

14-RET-FMK

Private 1920 - JPA; State Ind. Ed. - Kuhlmann
Journal of Psycho-Aesthetics, vol. 18, no. 3, March 1924

SOME RESULTS OF EXAMINING A THOUSAND
PUBLIC SCHOOL CHILDREN WITH A REVISION
OF THE BINET-SIMON TESTS OF INTELLIGENCE
BY UNTRAINED EXAMINERS. FIRST
ARTICLE.

BY F. KUHLMANN, *Faribault, Minnesota.*

In the spring of 1913 the public school children from first to the seventh grade, inclusive, of Faribault, Minnesota, were examined with my revision of the Binet-Simon tests by twenty of the teachers. The immediate object of the examinations was not to secure further norms for the tests, but to furnish the school authorities and teachers with a more accurate knowledge about the school children. Twenty teachers were chosen and given a very brief preliminary training in the use of the tests. This training consisted of my first demonstration of the use of the tests to them by examining a number of feeble-minded children at the Minnesota School for Feeble-Minded Children. After this the teachers were divided into groups of three in each group as examiner and the other two as observers. In this way each teacher examined a small number of feeble-minded children for practice. The difficulties that the examiner met and criticisms were discussed in each case. Each teacher spent between ten to twenty hours in this way. After this amount of practice, supplemented by informal discussions, they began the examination of the public school children. The examination was done mostly in the class-rooms during recesses and after school hours, when the rooms and halls were cleared of pupils. The results have three different lines of interest. First, the ability of the examiner without any special training to use the tests with a reasonable degree of accuracy. In connection with this question, it should be stated that the revised scale used contained

*See "A Revision of the Binet-Simon System for Measuring the Intelligence of Children," Journal of Psycho-Aesthetics, Monograph Supplements, Sept., 1914.

more detailed statement on how to give each test and how to interpret the responses than the authors of the tests supply. It is the scientific value of the results in showing further possibilities of improvements in the scale, and in showing the mental development of a representative group of public school children. The usefulness of such results to the school authorities and teachers. In reporting the results on the first two in the present study, advantage will also be taken of the occasion to bring up and consider rather in full the more fundamental problem of a scientific scale of tests which recent studies have brought to the foreground of discussion.

A. The Untrained Examiner.

Psychologists seem to be almost universally agreed that a psychological training and some practice in the use of the Binet-Simon tests are required of an examiner if he is to obtain accurate results. There is a similar concensus of opinion that an intelligent person without special training or practice will find the tests useful in obtaining a better understanding of a child's mental development than he can get in any other way, or at least without very prolonged and close observation. But it is not at all clear yet just what kind of training and how much practice and experience must be demanded. Is it a familiarity with the general principles of psychology, or with the mental development of the child, or a drill in laboratory technique that is needed? Will the practice derived from examining a dozen children suffice, or must it be twenty-five, fifty or several hundred? We have also as yet no definite idea as to how large errors may expect from the untrained examiner. The present study has had occasion to observe the difficulties that about twenty untrained examiners met in giving the tests, each examiner examining from several to twenty-five children. A study of the results by the twenty Faribault teachers and a few others also will throw some light on some facts about this question. An analysis of these observations will give some idea as to just what kind of training is required to make a successful examiner.

The principal main difficulty is lack of familiarity with the procedure for giving each individual test, which remains until a considerable number of children have been examined. The de-

tails of these directions in the whole system of tests many to be mastered in any other way than through the practice in giving the tests. To eliminate this factor entirely an average person probably needs to examine about 100 children. In the absence of this familiarity the unpracticed examiner has to resort to one of two methods. He either guesses at the directions and gives the tests usually in considerable modified ways, or he stops to read the directions first, giving each test. Either procedure is detrimental to the results. The former is entirely unpermissible because a slight change in the manner of giving a test may sometimes actually alter it. The latter is fatal to the proper attitude of interest of the child under examination. It is very essential to arouse and maintain the child's best efforts. For when these are not obtained or is lost, we have no longer any means of knowing how much the child's failure is due to lack of effort and how much is due to lack of mental development. Hesitations on the part of the examiner, or making the child wait between tests, very easily causes him to lose interest. For the sake of the examiner's practice, however, the latter is the course to follow. For guessing at the directions and giving the tests without directions makes it more difficult to learn to give them correctly.

A second difficulty which appears largely at the beginning is an inability to interpret the responses of the children correctly. It is not always easy to judge whether a response should be accepted as satisfactory or regarded as a failure to pass the test. The variety of responses for some tests is very large, and it is impossible to classify them all in such a way that the examiner can use the classification without error. In my experience with the tests an effort has been made to give directions for interpreting responses as well as for giving the tests, wherever it seemed at all called for. My experience since then, however, has shown that more are still needed to avoid difficulties in interpreting responses in some cases, for untrained examiners. Fortunately this difficulty does not affect the majority of the tests. In most cases the response is at once obviously satisfactory or obviously a failure. Where the difficulty is present it disappears in a measure with practice. Familiarity with the

Methods of responses obtained facilitates ready classification and improves the ability to judge them correctly. However, to judge them correctly sometimes implies a knowledge of some logical principle involved, or of some trait in the mental development of children. Where this knowledge is absent, the interpreting of responses will remain.

In this connection a third and more serious difficulty arises. These are sometimes of such a nature as to call for a variation in the procedure in giving a test. Each test aims at a definite object. It is to ascertain whether the child is capable of the performance involved. The directions for giving a test are intended to best bring out the child's ability in this performance, but they fit only the average child under average conditions. A failure to respond at all, or a response that can not be interpreted as either a "pass" or a "failure" calls for a variation in the procedure in giving the test. In the absence of a definite classification of these responses, directions on how to follow the regular procedure in giving a test can not be made.

My revision gives some such supplementary directions. It lays down the general rule to "follow them literally unless obvious reasons arising from unusual circumstances require that they can not be obtained with these directions." The untrained examiner meets grave difficulties in knowing how to follow the procedure under such circumstances in ways that are legitimate. Very frequently his variations so alter the nature of the response of the child is no longer of any value or interest. Practice in the use of the tests does not decrease this difficulty very much. To do this well involves a judgment that comes only from thorough psychological training and familiarity with laboratory methods.

Another matter concerns the untrained examiner's inability to conform himself in general to the requirements of the procedure. It is not enough merely to ask the child the questions of the test, but to give him the directions to do the things in the tests. The manner in which this is done is often more important than the exact words that are used. The examiner should assume the general attitude most natural to the child he is to examine. He must get down to the child's mental level in

each case. An attitude proper for a four or five-year-old for example, would be fatal to securing the best efforts from an eight or nine-year-old child. The procedure must be carried out in the general spirit of a game in order to arouse the child's interest, but must have enough seriousness in it for the older one not to cause him to regard it as mere play and of no consequence. To arouse and maintain the child's interest in a test to test it is often necessary to interpolate other questions and activities not directly involved in the tests. In certain exceptional cases, such as are found in a certain type of feeble-minded children whose attention is easily attracted and distracted by anything about them, the actual tests to be given often have to be worked in incidentally among other things that are used to lead up to the real tests to which responses are desired. Long continued practice in examining tends to develop a skill in this and in making the general adjustments here in question is, however, largely not a matter of practice or of psychological training. Some people have a natural ability to handle children and to get the best responses of which the children are capable. Others fail to do so, and seem not to improve much in this respect with practice in giving tests. We might say that the temperamentally unfit for examiners, similar to the case of a good student while in college who always remains a poor teacher in spite of knowledge and training. Adequate psychological training and practice in the use of the tests does not, therefore, always assure success as an examiner. Among the small number who have come under my observation there were at least two who never will make successful examiners, however long they continue their practice and training, while there are several who were remarkably successful from the beginning in securing the child's interest and best efforts in the examination.

These general facts come to light in merely watching untrained examiners give the tests. A study of the filled-out blank on which the Faribault teachers recorded the responses of children by "plus" and "minus" signs gives some additional important results. They reveal four very general faults in their work. They are: (1) irregularity of the results, when a child fails in two or more tests in a lower age-group and passes a

age-groups; (2) errors in counting up the mental age of individual plus and minus scorings; (3) not carrying far enough, but stopping with an age-group in which there were only two or three failures out of the five tests; (4) going with too high an age-group and not going back to lower ones, so that the record showed failure in one or more in the lowest age-group tried. In the records of sixteen teachers these faults appeared very frequently, several showing one fault to about every two children examined. The records of a few were entirely free from all of them, each having examined about forty children. With the exception of the first, one may trace the causes of these faults with a fair degree of certainty. Frequent very irregular results in the hands of an examiner in examining normal children simply shows that something has been done wrongly, but we can not trace it to its source. The probability is that the examiner failed to get the responses from the child he was capable of, through leading him in a wrong manner. In this case it is an error of the examiner's failure to assume the proper attitude toward the child. The child's failure to pass the tests that he passes may, on the other hand, be due to the examiner not following the directions for giving them and making them more difficult or to his misinterpreting the child's responses. But these two sources are not so apt to occur in age-groups quite below the child's mental age. They occur most frequently in tests that are already relatively difficult for him. Of the three faults one is at first sight inclined to attribute to carelessness. They are all due to failure to follow simple and clearly stated in every case, and in themselves easy to avoid.

But further analysis makes the fault a less personal one. The third and fourth are doubtless in part due to pressure of time to bear on the examiners to hurry, through lack of time and leisure in which to do the work, and possibly rivalry between the examiners to make rapid progress. The first, however, reduces itself to a failure to understand the importance and seriousness of the errors made, and to appreciate the necessity of proceeding carefully according to rule.

at every point. It is human nature to attempt short-cuts to attain desired ends, in place of going the round-about way outlined by rules and conditions to be learned and followed. The same tendency was repeatedly observed in watching experienced examiners testing children. In the absence of familiar directions some do not hesitate to guess at them and give the tests in altered forms, while others refused to proceed until they understood clearly just how to give them. The same is true of teachers of psychology always experience in making unskilled students follow the methods outlined in laboratory experiments reflects exactly the same thing. The only remedy is extended laboratory training. This alone can teach the individual the need of details in rules and methods, and it is unpermissible not to follow them if accurate results are to be obtained. Without such a foundation, practice merely the use of mental tests will not do much to supply the deficiencies in question.

After this analysis of observations, we are in position to give a more or less definite answer to the question as to the kind of training and how much is necessary to make a good examiner. The question divides itself into several different parts for several quite different things are required of the examiner. The successful examiner must have the following qualifications: (1) Thorough familiarity with all the rules and directions for giving the tests. (2) Familiarity with the nature of responses obtained from children and ability to interpret them correctly in all cases. (3) Ability to alter the procedure in giving a test in legitimate ways when unusual circumstances that demand it. (4) Ability to assume an attitude toward the child under examination that will arouse the child's interest and call forth his best efforts. (5) A proper appreciation of the absolute necessity of adhering strictly to all the rules and directions. We may now ask what kind and how much training of these qualifications calls for and summarize the answers already given above. The first requires practice in the giving of the tests. I am convinced that the average person would have to examine at least fifty children in order to become so familiar with all the details as to enable him to give all the tests

without hesitations in the procedure. The second calls for mainly practice in the use of the tests, but in a few cases for a general knowledge of the principles of psychology and mental development of children. The third calls for a minimum of psychology and laboratory training. No one could be regarded as really qualified to make alterations in the ways of giving mental tests who has not had several years of psychological work, including a minimum of a year of thorough laboratory training. The fourth calls for qualities that largely cannot be acquired by training. Extended practice in the use of tests and in handling children in general will increase an examiner's abilities along this line. The fifth, if not present already, can be acquired only through extended laboratory training and requires a minimum of a year's time.

It remains to point out that these qualifications are of very unequal importance, largely because some are constantly called for, while others are but rarely called for. The first, second, and fifth are constantly called for, and are therefore very important.

The fourth everyone possesses in a certain measure, though some examiners have naturally and do not need to be trained. My experience, however, points to the conclusion that the majority of untrained examiners will fail because of a lack of this qualification. The second is not needed for the majority of the tests, for the majority of the tests are of a nature that the response of the child is at once definite—a "pass" or a "failure." Instances in which a special knowledge is necessary are rare. The third is least needed. If the tests are correctly conducted in every respect, there is not the occasion for altering the way of giving a test very rarely or infrequently. Thus it is seen, in a word, that an examiner without much knowledge of psychology or laboratory training may be able to always get accurate results with the majority of the individual tests, and usually with the whole test; he may have all the other qualifications. But in a large number of cases they fail because of a lack of the fifth, that is, of practice in the use of tests does not supply.

When generating and analyzing the several sources of error in the use of tests and an examiner is subject to, one is apt to over-estimate

their influence on the general result, the mental age of the child as determined by such an examiner. We may, therefore, consider next the question as to the degree or range of error untrained examiners are liable to make in the mental ages obtained. This is not easy to determine with complete accuracy. It may be found approximately by comparing the results of those of a well-trained examiner. The mental ages obtained by the two examiners in examining the same children do not agree for every child to within a few points. Absolute agreement for any two examiners will always be the exception rather than the rule, for no examiner can always get all the contents of an examination under absolutely perfect control. The expert examiner will in the majority of cases not get the same results in examining the same children a second time. The writer has had occasion to compare the results of three hundred feeble-minded children examined by untrained examiners with his own results in examining the same children. These results might be given in statistical form except for the fact that two factors entered to destroy the value of a statistical comparison might otherwise have. These are, first, in most of these three hundred cases I used the 1908 series of tests, while the other examiners used my revision of the tests in nearly all cases. Secondly, the two examinations of a child were rarely made in more or less immediate succession. In many instances there was an interval of as much as two years. In some of these latter cases the child had undoubtedly made considerable mental progress during the interval between the two examinations. On the other hand, it should be stated that none of the examiners were entirely unfamiliar with the tests and uncertainties. All had at least read them through several times and had watched the testing of several children. Under these circumstances the comparison of my own results with those of the other examiners gave, in a word, the following: The maximum difference in the mental ages obtained for any child in the two examinations was two and two-fifths years. A difference of two years occurred several times, and a difference of one year quite a number of times. In the great majority of cases the difference was less than a year. Instances in which the

as two years or more were traced up further in order to find the cause. These were found to be (1) marked improvement in the intelligence of the child since the first examination as shown by the improvement in his school work and a second, third, examination; (2) failure of one examiner to obtain the best responses from the child he was capable of, as shown by a third testing; (3) failure of the examiner to carry the test far enough into the higher age-groups. These facts lead to the general conclusion that the untrained examiner may, because of his unfamiliarity with the tests and lack of psychological training, etc., make an error in the mental age of over two years but that in the majority of cases his result will be accurate within less than a year. They show also that the larger errors he makes are due mainly to his failure to follow the principles of testing which in themselves are easily enough followed and to his inability to so adjust himself to the general procedure in the attitude he takes towards the child as to call forth the child's best efforts in responding. The remedy for these faults has already been discussed.

Comparison of the Average Age With the Average Mental Age.

In considering the scientific aspects of the results obtained by Faribault teachers in examining the Faribault public school children we must always bear in mind that in all questions concerning the accuracy of the revised tests or the grades of intelligence of the children three factors enter. These are errors in mental ages due to inaccuracies in the tests, errors due to the examiner, and variations in the intelligence of the children from the average normal standard. In any given case in which the chronological and mental age of a child do not agree we can not say as to which of these three factors the discrepancy is due, or how the three factors combine to produce the discrepancy. However, we have a fair idea of the limits of the influence of these three factors, and in different results each has been wholly or in part eliminated. It is hoped that a study of the results given below and the comparison with the results obtained under other conditions will make some

contribution to the scientific phases of mental tests and testing.

The question as to the accuracy of the revised tests once into two separate ones. First, the accuracy of the on the whole, the measure of agreement between each logical age and the corresponding average mental age. For instance, is the average mental age of a group of average six-year-old children always just six years, as determined by the tests, and is the same true for each of the other chronological ages? This agreement might be close for one age and not for another. The degree of agreement for each age is the question to decide. It will show whether the scale of tests has any general tendency to measure too low or too high on any point. This agreement, however, might be perfect in the average when averages only are considered and yet the tests might be very inaccurate. For individual children the tests might sometimes measure too high, and sometimes too low, giving more or less frequent errors in the mental age, which errors would cancel each other in the average. The second question concerning the accuracy of the tests is, therefore, that of the frequency and range of error in the case of individual children. The distinction is of the first importance, both from the practical point of usefulness of the results and from the standpoint of perfecting the scale of tests. Where the scale measures too high or too low on the average the amount could be easily subtracted or added to the results in order to obtain the corresponding average mental ages. But in the case of the result for an individual child no correction could be made because it is not known when an error occurs. As regards perfecting the scale of tests, faults of the first kind could be remedied by shifting the tests from one age-group into another, but faults of the second kind might be difficult to overcome.

We will first compare the average chronological ages with the corresponding average mental ages. That is, we will compare the average age of the group of children who are all six years old with their average mental age, and do the same for the other ages up to the age of fifteen, inclusive. In this comparison the influence of errors in the mental ages due to

should be largely eliminated. It might be supposed to be largely eliminated on the basis of the assumption that these errors made in the mental ages are too large sometimes, and too small in such a way as to cancel each other. But it is not safe to make this assumption in entire safety, since there has been a general tendency of one or several examiners to make errors in one direction rather than in another. The irregular variations in the intelligence of the children was present because the children examined were not selected with reference to their normality. In the present comparison, however, the influence is quite negligible. If we assume that as high as 10 per cent., even, of the children were feeble-minded it would be seen in making the computation that their presence would not materially affect the averages that are here considered. We should have, therefore, a fair indication in these results whether the revised scale has a tendency to measure too low or too high at any given point.

The exact chronological ages were not obtained for quite a number of the children examined. After eliminating these there were 230 for which alone the statistical results are considered. The children were first grouped according to their chronological age, putting all who were five years and six months to six years and six months into the six-year group, the same being done for the other year-groups. The average ages of these groups when they come out very closely to exactly six years, seven years, and eight years. In recording the mental ages of the children fractions were used, and since the number of tests in each year-group in the revised scale is five, these fractions are fifths. These fractions were kept in computing averages. The comparison of average age with average mental age shows the ages of six to fifteen years.

TABLE I

6.16	7.03	8.02	9.08	10.01	11.02	12.02	Av.	13.04	13.96	14.92
6.24	7.36	8.28	9.28	10.19	10.88	11.07		11.58	10.89	11.67
+ .78	+ .33	+ .26	+ .20	+ .18	-.14	-.97	.41	-1.46	-3.57	-3.25
38	83	96	91	84	88	75		69	68	38

The comparison shows that the revised scale is still too low at its lower end, and too difficult at its upper end, a fact

that other observers have found to be true of the 1908 scale. It is especially easy for six-year-old children, who lack little over a test to measure a whole year too old. This result may be partly accidental, because of the small number of cases figuring in this age-group. After the six-year scale runs with quite satisfactory accuracy up to the eleven-year mark, where it more or less suddenly measures almost a year behind the chronological age. It will be noticed further that the mental age does not increase materially after the eleven-year mark. This might at first sight be taken to mean that development of intelligence begins to stop at this point. This interpretation, however, would be erroneous. The result is due to the fact that there is only one age-group of tests after the twelve-year mark. Eleven-year-old children have, therefore, less opportunity to pass extra tests beyond the age-group in which they pass all and thereby gain extra credits in mental age. The fault lies mainly in the method of counting up the mental age, which is very admirable for the rest of the scale, but needs to meet this difficulty at this point. That it lies only in the method of measure in the too great difficulty of the individual tests of these upper age-groups is shown by the further fact that nine- and twelve-year-old children pass the eleven- and twelve-year-old tests approximately as frequently as nine- and ten-year-old children pass nine- and ten-year-old tests. Figures on this point will be given later in other connections. With the present method of counting up the mental age, the only remedy lies in the introduction of more tests at this upper end.

By comparing this showing of the revised scale with that obtained by others who used the 1908 scale, some idea may be gained as to whether the revision has made any improvement in the scale with reference to this point. In attempting to make this comparison, however, we meet the fact that no two observers have obtained results under exactly the same conditions, and have stated them in different forms. This makes accurate comparison

*See especially Johnston K. L. An English version of M. Binet's tests for the measurement of intelligence. *Training School Record*, London, 1911. Terman, L. M., and Child, C. A. Tentative revision of the Binet-Simon measuring scale of intelligence. *Journ. Educ. Psychol.*, 1912. Bobertag, O. Ueber Intelligenzpruefungen (nach der Methode von Binet). *Zeitschr. f. angew. Psychol.*, 1911.

possible. On the other hand, it gives a means, though not a means of determining the possible influence of these variations on the statistical results that we wish to compare, but it supplies a means of analysis. The varying conditions that we have to deal are as follows: (1) The children are of different nationalities, requiring the tests to be adapted from the original French to the German and English languages and life. (b) They have been selected according to varying degrees of thoroughness with reference to their selection. Those of Binet and Simon (*), and Bobertag (†), were selected so as to include only children who were in the grade in which they normally belonged. Those of the other authors to be considered here were taken as found in the schools without further selection. (2) The examiners. In the case of Binet-Simon, and Bobertag, the children were examined by one or two well-trained and experienced psychologists. In Goddard's (‡), and in Terman and Childs' (§) results the children were examined by several partly trained examiners. In my own results they were examined by twenty partly-trained examiners. (3) The accuracy of the ages of the children. Goddard dropped fractions of a year in his cases. Binet-Simon's cases were exactly of the ages given. Bobertag's were all within two months of the ages given. Terman and Childs apparently took fractions of a year into account, but do not state how small fractions. In my results given in this paper fractions of a half month were taken into account. (4) The accuracy of the mental ages given of the children. Binet-Simon, Goddard and Bobertag dropped fractions of a year in their mental ages. Terman and Childs, by a special method of correction, counted half years, but usually dropped extra tests if they were not more than one or two. In my results the fractions of a year in the mental ages were counted by fifths, and the revised scale is the value of the individual test. Under these conditions, it so happens that no direct comparison

* *Études de l'intelligence chez les enfants. L'Année Psychologique, 1908.*

† *Quoted above.*

‡ *Children measured by the Binet measuring scale of intelligence. Ped. Sem.,*

§ *Quoted above.*

is possible between any two authors, the conditions have not been all the same for any two. However, the comparisons do show something, both with reference to the relative value of the revised scale in comparison with the 1908 scale and with reference to the influence of the different conditions of testing. Goddard's results for 1,547 non-selected public school children examined with the 1908 scale by five partly trained examiners offers the most direct comparison with my results when re-grouped and when fractions of a year in both the chronological and mental ages are dropped, as Goddard has done. In this comparison the varying condition is in the amount of training the examiners received and the difference in the number of children examined in the two cases. If the number of cases were large enough to average out accidental variations it should make no significant difference, for the purpose of the present comparison, whether fractions of a year were dropped in both the age and mental ages or whether they are taken into account in both. For, in the case of large numbers, the average age would be exactly the same, nearly six and a half, seven and a half years, etc., in place of seven years, etc. Likewise, the average mental age would be four-tenths of a year too small. For, according to the method of counting the mental age, five extra tests passed give a year to the mental age. Zero to four extra tests left are not counted. The average number of extra tests passed and not counted would therefore be two, or two-fifths, equivalent to two-tenths of a year. Thus, as compared with counting fractions of a year in both age and mental age, dropping them in both would make the average mental age always one-tenth of a year too small. Figures to be computed from one of Goddard's tables might be corrected by adding five-tenths of a year to the average chronological age and four-tenths to the average mental age. But no gain would be gained by such modification. In Table II are given my results re-grouped when all fractions in the ages and mental ages are dropped. Table III is derived from one of Goddard's tables and is to be compared with Table II.

TABLE II

6	7	8	9	10	11	12	Av.
6.65	7.44	8.44	9.10	9.98	10.59	10.87	
+ .65	+ .44	+ .44	+ .10	-.02	-.41	-1.13	.47
68	91	88	92	92	76	69	

TABLE III

6	7	8	9	10	11	12	Av.
6.33	7.13	7.74	8.82	9.66	10.12	10.48	
+ .33	+ .13	-.26	-.18	-.34	-.68	-1.52	.52
160	197	209	201	222	166	144	

comparison in these two tables shows the revised very slightly more accurate on the whole than the 1908. The improvement is all for the ages of nine to twelve. For the ages of six to eight the revised scale gives better results. The general average variation of the error from the age is .47 year for the revised scale and .52 for the 1908. But these figures are somewhat too unfavorable to the revised scale. By dropping the fractions of a year from the chronological and mental ages makes the revised scale show up as more accurate than it does in Table I, where these fractions are included. The error introduced in Table II is due to accidental variation in the results from small number of cases. This error should be less in Table III possibly in proportion as the number of cases is larger. The number of cases is approximately twice as many in Table III as in Table II. The varying factor of the error between the two scales also favors the 1908 scale. As was pointed out, twenty practically untrained examiners obtained the results for the revised scale, while five partly trained examiners obtained the results for the 1908. It is, therefore, fair to conclude that this comparison shows an appreciable improvement in the accuracy of the revised scale. A more definite idea of improvement in the accuracy may be gained by carrying the comparative analysis further.

Next consider the results of Terman and Childs in 1906 on 100 unselected public school children with the 1908 examinations being made by the authors, and two other examiners who presumably had some training for the work.

These authors compare the median ages with the mental ages. Table IV is taken from one of their tables.

TABLE IV

Med. Age	6.37	7.5	8.5	9.5	10.5	11.46	12.88
Med. Mental Age	6.5	7.5	8.0	9.0	10.0	10.0	10.5
Difference	+ .13	0	-.5	-.5	-.5	-1.46	-1.88
No. Cases	26	29	43	49	33	44	35

These figures show the same general tendency of difference seen in the preceding table. The general average variance of the mental age from the age is somewhat larger than Childs' results, being .70 year as compared with .52 year. Probably the smaller number of cases in Table IV and the use of the median in place of the average is more responsible for the difference than the other factors. The other varying conditions for the results in Tables III and IV lie in the fact that Childs and Childs took fractions of a year into account in the calculation of the mental ages. Thus, with reference to likeness of conditions as to examiners, Table IV should be compared with Table III. But with reference to likeness of conditions as to counting fractions of a year or not, it should be compared with Table I.

Bobertag in one of his tables gives the results of an examination with the 1908 scale of 180 selected school children, all of whom were within two months of the chronological age given. The children, all of whom were in their proper school grades, were first divided into three classes according to the quality of their school work. One hundred and eighty were chosen from the middle grade in such a way that an equal number belonged to each age, from seven to twelve inclusive. The results for these 180 children are given in Table V, taken from one of his tables.

TABLE V

Age	7	8	9	10	11	12	Av.
Av. Mental Age	7.16	8.43	9.00	9.97	10.65	11.43	
Difference	+ .16	+ .43	0.0	-.03	-.35	-.57	.26
No. Cases	32	28	30	30	32	28	

These results show the closest agreement of ages and mental ages of any, including those for the revised scale in Table I.

... closer for four out of the six ages, from seven to twelve, than in Table I, and has a general average variation of only .27 year, as compared with a general average variation of .33 year for these ages in Table I. Bobertag did not count the fractions in the average ages. When four-tenths of a year are added to his mental ages, a procedure that is, however, hardly permissible for a large number of cases, the agreement between ages and average ages becomes somewhat less, giving a general average variation of .38 year in place of .26 year. When four-tenths are subtracted from the average mental ages in my results in Table I, the figures are more directly comparable with Bobertag's results in Table V, the agreement becomes less for my results, giving a general average variation of .42 year in place of .33 year. Factors present that might produce this favorable result for the scale in Bobertag's results were (1) selection of the children with reference to their normality, and (2) examination of all by one and the same examiner, an experienced psychologist. These figures bring out the fact that the examiner's more accurate selection of the children with reference to normality than is obtained by taking merely public school children, or the two combined at least, are of greater importance in making the average ages and average mental ages equal than is that of the other variable factors with which we have had to deal in comparing results of different authors.

We may bring the main figures of the five tables together in one table for a more ready and final comparison. This is done in Table VI.

TABLE VI

	6	7	8	9	10	11	12	Av.
	+ .78	+ .83	+ .26	+ .20	+ .18	- .14	- .97	.41
	+ .05	+ .44	+ .44	+ .10	- .02	- .41	- 1.13	.47
	+ .33	+ .13	- .26	- .18	- .34	- .88	- 1.52	.52
IV	+ .13	+ .0	- .5	- .5	- .5	- 1.46	- 1.83	.70
		+ .16	+ .43	.0	- .08	- .35	- .57	.26

In this table the ages are given as just six, seven, eight, etc., and, for all authors, the fractions in the case of all authors being added. The other figures give merely the differences between the average ages and average mental ages. It appears

from these combined results that the revision of the scale made a marked improvement in its general accuracy, the improvements being in the upper ages, from ten to twelve inclusive. For the ages of eight and nine the improvement is less, and for the ages of six and seven the revision seems to have made the scale easier where it was already too easy. The favorable showing for the revised scale in the sixth year can perhaps be largely accounted for. It is probably an accidental variation made more possible by the much smaller number of cases, only 38, for this age than I had for the other ages. These children, who were all in the first grade, were brighter than the average, or the particular examiner who examined these children of this grade had a general tendency to get the ages too high. This explanation becomes more plausible when it is noted that in the revision no new tests were introduced into any age-group before the eighth, either from other groups or from the outside as entirely new tests. The tests responsible for giving a mental age of six to seven were, therefore, not made easier in the revision by introducing new tests. Age-group six, however, was made some easier by dropping one of the tests that was found too difficult for this age. In age-group seven, which enters in giving mental ages to seven, one test that was too difficult was pushed forward to age-group eight. It is not likely that this amount of revision is responsible alone for the present difference between the chronological age of six and the corresponding average mental age. However, it is found on further examination that there were a few children in this small group less than six years old who gave an exceptional result. We are left with the general conclusion that the whole scale that the revision has made larger improvements than the figures in these tables indicate directly, and that the presence of children in the group examined who varied considerably from the average normal, and the lack of training of the examiners are responsible for making the agreement between the average ages and the average mental ages poorer than it should be. How much of this is due to the varying factors of abnormality of the children and how much to the examiners

decided. Later analysis, however, will show that the scale played a very large role.

Frequency and Range of Error in the Mental Ages.

If the scale of tests were without exception always as accurate in the examination of each individual child as it is shown to be in the average results, there would be little left to be desired. If this were the case, indeed all that we would need to do to obtain absolute accuracy would be to add to or subtract from the mental age obtained for any child the amount the average results are seen to vary from absolute accuracy. But the matter of the frequency and range of error in the mental ages when considering individual children is an entirely different question. The data so far considered really gives us very little information as to the accuracy of the tests in this respect. For some of the tests would still be of the greatest value if on the whole they gave accurate results but frequently made large errors in individual cases. They would still serve to give us information as to the general status of any large group of individuals. We could compare different schools, or compare juvenile delinquents as a class with normals, or immigrants with normal American-born, etc., and know exactly the general status of the group as a group in each case. We need not dwell on the value and significance of such data if we are satisfied. But this can be obtained now with the Binet-Simon

tests. If we want information about the individual so much more than we do about a group as a group, the importance of the individual question increases in the same measure. Besides, when we know each individual always with whom we are finally dealing, we know each individual we necessarily know the group to which he belongs. The question as to the frequency of error in the tests is also not of the same significance and importance in itself, its frequency would not alone be of the tests much. For no matter in how many instances it occurred, we could always feel certain that it was of no consequence in our dealings with the individual. If the

range of error were large, however, its relative infrequency would not entirely compensate for its large range. For the case in our dealing with the individual on the basis of the results we would occasionally do him great injustice by not knowing when it occurred. The two questions will, however, be considered together, since the same data answer both questions well. In connection with our own results we must note that their value in answering the present question is affected by two of the several factors pointed out above concerning the error in causing discrepancies between the ages and mental ages. These are errors made by the examiners because they were untrained, and variations in the children from the normal intelligence because they were non-selected or unrepresentative. We may again compare our results with those of other studies which used the 1908 scale. If this comparison shows no greater frequency and range of error for the revised scale than the old scale it indicates an improvement in the revision over the old scale; if it shows a greater frequency and range of error no definite conclusion can be drawn.

1. Comparison of the Distribution of the Mental Age with the Normal Distribution Curve. The method most employed to decide the frequency and range of error in the mental age has been to examine a large number of school children and to determine the number that varied in their mental age from their age, and the range of this variation. It has been assumed that the frequency of these variations in the different amounts must have the same character as the normal distribution curve in order to prove that the tests measure accurately without an undue number of exceptions in individual cases. The normal distribution curve, as applied to grades of human intelligence, assumes that the majority of individuals have a middle or average grade of intelligence, and that the number with an intelligence below the average equals the number with an intelligence above the average grade. More definitely, Binet and Simon and others assume that the majority of children tested should test equally at age, and that the number of retarded should equal the number of advanced. We will give the results on the distribution curves, and discuss the validity of the several assumptions.

involved. It will be shown, first, that this method of the accuracy of the tests is entirely too rough and wholly unable to indicate anything more than that the tests give a whole more or less accurate results, and that the degree of accuracy thus proven is less than is generally conceded to the test as a means of detecting the smaller inaccuracies that we wish to know of this method is worthless. It will be shown, secondly, that the assumption as to the majority passing is entirely inadequate as it stands, and that the assumption of equality of the number of retarded and advanced is wrong in the first place. It will be shown, thirdly, that the procedure in getting results and the forms in which the results have been stated heretofore are in themselves inadequate to show the real facts in regard to the distribution curves.

The distribution curve for the scale as a whole. Some have massed the results for all the ages together in complete distribution curve for the mental ages, and simply given the total number of children mentally at age, the number retarded or advanced one, two, etc., years. It is pointed out that results thus treated can not show the accuracy of the tests, but that the figures must be given for each chronological age. This will be discussed at the moment. The results will first be given in this form, bringing out some facts that can best be shown in this form. The following table gives a comparison of my results with other authors.

TABLE VII

No. Cases	% Retarded	% At Age	% Advanced
664	18	65	19
660	31	35	34
142	30	48	22
132	37	42	21
359	25	45	30
161	25	52	23

When the ages of six to twelve alone are considered, the case of Bobertag's results, where the ages are six to twelve, inclusive. It was noted before that the mental ages were low, and in a smaller degree at twelve and eleven. This is apt to be too small because the scale does not ex-

tend beyond the thirteenth year. To include the third year would have erroneously increased the percentage of retarded children. The figures for the different authors are derived from tables they give, except in the case of Bobertag who gives his results in the form used in this table. In the case of my figures, in which 65 per cent. pass at age, fractions of a year in both the age and mental age are taken into account in classifying each child as "retarded," "at age" or "advanced." In the second set of figures in which 35 per cent. pass at age, fractions of a year are dropped in both age and mental age. In making comparisons between figures in this table all the conditions under which the results of the different authors have been obtained and the different ways in which they have been expressed must again be carefully considered. When this is done it is seen at once that counting fractions of a year in both ages and the mental ages is of the greatest importance in settling the present question. In my own results, when all conditions remain exactly the same, the percentage passing at age drops from 65 per cent. to 35 per cent. when these fractions of a year are dropped. The next poorest in the table is 25 per cent. in Goddard's results, which is the only other case in which these fractions were left out of account for both age and mental ages. In Binet-Simon's, Terman and Childs' and Bobertag's results the chronological ages are all more or less accurately taken into account. Terman and Childs count half years in the mental ages in addition. This makes their results rank above those of Goddard in this table, though it has been seen before that in the agreement of average age and mental age they ranked considerably below those of Goddard. Those of Binet-Simon, and of Bobertag rank above those of Terman and Childs because in the latter case the children were not selected with reference to average normality, combining with the fact that there were several examiners, some of whom probably lacked in training. Those of Bobertag, again, rank above those of Binet-Simon because the children were more accurately selected, only those who did average school work and were in the grades in which they belonged, being included by Bobertag for these figures. The first general con-

Table VII, therefore, is that unless fractions of a year in chronological and mental ages are taken into account, and unless children are more carefully selected with reference to normality than has been done the results can be of no great value for the accurate determination of the percentage of children who pass at age, etc., with the tests. With reference to the effect of the revision of the scale, we are left with two alternatives. The results of my figures, in which 35 per cent. pass at age, may be compared best again with those of Goddard, the only difference in the conditions here being in the examiners, as already pointed out. Since the revised tests give the poorer showing, it seems reasonable to attribute it to either the revision or to the examiners. Considering the results of the revised tests in connection with the previous question, the presumption is in favor of the latter, which then means that lack of training of the examiners induces frequent errors in the mental ages of the children examined. The degree of this influence can be seen from the present analysis, since the revised tests themselves give much or only a little less frequent errors in mental ages than the 1908 scale.

The distribution of mental ages for each age. The differences in the percentages passing at age for the different authors are due to the same factors which were found to be responsible for differences in the agreement between the average chronological and the average mental ages. The percentages not passing at age may be the result of the scale measuring too high at the lower end and too low at the upper end. This is without doubt the more correct case. We see, therefore, the necessity of plotting this distribution curve separately for each chronological age. We shall in consequence expect that, since the number of retarded does approximately equal the number of advanced when the results of all ages are considered together, the number of advanced will be larger than the number of retarded for younger ages, and smaller than the number of retarded for older ages. However, the range and frequency of variation in mental age from the age in the individual children may vary almost entirely independently of this connection, as was al-

ready discussed above. We will give next the distribution curve for each age separately.

Terman and Childs, Bobertag, and Stern, have shown that in the results obtained the advanced exceeds the retarded in number for the lower ages, and the retarded exceeds the advanced in number for the upper ages. Our own results are given in Tables VIII to XI. Table VIII gives the number of children for each chronological age who pass at age retarded or advanced one, two, etc., years.

TABLE VIII

	-4	-3	-2	-1	0	+1	+2	+3	+4
6					24	12	1	1	
7				4	61	12	5		
8				9	63	14	8	1	
9			2	10	51	25	3		
10		1	8	7	53	11	2	2	
11			4	14	64	6			
12		1	7	24	42	1			
13		2	21	25	21				

The first vertical column on the left gives the ages of the children, from six to thirteen years. The children six years old, for example, ranged from five years and six months to six years and five months, inclusive, the average age of the group of 38 being 6.16 years, as given in Table I. In determining the amount a child was retarded or advanced the measurements were exact to within half a month, and the mental ages were exact to within a fifth of a year, or 2.4 months. As in the preceding tables, a child is regarded as passing at age if he is retarded or advanced less than a whole year. A child is then regarded as one year retarded if the retardation is one year or over, and so on. In Table IX are given the percentages of children for the figures in Table VIII.

TABLE IX

	% Retarded	% At Age
6	0	63
7	5	74
8	10	66
9	13	56
10	19	63
11	20	73
12	43	56
13	70	30

last two tables show with reference to the revised others have pointed out in regard to the 1908 scale. The number of advanced exceeds the number of retarded for all ages, and the retarded exceeds the advanced for the result of the scale being too easy on one end and too difficult and also too short, on the other end. In this is seen that drawing any conclusion from the results for all ages together. A comparison of these figures with those obtained when given separately for each age, becomes very difficult in attempting to determine the relative accuracy of the present and the 1908 scale, because of the varying conditions which we have to contend. But we may give the following chiefly to bring out more clearly what is required to determine the present question as to the frequency of error in the mental ages obtained with the tests. The amount of variation of the mental age from the age is clearly seen from merely inspecting such figures as are given in Tables VIII and IX. In order to make comparison we will use some index of variation. This may be obtained by dividing for any age the total difference in age and mental age by the total age of the cases considered. Thus, taking the age of ten in Table VIII, for illustration, the total difference between age and mental age is 1 times 3, plus 8 times 2, plus 11 times 1, plus 2 times 2, plus 2 times 3, equals 36. Divided by 10 times 84, equals .056, or 5.6 per cent., which is the index of variation, combining range and frequency into one index. Using this index throughout, we have the following table of indexes of variation, comparing our results with those of others. The indexes given for K^1 are for the results in Table VIII. In these the fractions of a year in age and mental ages are taken into account. The indexes for K^2 are for my results when fractions of a year in mental ages are dropped, the fractions of a

TABLE X

	6	7	8	9	10	11	12	Av.
K^1	7.5	4.5	5.5	5.5	5.6	2.9	4.7	5.2
K^2	8.1	2.8	6.5	6.4	6.4	4.1	8.3	5.4
	11.8	9.4	15.6	13.6	10.1	5.9	9.3	10.8

G.	13.6	7.8	9.1	8.5	7.8	9.5	12.7
T. & C.	7.2	7.8	8.7	5.8	4.6	6.5	10.6
B.-S.	8.9	9.5	4.4	5.6	6.1	4.1	9.9
B.	6.5	7.7	7.1	3.5	6.9	6.7	4.8
A. D.	9.1	8.2	7.3	5.9	6.4	6.7	9.5

year in the ages being kept, thus making the dextes more directly comparable with those of Bine and of Bobertag. The indexes for K^3 are for my results; fractions of a year are dropped in both the ages and the ages, making these indexes more directly comparable with those of Goddard. The accuracy of the indexes for the results of man and Childs probably suffers somewhat because the age was used as the average age in multiplying by the number of cases to get the total age. But they are more accurate than they would have been if the ages had been used as exact years, etc., years, and it makes the results more comparable with those of K^1 . For K^3 , G., B.-S., and B., the ages were taken as 6, 7, etc., years. This is of course inaccurate for K^3 , the exact ages being higher, but it makes the results more comparable with each other. It is accurate for B.-S., and for B. the ages of their children were exactly 6, 7, etc., years. we may compare the indexes of K^1 with those of T. and Childs; of K^2 with those of B.-S., and B.; of K^3 with those of G. in judging the validity of these comparisons, and to more fully detect the factors most responsible for large indexes of variation we may add the following summary of the conditions that are for and against variations.

	Against	Rank	For
K^1	Exact ages Exact mental ages Revised scale	1	Many, and untrained experiments Non-selection of children
K^2	Exact ages Revised scale	2	Many, and untrained experiments Inexact mental ages Non-selected children

Revised scale	7	Many, and untrained examiners Inexact ages Inexact mental ages Non-selected children
Few, partly trained examiners	6	1908 scale Inexact ages Inexact mental ages Non-selected children
Few, partly trained examiners	5	1908 scale Slightly inexact ages Slightly inexact mental ages Non-selected children
Two trained examiners Exact ages Selected children	4	1908 scale Inexact mental ages
One trained examiner Exact ages Carefully selected children	3	1908 scale Inexact mental ages

may use the average indexes of variation given in the vertical column on the right in Table X as a basis for comparing the revised scale with the 1908 scale, and to determine the relative importance of the different factors producing the error. The most striking result is then the fact that the error falls from the first to the seventh or last in rank order in a range of two factors, dropping fractions of a year in the average of the mental ages. From this fact alone it follows that the results can not be considered on this question of accuracy of frequency of error in the mental ages, since he has not taken into account in either age or mental age. In a smaller measure true of the results of Terman

and Childs. Here are combined in the influences of dropping fractions of a year in the mental ages and of the error introduced by the necessity of my using their median ages instead of the average ages. It is noteworthy that Goddard's results and Childs' results rank sixth and fifth, respectively, with reference to range and frequency of error in mental ages. This leaves the comparison of the results for K² with those for B.-S., and B. Cancelling out the factor of "exact ages" which is common to all three, leaves the influence of the "revised scale" for K² against the influence of "expert examiners," "selected children" for B.-S. and B. Since the variations in rank for K² as compared with the third and fourth for B., and B.-S., the conclusion is suggested that the influence of the scale has been more influential in reducing range and frequency of error in the mental ages than have the factors of expert examiners, and the selection of the children with reference to normality. This conclusion is made somewhat more certain by the fact that the revised scale gives careful, detailed directions on how to give each test and how to interpret responses which are lacking in the 1908 scale. This, of course, reduces the importance of the training of the examiner, and leaves the influence of the factor of the selection of the children with reference to normality. This conclusion is, however, probably too favorable to the revision of the scale. Its favorable showing in this comparison is in part due to the fact that the revised scale tends to give too high rather than too low more than does the 1908 scale, as was seen in comparing the average ages with the average mental ages, above. Hence, dropping the fractions in the mental ages as is here done in the results for K² is particularly favorable for the revised scale. In this connection it may be noted that the variations that we are here discussing seem to be largely dependent on and due to the same factor, the difference between the average ages and average mental ages. In the last horizontal column in Table X are given the indexes of variation for all the results of the 1908 scale, including those for K¹, K² and K³. These average indexes of variation decrease at first, reaching their minimum of 5.9 at the ninth year, and then increase again, thus running roughly

the lack of agreement between average ages and average mental ages.

The final conclusion up to this point from this lengthy and detailed analysis, with reference to the improvement the revised scale has made, is simply that the revised scale seems to give more accurate results on the whole, especially at the higher ages, and that it also reduces the frequency of error or both in the mental ages when individuals are considered. No exact idea of the amount of improvement can be gained from the complexity of the conditions which comparisons had to be made. We are now ready to test the validity of the assumptions made that the number of children should pass at age, and that the number of retarded should equal the number of advanced, and the general shape of the distribution curves to show the accuracy of the test and other questions will be taken up in a second

to appear in the next number of this Journal.