption underlying the universal use of the Binet Scale is that these are the same mental qualities, differences of which differentiate older normal children from younger. This is the same as saying that children, in maturing mentally, virtually pass through the various stages of intelligence from what would be for adults feeble-mindedness up to normal-mindedness—they mature in intelligence. It has been found, however, that in terms of "mental age" derived from the number of tests passed, children mature in intelligence at different rates. Those who mature rapidly are called "bright," those who mature slowly, "dull." The general conception is also that the yearly increments of intelligence decrease with age, so that intelligence approaches a maximum at a certain age, perhaps not far from 18, and presumably approximately the same for all persons, this being called the "age of maturity of intelligence." Test results suggest that if we had perfect measures of the intelligences of a very large number of children of each age we would find them distributed for each age approximately according to the law of normal distribution; and, as would follow from what was said above, children who are above or below normal (median for their age) have degrees of intelligence comparable to that normal for older or younger children. Children of the same degree of intelligence (mental age) but of different ages may be said to differ in "brightness."

Although originally, no doubt, the word intelligence was intended to apply only to what we have called "brightness," practically every writer on Binet testing has come to use the word, intelligence, also to mean that which is measured by the mental age (of children as well as of adults). Therefore, it has seemed

³Since writing this article (April, 1916) the writer has undertaken to make a point scale along the lines indicated. It has seemed advisable, therefore, to omit some mathematical proofs and other detailed considerations which will be given in the description of the completed scale.

necessary to adopt a new term, brightness, to denote that in which children may differ who have the same intelligence—mental age. Brightness would therefore be measured by an intelligence quotient.

Standard 10-year intelligence being generally defined as the median 10-year intelligence, or that degree of intelligence possessed by 50% of unselected 10-year-olds, a child of any age possessing this degree of intelligence is said to have a mental age of 10 years. This method of characterization of an individual's degree of intelligence, however, is not adequate, since by it there is no way of expressing the degree of intelligence which is above that normal for adults.

While it will not be known what exact relation exists between a child's degree of brightness at one age and his degree of brightness at another age until the same children have been tested repeatedly in successive years with very exact tests, it is quite probable, and we shall assume, that degrees of "brightness" are theoretically constant for a given child, being a fixed character of the endowment, so that a child who is really below normal at one age will be so at all ages, and that an adult who is above normal was as much above normal relatively at any stage of development. The indications of data thus far gathered tend to confirm this theory. Its truth is, in fact, the foundation of our hopes in testing intelligence, for if it is not true, in part at least, we can not prognosticate, and intelligence measurement will be of no great value.

In stating that intelligence is assumed to approach a maximum, or maturity at some age, say 18, we are, of course, considering what we might call the *absolute* amount of intelligence, as measured in theoretically equal units, as distinguished from being measured on a year scale. Yearly increments of growth of intelligence are equal, of course, in the sense that years are equal, but not in absolute amount, unless perhaps during childhood.

It is customary among writers to denote the degree of brightness of a child in terms of an "intelligence quotient" (I Q), which is the quotient obtained by dividing the actual mental age of an individual by the mental age normal for his chronological age.

⁴ The expression "that degree of intelligence possessed by," will be used hereafter to mean "that degree of intelligence which is equalled or exceeded by the degree of intelligence possessed by," etc.

This amounts to dividing by the chronological age in all cases in which the latter is not above the age of maturity of intelligence. Thus a child of 10 having a mental age of 12 has an I Q of $12-10$ or 1.20 , or an adult of 30 years having a mental age of 15 has an I Q of $15-18$ or $.80$ (assuming 18 years to be the age of maturity of intelligence). This method involves two difficulties; first, by virtue of the above mentioned limitation on the mental age method of expressing degrees of intelligence and second, because an I Q of $15-18$ is very probably not equal to an I Q of $5-6$ since 15-year mental age is very probably much closer to 18-year mental age relatively than 5-year mental age is to 6-year mental age.

The first of these difficulties has been met by assigning arbitrary mental ages above that of maturity to designate the degrees of intelligence above that normal for the age of maturity. At best, however, this method is objectionable.

Terman and Stern have found, however, by successive testing of the same individuals that I Q's gotten by dividing mental by chronological age do remain fairly constant up to the age of 14, which, if we make the hypothesis that degrees of brightness are constant, would indicate that the yearly increments of intelligence are to be considered as equal absolute units up to about the age of 14.

For the purpose of getting a clear idea of the nature of the intelligence of an abnormal child, it is well to express his degree of intelligence in terms of the normal intelligence for other ages or, in other words, in terms of mental age. Thus, it is very significant to say that a 10-year child has an intelligence which is normal for a child of 7 years or of 13 years, or, as we say, that he has a mental age of 7 years or of 13 years. But on account of the above mentioned defects of this method for purposes of comparison of degrees of brightness of children of different ages, it would seem very desirable, if a 10-year child deviates from the average, to be able to say that he has an intelligence which exceeds that possessed by 65% of 10-year-olds or only 35% of 10-year-olds, etc.

If it may be assumed, as would seem reasonable in the light of facts to date, that if a 10-year child exceeds 65% of 10-year-olds in intelligence, he has exceeded and will exceed 65% of children of his age when at any other age, including adulthood,⁵

⁵ The determination of the degree of approximation of this assumption to the truth, as has been said, awaits the testing of the same individuals through successive years.

then this method of expressing the degree of brightness of an individual becomes very convenient for purposes of comparison or prognostication. Thus we may say that a 10-year-old exceeding 65% of 10-year-olds in intelligence is of the same degree of brightness as a 13-year child or as an adult if either of these exceeds the same percentage of persons of his age in intelligence. Also we may say that he is "brighter" than an individual of any age who exceeds a lesser percentage and "duller" than one who exceeds a greater percentage. The maximum ultimate attainments in test performances of any child upon reaching maturity can thus be predicted from his rank at any age and a knowledge of the ability of adults of the same rank.

For reasons which will appear later, it will be found best to deal principally with intelligence as measured on an absolute scale. Since in measuring intelligence we begin by finding the degrees of performance in certain tests which a person may accomplish, and end by stating his mental age or the percentage of persons of his chronological age whom he exceeds in intelligence, it will be seen therefore that an absolute scale of intelligence may be graduated somewhat arbitrarily since it is only a means to an end, upon the attainment of which the arbitrary scale itself is no longer in evidence.

For the purpose of expressing the different absolute degrees of intelligence in tangible units, let us begin with that degree which is normal for adults—that which is possessed by 50% of unselected adults—and call this an absolute intelligence of grade 100. Then assuming the distribution of absolute grades of intelligence to be normal we could represent that grade possessed by 75% of adults as a grade of 90 and that grade possessed by but 25% of adults as a grade of 110. The unit difference in absolute intelligence is thus defined as $\frac{1}{10}$ of the quartile of the distribution of absolute degrees. According to this definition, the correspondence between the grades of general intelligence denoted by the values 60, 70, etc., representing respectively —4Q (Q = quartile), —3Q, etc., of the distribution, and the percentages of adults which theoretically would possess those values would be as shown in Table I.

TABLE I

Grade of general intelligence.....	60	70	80	90	100	110	120	130	140
Point in the distribution.....	—4Q	—3Q	—2Q	—Q	C.T.	+Q	+2Q	+3Q	+4Q
Percentage of adults possessing these..	99.7	98	91	75	50	25	9	2	0.3

So far we have only defined hypothetical absolute degrees or grades of intelligence in terms of an arbitrary scale. We may be permitted to do this irrespective of what degrees of performance in any test denote the grades of absolute intelligence and irrespective of the percentage of children of any age possessing any absolute grade of intelligence. It is not even essential that the distribution of grades of adult intelligence be normal as measured in any of the natural units such as dollars of earning capacity, or digits remembered, or words in vocabulary or even mental age. The absolute units of intelligence are purely hypothetical and are so defined that the distribution of grades of adult intelligence *will be normal*.

In the sense in which the absolute grades of intelligence possessed by unselected adults fall into a normal distribution, the absolute increments of intelligence composing the units of the absolute scale are equal increments, and in this sense, therefore, an individual possessing the absolute grade of 120 is as much more intelligent than one possessing a grade of 110 as the latter is more intelligent than one possessing a grade of 100, etc. This suggests a more mathematical measure of the degree of brightness.

Thus, to express the degree of brightness of an adult of absolute grade 120 in comparison with the normal adult grade of 100, we might divide 120 by 100 and say the individual had a degree of brightness or *coefficient of brightness* of 1.20. An adult possessing a grade of 110 would therefore have a coefficient of brightness of 1.10, etc. The first is therefore as much brighter than the second as the second is brighter than a normal adult. Such a fact could not be determined when brightness is expressed in terms of percentage rank in intelligence. The coefficient of brightness of a child would then be expressed as the quotient obtained by dividing the absolute grade of intelligence possessed by him by the absolute grade which was normal for children of his age.

Having so made an arbitrary absolute scale of intelligence, however, it would not follow that an individual's coefficient of brightness in these terms would be constant. That is, we do not know that an intelligence grade of 80 in these terms would bear the same proportion to one of 100 that one of 60 would to one of 75, although the first is in each case $\frac{4}{5}$ of the second in terms of our arbitrary scale. This is because we do not know

that we have chosen our units so that zero on the scale represents "just not any intelligence" or true absolute zero.

A possible method would be the following if we may assume that the distributions of intelligences of children of any given age, when measured on an absolute scale, are "normal," and that if the units are properly chosen these distributions will have the same coefficients of variation (such coefficient being taken as the quartile value divided by the median). Such an assumption is merely a corollary to the assumption that degrees of brightness are constant for a given individual. The method would then be merely to find the difference between median adult intelligence and true absolute zero, as a multiple of the Q value (quartile distance on the scale) and then to change the number of units in Q from 10 (arbitrarily assigned) to that number which will make the above mentioned difference 100.

Terman^a has pooled the I Q's of 1000 unselected children and finds them distributed as shown in Table 1A. Here we note that 75% of individuals reach an I Q of .92 while 25% reach 1.08, meaning that the quartile distance on the scale of I Q's is .08 points. The percentages corresponding to the positive and negative multiples of Q are very near the theoretical values to be expected in a normal distribution as may be seen by comparing Tables 1A and 1B. If the yearly increments of intelligence are fairly equal in absolute units up to the age of 14, these data would incline us to believe that the median intelligence of individuals of any given age in true absolute units is not far from $12\frac{1}{2}$ times the quartile distance in the distribution of those values. This would mean that the Q of the distribution of adult intelligences should perhaps be 8 units. If that were the case (which let us suppose for illustration) the percentages of adults possessing the various grades of intelligence by the true absolute scale would be as shown in Table 1B.

TABLE 1A

Showing the Percentages of 1000 Unselected Children Reaching Various Intelligence Quotients, as Found by Terman

Percentages of Children.....	.99	.98	.97	.95	.90	.85	.80	.75	.66 $\frac{2}{3}$	.50
Corresponding I Q's.....	.70	.73	.76	.78	.85	.88	.91	.92	.95	1.00
Percentages of Children.....	.33 $\frac{1}{3}$	25	20	15	10	5	3	2	1	
Corresponding I Q's.....	1.06	1.08	1.10	1.13	1.16	1.22	1.25	1.28	1.30	

^a LEWIS M. TERMAN. *The Measurement of Intelligence*, 1916.

TABLE 1B

Showing the Theoretical Percentages of Adults Attaining the Various Absolute Grades of Intelligence, the Quartile Value of the Distribution Being Taken as 8 points with Median at 100

Grades.....	60	62	64	66	68	70	72	74	76	78
Percentages	99.96	99.93	99.88	99.79	99.65	99.43	99.09	98.58	97.85	96.82
Grades.....	80	82	84	86	88	90	92	94	96	98
Percentages	95.41	93.54	91.13	88.11	84.41	80.04	75.00	69.35	63.21	56.70
Grades.....	100	102	104	106	108	110	112	114	116	118
Percentages	50.00	43.30	36.79	30.65	25.00	19.96	15.59	11.89	8.87	6.46
Grades.....	120	122	124	126	128	130	132	134	136	138
Percentages	4.59	3.18	2.15	1.42	0.91	0.57	0.35	0.21	0.12	0.07
Grade.....	140									
Percentage.	0.04									

CONSIDERATIONS REGARDING THE SELECTION AND USE OF TESTS

Before we can find the percentages of children of different ages who possess the different absolute grades of intelligence as defined, it will be necessary to have an objective means of determining the grades of intelligence of adults and children. It is this purpose, of course, which "tests" serve. Now the tests may be such as are scored simply as right or wrong and the grade of intelligence indicated by the number of tests "passed" or answered correctly. This constitutes what is called the "all-or-none method" which is used to a large extent in the Binet Scale and which is condemned by Yerkes and Bridges.

Another method is that of defining degrees of performance in a single test and determining the grade of intelligence from the degree of performance achieved. It is this method which is believed by the writer and presumably by Yerkes and Bridges to be superior to the all-or-none method, since, as has been said, the degree of a performance such as the quickness of putting together a form board, is of additional significance to the mere fact of its having been put together and since it is no more trouble appreciably to take the time than not to do so.

It should be noted that the second method is used to some extent also in the Binet scale, in that different degrees of performance are recognized and scored as such in a number of tests, such as the vocabulary, fables, memory for ideas, picture interpretation, etc. These will be considered at length later.

Since it is evident that where different degrees of performance in a test may be easily determined and are believed to be significant, they should be taken into account, let us consider the making of a scale or series of tests for measuring the degrees of each of the various abilities of an individual which are deemed to depend largely on intelligence. For purposes of illustration and comparison with the Yerkes scale we shall confine ourselves to those tests which have been tried out, particularly those which are used in the Binet Scale.

In selecting tests and determining upon the degree of exhaustiveness of each, many considerations should be borne in mind, which we need not discuss. It is probable that some of the lengthy tests of the Binet Scale, such as the fables, code, and vocabulary tests should be either shortened or replaced by others, while some of the short ones, such as the dissected sentence test, could be made to afford a greater variety of degrees of performance. It is probable also that a number of new tests could be added to the Binet list to advantage. For the sake of illustration let us assume that we shall find it advisable to include in our series of tests those in Table II.

TABLE II
Possible List of Tests Used in the Series

1. Vocabulary	10. Memory for ideas
2. Completion	11. Memory for designs
3. Opposites	12. Dissected sentences
4. Similarities	13. Resourcefulness
5. Analogies	14. Fables or proverbs
6. Logical reasoning	15. Picture interpretation
7. Arithmetical reasoning	16. Formboard
8. Memory for digits	17. Associations
9. Memory for sentences	18. Sense discrimination

CONSIDERATIONS REGARDING THE COMBINING OF THE RESULTS OF THE TESTS

Our next task is to combine the measures of ability by the various tests so as to get the best measure of intelligence. This cannot be done, of course, by merely adding together or averaging the scores of the several tests. These must be reduced to a common basis first. One method which might be used would be to equate the averages to the standard or median deviations of the distributions of scores made in the several tests by the same group of individuals. This method would be unsafe, however, until such time as the elements of the several tests

have been very carefully graduated in difficulty so that for any given test the increments of difficulty between the elements are very nearly equal, thus guaranteeing comparability of the distributions of scores in the several tests.

The method advised is, therefore, that the common basis to which to reduce all test scores is the absolute scale of ability. That is, we should find the absolute grade which should be assigned to each degree of performance in each test series. Thus, any degree of performance achieved by 50% of adults would be considered as grade 100 in that test and would be equal to grade 100 in every other test. Degrees of performance in the several tests achieved by 75% of adults would similarly be considered as of grade 90 and therefore equal, etc.

It would be necessary, no doubt, to supplement this method in the cases of the lower degrees of performance by other means since very large percentages of adults would be untrustworthy. Degrees of performance below 80, say, might be equated to one another by finding their relation to an age scale, that is, by finding the degree of performance in each test which may be achieved by 50% of the children of the various ages; or rather, perhaps, to find the age at which just 50% of children could achieve each of the designated degrees of performance in each test. A method which could be employed for this purpose is explained by the writer elsewhere.⁷

Another method, which would probably be more suitable in cases where there are a considerable number of degrees of performance in a single test, would be to find the curve of relation or "*regression curve*"⁸ between age and the several degrees of performance, letting ages be the abscissas and finding first the median degrees of performance for each division of the age scale as a basis for drawing the curve. From this curve then, the precise age could be found for which each degree of performance was most probably standard. In this way a series of scales (virtually diminutive Binet Scales) could be made, each complete in itself in a small way.

The question may now arise in the mind of the reader why we do not rest with finding the mental age of the child in each test and determining a representative value of these mental ages.

⁷ *Some Logical Aspects of the Binet Scale*, Psychological Review, March and May, 1916.

⁸ See WM. BROWN, *The Essentials of Mental Measurement*, 1911.

The reasons for converting the test score from mental ages into absolute grades of intelligence or of the specific ability, are as follows: First, it is not correct to average the scores as mental ages in all cases since the absolute average of a mental level of 16 years and one of 12 years is not one of 14 years but is less than 14 years, that is, 16-year ability is, absolutely considered, closer to 14-year ability than is 12-year ability. A second reason is that we cannot express degrees of ability above normal adult ability in terms of any mental age; and a third reason is that the scores of the different tests should be weighted to take into account different degrees of efficiency of different tests.

After having found the curve of relation between age and degree of performance in each of the tests, then through the medium of the age scale the correspondence of degrees of performance in each test scale can be found with those of every other. It would then remain to transmute measures of the various degrees of performance in each test from age values into units of the absolute scale. Various methods could be suggested for accomplishing this. One would be to find the percentage of adults which pass that degree of performance in each test corresponding to ten-year ability, to find the absolute value corresponding to the median of these percentages and to consider this as being the absolute value which corresponded to ten-year ability, etc. A more refined method might be used which involved a comparison of the variabilities of the distributions of grades of intelligence of the children of the different ages for the purpose of determining the curve of growth of intelligence. Thus we might find the correspondence of the age scale determined as suggested above, to be that shown in Table III.

TABLE III

3	4	5	6	7	8	9	10	11	12	13	14	15	16	18
20	27	34	41	48	55	62	68	75	81	87	92	95	98	100

We have assumed that the distributions of absolute grades of intelligence attained by the children of each age were normal distributions having variabilities proportional to the values of their central tendencies (this will be the case if our hypothesis regarding the growth of general intelligence is true), and have considered it possible to so graduate our absolute scale that the quotient obtained by dividing the actual absolute grade of intelligence of any child by the absolute grade which is normal for

children of his age would give us an absolute intelligence quotient (A I Q) which is constant for a given child throughout all ages and may be taken as the "coefficient of brightness" of the child and as a measure of the degree of brightness which he will have also when an adult. By means of this "absolute intelligence quotient" obtained for any child we may predict the probable ultimate absolute grade of intelligence which he will reach and thereby predict the probable maximum intellectual attainments which he will achieve.

WEIGHTING THE SCORES ACCORDING TO VALUE AS MEASURES OF INTELLIGENCE

If we merely found the average or median level of ability attained in a series of tests by a given child, we would be assuming that all tests were of equal significance as measures of intelligence. As has been stated, however, certain tests are much more significant than others. This may be due to finer scale gradations or to more time devoted to giving the test or to the fact that the ability tested is less dependent upon factors other than those which we consider as constituting intelligence. The tests should be weighted, therefore, according to their significance as measures of intelligence.

There are several ways, of course, in which this could be done. One would be to make use of a criterion of intelligence apart from the tests, to which they would be compared. Thus we might get as reliable an estimate as possible of the intelligences of pupils by their teachers and find the manner of combining the results of the tests which would give the best approximation to that estimate. The method generally employed in such a case is that of partial correlation and the regression equation.⁹

Another and perhaps better method of weighting the test results could be employed in which no external criterion was used. Thus McCall¹⁰ finds the amount of correlation between each test and a measure obtained by combining the measures of all the tests when these were weighted by "guess." It may be shown, however, that a very good approximation of the relative degrees

⁹ See G. UDN YULE, *An Introduction to the Theory of Statistics*, Chapter 12, or TRUMAN L. KELLEY, *Bulletin of the University of Texas*, 1916, No. 37.

¹⁰ W. A. MCCALL, *Correlations of Some Psychological and Educational Measurements with Special Attention to the Measurement of Mental Ability*. School and Society, January 6, 1917.

in which each test measures a factor common to all the abilities tested may be found by a consideration of the amounts of inter-correlation between the tests taken singly. Proof of this may be derived from the formula for partial correlation.

Assuming that we have satisfactorily determined the weightings to be given each test result, an illustration will be given of the determination of the weighted average of the levels of attainment in the various tests, supposing that an abbreviated series of tests is being used consisting of the tests named in Table V. Let us assume that a child attained a degree of performance in each test which would yield the grade scores designated in column 2, and let us suppose the respective weights of the tests as determined in an approved manner are as shown in column 3. (These are made small numbers for convenience of illustration.)

TABLE V

1	2	3	4	5	6
Tests	Test Grade Scores	Weights	Weighted Scores	Altered Weights	"Points"
1. Digits.....	77	×	2 = 154	1.0	77
2. Similarities.....	83	×	3 = 249	1.5	125
3. Analogies.....	80	×	3 = 240	1.5	120
4. Completion.....	88	×	4 = 352	2.0	176
5. Discrimination...	65	×	1 = 65	0.5	32
6. Sentences.....	72	×	2 = 144	1.0	72
7. Vocabulary.....	95	×	5 = 475	2.5	238
Totals.....	7)560	20	20)1679		840
Average	80	Weighted average	84		

Now, of course, the sum of the weights may be made equal to any number so long as the proportions are kept the same. They may be altered, therefore, so that their sum equals 10 (using decimal fractions). Moreover, the values of the grade scores for each test may be multiplied by the proper weight thus found and a new table made for each test in which the weighted score in "points" can be read directly, opposite the degree of performance attained. Thus for the sentence test, the correspondence of the degrees of performance (number of sentences repeated) and the absolute grade of ability reached (measured on the absolute scale) and the number of "points" (weighted absolute grade) might be as shown in Table VI.

According to this scheme, the weighted average of the absolute scores of a person achieving median adult performance in

each test would be 1000 "points." That is, 1000 points equals a score of 100 on the absolute scale and, as may be shown, all other scores in points would be just 10 times the corresponding scores on the absolute scale. The latter might now be discarded having served its purpose as the basis for the construction of a "point scale" consisting of weighted scores corresponding to the degrees of performance in each test.

THE CHART

In the Chart we have this Point Scale in the center with 1000 points as median adult intelligence. Values of the grade may be considered as $\frac{1}{10}$ of each number of points. On the left are curves postulating the growth of absolute intelligence. On the right are scales of degrees of performance in nine tests assumed to constitute an abridged series with their corresponding numbers of points.

TABLE VI

Hypothetical Corresponding Scales for a Sentence Test

Number of sentences.....	1	2	3	4	5	6	7	8	9	10	etc.
Absolute grade of ability required.....	20	22	25	30	39	52	65	78	89	100	etc.
Points (Weight 2/6.7).....	6	7	8	10	13	18	22	26	30	33	etc.

Thus we see that to find the number of points designating the absolute degree of general intelligence of any individual, it is necessary merely to find the degree of performance which he can achieve in each test, note the number of points in the chart corresponding to this degree of performance, find the sum of those numbers of points for all the tests and this sum will be the "point-score" which denotes the absolute degree of intelligence of the individual in the determination of which due consideration has been given to the different degrees of efficiency (reliability and correlation with intelligence) of the various tests and which point scores are expressed in units which are equal according to the theory of normal distribution of grades of adult intelligence.

From this "point-score" which may be used to denote the absolute degree of general intelligence and the age of an individual, we may determine upon which curve of growth he is progressing, so to speak. Thus a 12-year-old with 880 points is progressing on the "75% curve" showing that he exceeds 75%

THE CHART FOR A Scale of Absolute Degrees of Intelligence

Age Scale	5-3	4-	5-	6-	7-	8-	9-	10-	11-	12-	13-	14-	15-	16-	18=
No Words Vocabulary (wt.5) Points	5 33		8 50	10 67	15 82	20 90	25 102	30 113	35 125	40 136	45 143	50 152	55 157	60 164	65 167
No Words Completion (wt.5) Points	10 35	20 46	25 54	30 62	40 80	45 88	50 96	55 104	60 118	65 133	70 148	75 163	80 174	85 188	90 200
Number Opposites (wt.4) Points	2 29	5 39	10 44	15 53	20 59	25 66	30 72	35 79	40 86	50 105	55 113	60 125	65 137	70 151	75 166
Number Reasoning (wt.4) Points	2 28	3 35	4 40	5 44	6 52	7 56	8 63	9 70	10 78	12 100	13 110	15 126	16 134	17 144	18 156
Number Analogies (wt.3) Points	2 21	4 26	5 31	6 35	7 39	8 43	9 47	10 53	11 58	13 72	14 83	15 93	16 100	17 107	18 115
Number Digits (wt.3) Points	3 20	4 27	5 34	6 34	7 40	8 48	9 54	10 60	11 68	13 77	14 84	15 91	16 99	17 107	18 111
Syllables Sentences (wt.3) Points	6 20	12 27	14 38	16 45	18 57	20 60	22 75	24 84	26 92	28 100	30 108	32 116	34 123	36 127	38 135
No. Counts Designs (wt.2) Points	10 14	20 19	30 23	40 27	50 31	55 35	60 40	65 47	70 53	75 60	80 65	85 71	90 76	92 76	94 77
No. Words Association (wt.1) Points	18 7	20 10	23 13	25 17	30 20	37 23	45 27	55 33	65 40	75 48	85 57	90 63	95 70	100 77	105 83

of persons in brightness, will attain an ultimate degree of intelligence represented by 1100 points and therefore has a "coefficient of brightness" of $1100/1000$ or 1.10 and may be expected to ultimately achieve the degrees of performance in the tests opposite the 1100 points on the Point Scale. In the finished chart, of course, a much larger number of curves should be drawn, or else these must be supplemented by a table composed of portions similar to Table IB for every fractional year of age.

It is quite possible that in the actual making of a point scale as outlined above, some modifications in the arrangement of tests would be necessary, since certain tests which are excellent for persons above a certain age might be incapable of being used with very young children, such as those who could not read in the case of the completion test. In this case it would be necessary merely to substitute a suitable test in the lower part of the scale, the reliability of which was such that it would be weighted the same. In fact if the reliability of a single test were different for different extremes of the scale it could be divided, and each half could be completed with tests having the proper reliability. Scale arrangement on the Chart is of no consequence in giving the tests.

In postulating the absolute grades of intelligence which corresponded to certain degrees of performance in three of the tests, the vocabulary, digit and sentence tests in the Chart, reference was made to the Stanford Revision of the Binet Scale and the numbers of words of vocabulary, digits repeated on third trial, and syllables in the sentences repeated, were placed opposite the mental age corresponding to the age group in which these degrees were found. Thus since "Repeats 4 digits" was found in age group IV, 4 was placed opposite mental age 4 years.

THE MAKING OF A "BINET POINT SCALE"

Before taking up the specific criticism of the Yerkes-Bridges Point Scale, let us see how the Stanford Revision of the Binet Scale compares with our suggested point scale. For this purpose let us convert it into a point scale as far as possible. For the sake of conciseness let us denote the various tests by their numbers and catch names referring the reader to the test blank for a more complete knowledge of the tests.

TABLE VII
A Binet Point Scale

Point Scale	Mental Age	Points	List 1	Points	List 2	Points	List 3
12	III	2	(1) Body	2	(3) Pictures	2	(2) Objects
24	IV	2	(1) Lines	2	(2) Forms	2	(3) Pennies
36	V	2	(1) Weights	2	(4) Definition	2	(3) Comparison
48	VI	2	(1) Right and Left	2	(2) Pictures	2	(3) Pennies
60	VII	2	(1) Fingers	2	(2) Pictures	2	(4) Knot
72	VIII	2	(6) Vocabulary 20	2	(5) Definition	2	(4) Similarities
84	IX	2	(2) Weights	2	(1) Date	2	(5) Words
96	X	2	(1) Vocabulary 30	2	(3) Designs	2	(2) Absurdities
120	XII	4	(1) Vocabulary 40	4	(3) Fables	4	(8) Similarities
144	XIV	4	(1) Vocabulary 50	4	(2) Induction	4	(5) Arithmetic
168	XVI	4	(1) Vocabulary 65	4	(3) Fables	4	(6) Boxes
192	XVIII	4	(1) Vocabulary 75	4	(2) Paper	4	(5) Digits 7 back.

A Binet Point Scale—Continued

Point Scale	Mental Age	Points	List 4	Points	List 5	Points	List 6
12	III	2	(7) Digits 3	2	(5) Name	2	(6) Syllables 6-7
24	IV	2	(6) Digits 4	2	(5) Compreh. 1	2	(7) Syllables 12-13
36	V	2	(6) Commissions	2	(5) Patience	2	(2) Colors
48	VI	2	(5) Coins	2	(4) Compreh. 2	2	(6) Syllables 16-18
60	VII	2	(3) Digits 5	2	(5) Differences	2	(7) Days
72	VIII	2	(2) Counts	2	(3) Compreh. 3	2	(1) Ball and Field
84	IX	2	(4) Digits 4 back.	2	(6) Rhymes	2	(7) Months
96	X	2	(7) Digits 6	2	(5) Compreh. 4	2	(8) Syllables 20-22
120	XII	4	(6) Digits 5 back.	4	(2) Definition	4	(4) Ball and Field
144	XIV	4	(7) Digits 7	4	(3) Differences	4	(6) Clock
168	XVI	4	(5) Digits 6 back.	4	(2) Differences	4	(7) Syllables 28
192	XVIII	4	(3) Digits 8	4	(6) Ingenuity	4	(4) Memory

Suppose we consider the first test scale as made up of III, (1) Body; IV, (1) Lines; V, (1) Weights, etc., as shown in Table VII. We could in this way make up six "lists" of tests, each list corresponding to a "test" in the Chart and each test corresponding to a degree of performance in a test in the Chart.

Different degrees of the same ability are recognized in the Binet Scale in the following: Digits, Syllables, Definition, Comprehension, Pennies, Pictures, Vocabulary, Similarities, Weights, Time orientation, Designs (Forms), Ball and Field, Fables, and Differences. So far as possible, therefore, those of a kind have been placed in the same lists. Thus we see that List 4 is almost entirely made up of degrees of memory for digits.

In so far as any list is composed of tests representing different degrees of the same ability, such as vocabulary or memory for digits, it may be said that the ability to achieve a certain degree of performance, such as defining 30 words, or repeating six digits, presupposes possession of the ability to achieve all lesser degrees of performance in the same list and none of the greater degrees. This cannot be said, however, of any of the lists where the abilities are different, since a child might conceivably succeed in the tests III, IV, V, VIII and IX of List 2 and fail in VI, VII, as well as those above IX. It is perhaps fairly safe to say that his "average ability" in the tests of a single list is approximately represented by the age corresponding to his point-score in the list, even though the tests passed be not in consecutive order.

To use the Binet Point Scale, the tests constituting the regular Binet Scale would be given exactly as is usual. One should then note how many tests in each list were passed and find the sum of the points corresponding to each test passed, as indicated in the column headed Points. This sum would constitute the "point-score." The correspondence between the Point Scale and the Mental Age Scale at the left of the table has been arranged on the assumption that the passing of six tests (12 points) indicates mental age III, that the passing of twelve tests (24 points) indicates mental age IV, etc. A point score of 66 points ($60 + 6$) indicates a mental age of seven years (VII) and six months, a point score of 164 points ($144 + 12 + 8$) indicates a mental age of fifteen years and eight months, etc. Each point corresponds to an increment of one month in mental age. (Mental ages XVI and XVIII are fictitious.)

SPECIFIC CRITICISM OF THE YERKES-BRIDGES POINT SCALE

It may be seen, therefore, that to convert the present Binet Scale into a point scale is merely to reorganize it on paper. As for the administering of the tests, as above mentioned, this would be done in exactly the same manner for our Binet Point Scale as for the Binet Scale. Even the scoring would be done in the same way. The only difference is in the manner of expressing the result—whether the degree of maturity of the child be spoken of in terms of “mental age” or “points.” As a matter of fact the upper mental ages are merely arbitrary designations anyway, as has been said. In other words, we may say that *the present Binet Scale is already a point scale*, for all the purposes for which Yerkes’ and Bridges’ alteration¹¹ of the Binet Scale could be used. If Yerkes and Bridges had let each month of mental age by the present Binet Scale be called a “point” or given one “point” for passing each test, they would have had a series of “point scores” which would have “lent themselves readily to statistical work,” as they say, including the use of different norms for different classes if desired, the comparison of mental ability with physiological maturity, etc. At the same time, the Binet scores thus converted would be more valid for the following reason.

Obviously, if a “point” has any meaning, it must stand for an increment of intelligence or intellectual maturity having the same value as that of every other “point,” as nearly as possible. In the Binet Point Scale one “point” indicates one month’s normal growth in intellectual maturity. In so far as months of normal growth are equal, the values of “points” in the Binet Point Scale are equal and therefore statistically valid. Those points coming near together in any part of the scale may be assumed to be practically equal (if the tests are properly standardized), as rate of growth probably changes gradually. In so far as the values of the points in the two ends of the scale are not equal (on account of slowing down in rate of growth) this error can only be corrected by some such method as we have outlined for making a point scale, in which “points” refer to equal absolute increments of intelligence.

¹¹ The “Point Scale” is practically the Binet Scale remodeled.

In Yerkes' and Bridges' alteration of the Binet Scale, however, even the time basis of equality in point values has been lost by a seemingly complete disregard for the necessity of having such values equal. One has but to examine Yerkes' and Bridges' alteration to note the absence of any apparent system in the assigning of points to increments of mental ability. For example, the abilities required to repeat series 3, 4, 5, 6, and 7 digits are assigned 1, 2, 3, 4, and 5 points, respectively. Now the assumption implicit in this arrangement, taken by itself, is that the increments of mental growth from each such ability to the next are all equal. Not only is no basis presented upon which this arrangement was made, but the indications of data strongly suggest that such is not the case, since the amounts of growth necessary to gain these four increments of ability, 3 to 4, 4 to 5, 5 to 6, and 6 to 7, digits are found to be approximately 1, 3, 3, and 4 years respectively, as indicated by the placement of these tests in the Binet Scale.

Similarly, one point is given in Yerkes' and Bridges' alteration for defining in terms of use each of the following words: chair, horse, fork, baby. Obviously, the difference between the ability to define one of these and the ability to define four of them in terms of use only, is not to be compared with the difference between ability to repeat four digits and the ability to repeat seven digits, though these differences are represented in both cases by increments of 3 points. We believe that these illustrations are sufficient to show that in altering the Binet Scale to make a point scale, Yerkes and Bridges have not only gained nothing over the present Binet Scale, since it is itself already a point scale for those who wish to say "points" instead of "months," but they have materially detracted from its value by a disarrangement which renders the value of a "point" quite meaningless. Such improvements as (1) the expressing of the degrees of intelligence in absolute units so that the value of one point is equal to the value of every other point and so that a given degree of brightness is expressed by the same coefficient at all ages, (2) the defining of the degrees of intelligence above that normal for adults in terms of the percentages of adults possessing these degrees, and (3) the weighting of the results according to the efficiency of the tests, are not even suggested in Yerkes' and Bridges' alteration of the Binet Scale.

Since the above was written, there has appeared in this JOURNAL an article¹² in which the defects of the Yerkes-Bridges arrangement of the Binet Scale are even more clearly shown than can be done by a consideration of the scale construction.

On page 604 of that article, the writers note the extraordinary fact that "During the fourth year of life, intellectual ability (always understood as measured by the point scale) increases more than fifty per cent.; during the fifth year of life, by scarcely more than thirty per cent.; during the seventh year, by approximately fifteen per cent." This shows the kind of results that may be gotten if one point is given to each of too many small increments of ability in the easy part of the scale, as suggested above in the case of defining words. It also shows what is deemed an altogether unwarranted confidence on the part of the writers in the power of numbers of arbitrarily assigned points to "indicate *directly* the rate, or annual increments of intellectual growth," (p. 603, italics mine) whereas the appearance of such questionable results should have immediately suggested to the writers the probability of an erroneous construction of the scale.

The obvious aim of intelligence measurement applied to children is to get a measure of the brightness of one child which is *comparable* with measures of brightness of other children of different ages. Presumably therefore, a measure of brightness which is valid, must stand for the *same degree of brightness* at all ages, which means that approximately the *same percentages* of children of the different ages should obtain any given measure of brightness.

Let us see how well this condition is fulfilled in the Yerkes-Bridges arrangement. In Table III in the above mentioned article a "coefficient" of 1.20 or more is attained by percentages of children of ages 4-5, 6-7, 8-9, 10-11, 12-13, 14-15, 16-17, 18-on, respectively, as follows: 34.5%, 33.0%, 15.8%, 13.7%, 5.5%, 6.5%, 0%, 0% (total number of individuals in the "normal group" upon which the measurements were taken = 1282). Certainly these "coefficients of intelligence" do not represent the *same degree of brightness* for the different ages. Obviously, therefore, they are not *comparable* with one another. Can they, therefore, be said to measure brightness or "intelligence" at all?

¹² ROBERT M. YERKES AND LOUISE WOOD. *Methods of Expressing Results of Measurement of Intelligence: Coefficient of Intelligence*. December, 1916, 593-606.

One can easily imagine a thoughtful investigator obtaining this amount of irregularity in what he hoped would be a valid measure of brightness, but it is inconceivable that in the face of such evidence of failure, one should be led into making the statement (p. 598) that "extraordinary intellectual ability is fairly common up to eight years, and then becomes very uncommon."