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 demonstrating the use of the tests to them by examining a number of feeble-minded children at the Minnesota School for Feeble-Minded. 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 fifteen 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 examining 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, it should be stated that the revised scale used contai...
more detailed statement on how to give each test and how to interpret the responses than the authors of the tests supply. 3, the scientific value of the results in showing further improvements in the scale, and in showing the mental de- v elopment 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 article, advantage will also be taken of the occasion to bring up and consider rather in full the more fundamental prob- lems 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 logical training and some practice in the use of the Bine- mon tests are required of an examiner if he is to obtain the results. There is a similar concensus of opinion that intelligent person without special training or practice will find the tests useful in obtaining a better understanding of a mental development than he can get in any other way, or without very prolonged and close observation. But it is not all clear yet just what kind of training and how much practice must be demanded. Is it a familiarity with 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 suffice, or must it be twenty-five, fifty or several hun- dred? We have also as yet no definite idea as to how large errors we may expect from the untrained examiner. The present writer has had occasion to observe the difficulties that about trained examiners met in giving the tests, each examiner from several to twenty-five children. A study of the scoring by the twenty Faribault teachers and a few others also brings to light some facts about this question. An analysis of observations will give some idea as to just what kind of training is required to make a successful examiner. The initial main difficulty is lack of familiarity with the directions for giving each individual test, which remains until a large number of children have been examined. The de-
tails of these directions in the whole system of tests are too many to be mastered in any other way than through practice in giving the tests. To eliminate this factor entirely, the average person probably needs to examine about fifty children. In the absence of this familiarity the unpracticed examiner has to resort to one of two methods. He either skips the directions and gives the tests usually in considerably modified ways, or he stops to read the directions first before giving each test. Either procedure is detrimental to accurate results. The former is entirely unpermissible because a slight change in the manner of giving a test may sometimes 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 this is 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 examiner's practice, however, the latter is the course to follow. For guessing at the directions and giving the tests wrongly 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 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 untrained can use the classification without error. In my revision of the tests, an effort has been made to give directions for interpreting responses as well as for giving the tests, wherever 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 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 ability to judge responses correctly sometimes implies knowledge of psychological principles or traits in the mental development of children. Where this knowledge is absent, interpreting responses will remain incorrect.

In this connection a third and more serious difficulty arises. Responses 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 objective. It is to ascertain whether the child is capable of the performance involved. The directions for giving a test are devised to best bring out the child's ability in this performance; they fit only the average child under average conditions. A failure to respond at all, or a response that cannot 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 classification of these responses, directions on how to vary the regular procedure in giving a test cannot be made.

My revision gives some such supplementary directions down the general rule to "follow them literally until from obvious reasons arising from unusual circumstances can not be obtained with these directions." The untrained examiner meets grave difficulties in knowing how to vary the procedure under such circumstances in ways that are legitimate. Very frequently his variations so alter the response of the child that it is no longer of any value or significance. 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 in conformity with laboratory methods.

Another matter concerns the untrained examiner's inability in general to meet the requirements of the procedure. It is not enough merely to ask the child the questions of the test and 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 must assume the general attitude most natural to the child he is examining. He must get down to the child's mental level in
each case. An attitude proper for a four or five-year-old child, 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 consequence. To arouse and maintain the child’s interest from test to test it is often necessary to interpolate other questions and activities not directly involved in the tests. In certain extreme cases, such as are found in a certain type of feeble-minded children whose attention is easily attracted and distracted by everything 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, they continue their practice and training, while there are some 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 blanks 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 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
individual plus and minus scorings; (3) not carrying
far enough, but stopping with an age-group in which
were only two or three failures out of the five tests; (4)
by with too high an age-group and not going back to
er ones, so that the record showed failure in one or
in the lowest age-group tried. In the records of sixteen
teachers these faults appeared very frequently, sever-
g one fault to about every two children examined. The
of a few were entirely free from all of them, each hav-
ined about forty children. With the exception of the
one may trace the causes of these faults with a fair
of certainty. Frequent very irregular results in the
an examiner in examining normal children simply shows
thing has been done wrongly, but we can not trace it
ce. The probability is that the examiner failed to get
responses from the child he was capable of, through
ing him in a wrong manner. In this case it is an error
the examiner’s failure to assume the proper attitude
the child. The child’s failure to pass the tests that he
ss may, on the other hand, be due to the examiner not
the directions for giving them and making them more
or to his misinterpreting the child’s responses. But
these two sources are not so apt to occur in age-
below the child’s mental age. They occur most
in tests that are already relatively difficult for him.
three faults one is at first sight inclined to attribute
lessness. They are all due to failure to follow sim-
clearly stated in every case, and in themselves easy
But further analysis makes the fault a less personal
third and fourth are doubtless in part due to pressure
bear on the examiners to hurry, through lack of
and leisure in which to do the work, and possibly
ry between the examiners to make rapid progress.
however, reduces itself to a failure to understand
and seriousness of the errors made, and to ap-
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 lined by rules and conditions to be learned and followed. The same tendency was repeatedly observed in watching examiners testing children. In the absence of familiarity with the directions some do not hesitate to guess at them and alter the tests in altered forms, while others refused to proceed until they understood clearly just how to give them. The teachers of psychology always experience in making undergraduate 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 is unpermissible not to follow them if accurate results are to be obtained. Without such a foundation, practice merely in use of mental tests will not do much to supply the deficiency in question.

After this analysis of observations, we are in position to give a more or less definite answer to the question as to what kind of training and how much is necessary to make a good examiner. The question divides itself into several different ones, 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 variety 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 arise that demand it. (4) Ability to assume an attitude towards 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 of testing. 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 actual use of the tests. I am convinced that the average person will examine at least fifty children in order to become so familiar with all the details as to enable him to give all the test
results of examining public school children

without hesitations in the procedure. The second requires mainly practice in the use of the tests, but in a few cases calls for a general knowledge of the principles of psychology and mental development of children. The third calls for a knowledge of psychology and laboratory training. No one could be really qualified to make alterations in the ways 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 are acquired by training. Extended practice in the use of the tests and in handling children in general will increase an examiner's abilities along this line. The fifth, if not present at all, can be acquired only through extended laboratory training, 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, others are but rarely called for. The first, fourth and fifth are constantly called for, and are therefore very essential. Everyone possesses in a certain measure, and some examiners have naturally and do not need to acquire training. My experience, however, points to the conclusion that the majority of untrained examiners will fail because of lack of this qualification. The second is not needed for the majority of the tests, for the majority of the tests are of such a nature that the response of the child is at once "pass" or a "failure." Instances in which a special psychological knowledge is necessary are rare. The third is least needed. If the tests are correctly conducted in every respect, the occasion for altering the way of giving a test is relatively infrequent. Thus it is seen, in a word, that 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 his he may have all the other qualifications. But in a number of cases they fail because of a lack of the fifth, which mere practice in the use of tests does not supply.

In enumerating and analyzing the several sources of error 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 to which 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 a well-trained examiner. The mental ages obtained by the two examiners in examining the same children 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 conditions 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 about 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 such 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. Some of these latter cases the child had undoubtedly made some 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 until all had at least read them through several times and watched the testing of several children. Under these circumstances the comparison of my own results with those of other examiners gave, in a word, the following: The maximum difference in the mental ages obtained for any child in the 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 im-

provement in the intelligence of the child since the first exami-

nation shown by the improvement in his school work and a

second, third, examination; (2) failure of one examiner to

get the best responses from the child he was capable of, as

by a third testing; (3) failure of the examiner to carry

far enough into the higher age-groups. These facts lead
to the general conclusion that the untrained examiner may,
behind his unfamiliarity with the tests and lack of psycholog-
ing, etc., make an error in the mental age of over two

years, but that in the majority of cases his result will be accur-

ate within less than a year. They show also that the larger

errors he makes are due mainly to his failure to follow the sim-

ple rules of testing which in themselves are easily enough fol-

lowed and to his inability to so adjust himself to the general pro-

ceedure in the attitude he takes towards the child as to call forth

the child's best efforts in responding. The remedy for these faults

already been discussed.

Comparison of the Average Age With the Average Mental

Age.

considering the scientific aspects of the results obtained

by the Faribault teachers in examining the Faribault public

children we must always bear in mind that in all questions

dealing with the accuracy of the revised tests or the grades of in-

tellectual development of the children three factors enter. These are errors

in mental ages due to inaccuracies in the tests, errors due to

examiner, and variations in the intelligence of the children

from an average normal standard. In any given case in which

age and mental age of a child do not agree we can not say

to which of these three factors the disagree-

due, or how the three factors combine to produce

discrepancy. However, we have a fair idea of the limits of

influence of these three factors, and in different results each

has been wholly or in part eliminated. It is hoped that

use of the results given below and the comparison with

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 falls once into two separate ones. First, the accuracy of the tests on the whole, the measure of agreement between each chronological 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 poor 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 a point. This agreement, however, might be perfect in every case when averages only are considered and yet the tests might be very accurate. For individual children the tests might sometimes measure too high, and sometimes too low, giving a certain frequency of 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. This 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 correct average mental ages. But in the case of the result for the individual child no correction could be made because it is never known when an error occurs. As regards perfecting the tests, faults of the first kind could be remedied by shifting 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 eliminated on the basis of the assumption that these made in the mental ages are too large sometimes, and too small in such a way as to cancel each other. But not make this assumption in entire safety, since there been a general tendency of one or several examiners errors in one direction rather than in another. The influence of variations in the intelligence of the children was present the children examined were not selected with refer their normality. In the present comparison, however, influence is quite negligible. If we assume that as high per cent., even, of the children were feeble-minded it been in making the computation that their presence not materially affect the averages that are here consider should have, therefore, a fair indication in these results whether the revised scale has a tendency to measure too low at any given point.

Exact chronological ages were not obtained for quite of the children examined. After eliminating these there 730 for which alone the statistical results are con. The children were first grouped according to their all who were five years and six months to six years months into the six-year group, the same being done other year-groups. The average ages of these groups come out very closely to exactly six years, seven. In recording the mental ages of the children fractions were counted, and since the number of tests in each 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 ages of six to fifteen years.

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Average Age</th>
<th>Average Mental Age</th>
<th>Difference</th>
<th>Number of Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.16</td>
<td>6.16</td>
<td>6.94</td>
<td>+.78</td>
<td>38</td>
</tr>
<tr>
<td>7.03</td>
<td>7.36</td>
<td>7.36</td>
<td>+.33</td>
<td>83</td>
</tr>
<tr>
<td>8.02</td>
<td>8.02</td>
<td>8.28</td>
<td>+.26</td>
<td>96</td>
</tr>
<tr>
<td>9.08</td>
<td>9.08</td>
<td>9.28</td>
<td>+.20</td>
<td>91</td>
</tr>
<tr>
<td>10.01</td>
<td>10.19</td>
<td>10.88</td>
<td>+.79</td>
<td>13.04</td>
</tr>
<tr>
<td>11.02</td>
<td>10.19</td>
<td>11.07</td>
<td>-.14</td>
<td>13.96</td>
</tr>
<tr>
<td>12.02</td>
<td>9.98</td>
<td>11.58</td>
<td>-1.46</td>
<td>14.92</td>
</tr>
<tr>
<td>Av.</td>
<td>10.01</td>
<td>10.19</td>
<td>-3.57</td>
<td>10.89</td>
</tr>
<tr>
<td></td>
<td>10.19</td>
<td>11.07</td>
<td>-3.25</td>
<td>11.67</td>
</tr>
<tr>
<td></td>
<td>+.18</td>
<td>-1.46</td>
<td>-1.46</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-3.57</td>
<td>-3.25</td>
<td>-3.25</td>
<td></td>
</tr>
</tbody>
</table>

This comparison shows that the revised scale is still too easy 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 mentally. This result may be partly accidental, because of the small number of cases figuring in this age-group. After the six-year the scale runs with quite satisfactory accuracy up to the year, where it more or less suddenly measures almost a year behind the chronological age. It will be noticed further that mental age does not increase materially after the eleven year. 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 group. Eleven year and older 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 necessarily meets this difficulty at this point. That it lies only in a measure in the too great difficulty of the individual tests in these upper age-groups is shown by the further fact that and twelve-year-old children pass the eleven and twelve old tests approximately as frequently as nine and ten-year-old tests. Figures on this will be given later in other connections. With the present rule for counting up the mental age, the only remedy lies in the addition 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 be gained as to whether the revision has made any improvement in the scale with reference to this point. In attempting to this comparison, however, we meet the fact that no two authors have obtained results under exactly the same conditions, or stated them in different forms. This makes accurate co*

possible. On the other hand, it gives a means, though rough, of determining the possible influence of these variations on the statistical results that we wish to compare, supplies a means of analysis. The varying conditions with which we have to deal are as follows: (1) The children tested. They have been 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 with varying degrees of thoroughness with reference to their normality. Those of Binet and Simon (*), and Bobertag (†), were so as to include only children who were in the grade in which they normally belonged. Those of the others to be considered here were taken as found in the schools without further selection. (2) The examiner. In the case of Binet-Simon, and Bobertag, the children were examined one or two well-trained and experienced psychologists. In Goddard’s (+), and in Terman and Childs’ ($) results, examined by several partly trained examiners. In my own results they were examined by twenty practically untrained examiners. (3) The accuracy of the ages given by the children. Goddard dropped fractions of a year entirely. 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 fractions of a half month were taken into account. (4) Accuracy of the mental ages given of the children. Binet-Simon and Bobertag dropped fractions of a year in the mental ages. Terman and Childs, by a special method of procedure, counted half years, but usually dropped extra tests if they were not more than one or two. In my results of a year in the mental ages were counted by fifths, the revised scale is the value of the individual test. Under these conditions, it so happens that no direct comparison

---

(*) Le developpement de l'intelligence chez les enfants. L'Année Psychologique, 1908.
(†) Reference quoted above.
($) Two thousand children measured by the Binet measuring scale of intelligence. Ped. Sem., quoted above.
is possible between any two authors, the conditions not been all the same for any two. However, the comparison shows something, both with reference to the relative accuracy of the revised scale in comparison with the 1908 scale, and with reference to the influence of the different conditions. 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 ages and mental ages are dropped, as Goddard has done. In this varying condition is in the amount of training the examiner and the difference in the number of children examined in the two cases. If the number of cases were large enough to remove accidental variations it should make no significance, for the purpose of the present comparison, whether fractions of a year were dropped in both the age and mental age, or whether they are taken into account in both. For, in of large numbers, the average age would be exactly nearly six and a half, seven and a half years, etc., in place of six, seven years, etc. Likewise, the average mental age would be exactly four-tenths of a year too small. For, according to the rule of counting the mental age, five extra tests passed give one year to the mental age. Zero to four extra tests left over are not counted. The average number of extra tests passed not counted would therefore be two, or two-fifths, equivalent to one-tenth of a year. Thus, as compared with counting fractions of a year in both age and mental age, dropping them in both makes 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 age and four-tenths to the average mental age. But little 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.
The comparison in these two tables shows the revised scale very slightly more accurate on the whole than the 1908 scale. The improvement is all for the ages of nine to twelve, for the ages of six to eight the revised scale gives poorer results. The general average variation of the mental age from the age is .47 year for the revised scale and .52 year for the 1908. But these figures are somewhat too unfavorable to the revised scale. By dropping the fractions of a year and mental ages makes the revised scale show up better than it does in Table I, where these fractions are kept. The error introduced in Table II is due to accidental variations resulting from small number of cases. This error should be reduced in Table III possibly in proportion as the number of cases is larger. The number of cases is approximately twice as large in Table III as in Table II. The varying factor of the examiner for the two scales also favors the 1908 scale. As was quoted above, 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 revised scale. A more definite idea of improvement in accuracy may be gained by carrying the comparison further.

Next consider the results of Terman and Childs in examining 396 unselected public school children with the 1908 scale, the examinations being made by the authors, and two assistants 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 | 12.55 |
| Med. Mental Age | 6.5 | 7.5 | 8.0 | 9.0 | 10.0 | 10.0 | 10.0 |
| Difference | +.13 | 0 | -.5 | -.5 | -.5 | -1.46 | -1.83 |
| No. Cases | 26 | 29 | 43 | 49 | 33 | 44 | 35 |

These figures show the same general tendency seen in the preceding table. The general average varies the mental age from the age is somewhat larger than standard's results, being .70 year as compared with .52 year. 

Possibly 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 

and Childs took fractions of a year into account in the mental ages. Thus, with regard to likeness of conditions as to examiners, Table IV should be compared with Table III. But with reference to likeness of 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 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**

<table>
<thead>
<tr>
<th>Age</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>Av.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Av. Mental Age</td>
<td>7.16</td>
<td>8.43</td>
<td>9.00</td>
<td>9.97</td>
<td>10.55</td>
<td>11.43</td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td>+.16</td>
<td>+.43</td>
<td>0.0</td>
<td>-.53</td>
<td>-.35</td>
<td>-.57</td>
<td>.28</td>
</tr>
<tr>
<td>No. Cases</td>
<td>32</td>
<td>29</td>
<td>43</td>
<td>49</td>
<td>33</td>
<td>44</td>
<td>35</td>
</tr>
</tbody>
</table>

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 opposed with a general average variation of .33 year for these ages. Bobertag did not count the fractions in the mental ages. When four-tenths of a year are added to his mental ages, a procedure that is, however, hardly permissible for all number of cases, the agreement between ages and mental 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 V, the agreement becomes less for my results, giving a general average variation of .42 year in place of .33. These present that might produce this favorable result for Bobertag's 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 or the two combined at least, are of greater importance in making the average ages and average mental ages equal to 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**

<table>
<thead>
<tr>
<th>Age</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>Av.</th>
</tr>
</thead>
<tbody>
<tr>
<td>K. - Table I</td>
<td>+.78</td>
<td>+.53</td>
<td>+.38</td>
<td>+.30</td>
<td>-.18</td>
<td>-.14</td>
<td>-.07</td>
<td>.41</td>
</tr>
<tr>
<td>K. - Table II</td>
<td>+.65</td>
<td>+.44</td>
<td>+.44</td>
<td>+.10</td>
<td>-.02</td>
<td>-.41</td>
<td>-.13</td>
<td>.47</td>
</tr>
<tr>
<td>G. - Table III</td>
<td>+.33</td>
<td>+.13</td>
<td>-.26</td>
<td>-.18</td>
<td>-.84</td>
<td>-.88</td>
<td>-1.52</td>
<td>.33</td>
</tr>
<tr>
<td>T &amp; C - Table IV</td>
<td>+.18</td>
<td>+.00</td>
<td>-.50</td>
<td>-.50</td>
<td>-.50</td>
<td>-1.46</td>
<td>-1.83</td>
<td>.70</td>
</tr>
<tr>
<td>B. - Table V</td>
<td>-.16</td>
<td>+.43</td>
<td>-.00</td>
<td>-.08</td>
<td>-.35</td>
<td>-.57</td>
<td>-.57</td>
<td>.25</td>
</tr>
</tbody>
</table>

In this table the ages are given as just six, seven, eight, etc., by all authors, the fractions in the case of all authors being omitted. The other figures give merely the differences in 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 was 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 unfavorable showing for the revised scale in the sixth year perhaps be largely accounted for. It is probably an accident that 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 children of this grade had a general tendency to get the mental ages too high. This explanation becomes more plausible if 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 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 of six to seven, one test that was too difficult was pushed forward in age-group eight. It is not likely that this amount of revision is responsible alone for the present difference between the age of six and the corresponding average mental age. Moreover, it is found on further examination that there were children in this small group less than six years old who advanced in their mental ages by two years or more, 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 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 fact of normality of the children and how much to the examiners.
RESULTS OF EXAMINING PUBLIC SCHOOL CHILDREN

169

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

Frequency and Range of Error in the Mental Ages.

The scale of tests were without exception always as ac-
the examination of each individual child as it is shown
the average results, there would be little left to be de-
If this were the case, indeed all that we would need to do
an absolute accuracy would be to add to or subtract from
mental age obtained for any child the amount the average
are seen to vary from absolute accuracy. But the matter
frequency and range of error in the mental ages when
considering individual children is an entirely different

The data so far considered really gives us very little
the accuracy of the tests in this respect. For some
the tests would still be of the greatest value if on the
they gave accurate results but frequently made large
individual cases. They would still serve to give us
information as to the general status of any large group
nals. We could compare different schools, or com-
delinquents as a class with normals, or immigrants
with normal American-born, etc., and know exactly
status of the group as a group in each case. We
dwell on the value and significance of such data if
But this can be obtained now with the Binet-Simon

we want information about the individual so much
we do about a group as a group, the importance
question increases in the same measure. Besides,
individual always with whom we are finally dealing,
how each individual we necessarily know the group
belongs. The question as to the frequency of er-
question as to the range of error in the mental ages
the tests are also not of the same significance and
If the range of error were small enough to have
importance in itself, its frequency would not alone
be of the tests much. For no matter in how many
occurred, we could always feel certain that it was
ence 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 case in our dealing with the individual on the basis of results we would occasionally do him great injustice in knowing when it occurred. The two questions will, however, be considered together, since the same data answer both 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 entering 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 on this point. We may again compare our results with those of others who used the 1908 scale. If this comparison shows no greater frequency and range of error for the revised 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 Ages with the Normal Distribution Curve. The method most employed to decide the frequency and range of error in the mental ages has been to examine a large number of school children and to find 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 the majority of individuals have a middle or average grade of intelligence, and that the number with an intelligence below 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 out mentally at age, and that the number with retardation 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 testing the accuracy of the tests is entirely too rough and wholly inadequate to indicate anything more than that the tests give more or less accurate results, and that the degree of accuracy thus proven is less than is generally conceded to the method as a means of detecting the smaller inaccuracies that wish to know of this method is worthless. It will be secondly, that the assumption as to the majority passing entirely inadequate as it stands, and that the assumption as to 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 computing the distribution curve for the mental ages, and simply determined the total number of children mentally at age, the total number retarded or advanced one, two, etc., years. It pointed out that results thus treated can not show the real accuracy of the tests, but that the figures must be given for each chronological age. This will be discussed in a moment. The results will first be given in this form, bringing out some facts that can best be shown in this following table gives a comparison of my results with other authors.

### TABLE VII

<table>
<thead>
<tr>
<th>No. Cases</th>
<th>% Retarded</th>
<th>% At Age</th>
<th>% Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>154</td>
<td>15</td>
<td>65</td>
<td>19</td>
</tr>
<tr>
<td>450</td>
<td>31</td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td>142</td>
<td>30</td>
<td>48</td>
<td>23</td>
</tr>
<tr>
<td>322</td>
<td>37</td>
<td>42</td>
<td>21</td>
</tr>
<tr>
<td>252</td>
<td>25</td>
<td>45</td>
<td>30</td>
</tr>
<tr>
<td>161</td>
<td>25</td>
<td>53</td>
<td>23</td>
</tr>
</tbody>
</table>

Table the ages of six to twelve alone are considered, except in the case of Bobertag's results, where the ages are inclusive. It was noted before that the mental age at thirteen, and in a smaller degree at twelve and eleven, is apt to be too small because the scale does not ex-
tend beyond the thirteenth year. To include the thirteenth year would have erroneously increased the percentage of retarded children. The figures for the different authors arrived from tables they give, except in the case of Bobertag who gives his results in the form used in this table. In the first set 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 fractions of a year are dropped in both age and mental age. Making comparisons between figures in this table all the varying conditions under which the results of the different authors have been obtained and the different ways in which they have 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 deciding the present question. In my own results, when all other 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 Goddard's results, which is the only other case in which these fractions were left out of account for both ages 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 considered half years in the mental ages in addition. This makes their results rank above those of Goddard in this table, though it was seen before that in the agreement of average age and average mental age they ranked considerably below those of Goddard. Those of Binet-Simon, and of Bobertag rank above those of Terman and Childs because the children were not selected with reference to average normality, combining the fact that there were several examiners, some of whom probably lacked in training. Those of Bobertag, again, are above those of Binet-Simon because the children were accurately selected, only those who did average school work and were in the grades in which they belonged, being by Bobertag for these figures. The first general con
the VII, therefore, is that unless fractions of a year in and mental ages are taken into account, and unless are more carefully selected with reference to nor-
has been done the results can be of no great value
urate determination of the percentage of children who, etc., with the tests. With reference to the effect of the the scale, we are left with two alternatives. The of my figures, in which 35 per cent. pass at age, may best again with those of Goddard, the only differ-
conditions here being in the examiners, as already the revised tests give the poorer showing, it seems ve to attribute it to either the revision or to the exam-
considering the results of the revised tests in connection previous question, the presumption is in favor of the which then means that lack of training of the exam-
ences frequent errors in the mental ages of the in-
examined. The degree of this influence can
from the present analysis, since the revised tests themselves give much or only a little less frequent errors of ages than the 1908 scale.
tribution of mental ages for each age. The differ-
percentages passing at age for the different authors due to the same factors which were found to be re-
differences in the agreement between the average average mental ages. The percentages not passing at be the result of the scale measuring too high at the and too low at the upper end. This is without doubt the case. We see, therefore, the necessity of this distribution curve separately for each chrono-
We shall in consequence expect that, since the retarded does approximately equal the number of ad-
the results of all ages are considered together, the advanced will be larger than the number of retarded ages, and smaller than the number of retarded for ages. However, the range and frequency of variation al age from the age in the individual children may mostly 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 noted 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**

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Retarded</th>
<th>At Age 0</th>
<th>At Age +1</th>
<th>At Age +2</th>
<th>At Age +3</th>
<th>At Age +4</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>12</td>
<td>1</td>
<td>5</td>
<td>8</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>51</td>
<td>12</td>
<td>5</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>20</td>
<td>10</td>
<td>51</td>
<td>30</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
<td>8</td>
<td>7</td>
<td>53</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>14</td>
<td>7</td>
<td>53</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>50</td>
<td>24</td>
<td>43</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>21</td>
<td>25</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>22</td>
<td>25</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The first vertical column on the left gives the approximate ages of the children, from six to thirteen years. The six years old, for example, ranged from five years and six months to six years and five months, inclusive, the average age of this group of 38 being 6.16 years, as given in Table I. In computing the amount a child was retarded or advanced the ages used were exact to within half a month, and the mental ages 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 counted as one year retarded if the retardation is one year or over and less than two years, and so on. In Table IX are given the percentages for the figures in Table VIII.

**TABLE IX**

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>% Retarded</th>
<th>% At Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0</td>
<td>63</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>76</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>60</td>
</tr>
<tr>
<td>9</td>
<td>13</td>
<td>54</td>
</tr>
<tr>
<td>10</td>
<td>19</td>
<td>55</td>
</tr>
<tr>
<td>11</td>
<td>20</td>
<td>73</td>
</tr>
<tr>
<td>12</td>
<td>43</td>
<td>56</td>
</tr>
<tr>
<td>13</td>
<td>70</td>
<td>30</td>
</tr>
</tbody>
</table>
last two tables show with reference to the revised others have pointed out in regard to the 1908 scale. number of advanced exceeds the number of retard d for ages, and the retarded exceeds the advanced for the result of the scale being too easy on one end and too it also too short, on the other end. In this is seen drawing any conclusion from the results for all ages together. A comparison of these figures with those when given separately for each age, becomes very in attempting to determine the relative accuracy d and the 1908 scale, because of the varying condi-which we have to contend. But we may give the y chiefly to bring out more clearly what is required determine the present question as to the frequency of error in the mental ages obtained with the tests. amount of variation of the mental age from the age is seen from merely inspecting such figures as are tables VIII and IX. In order to make comparison ill use some index of variation. This may be ob- tained for any age the total difference in age and by the total age of the cases considered. Thus, tak- of ten in Table VIII, for illustration, the total dif- ference and mental age is 1 times 3, plus 8 times 2, plus mu 11 times 1, plus 2 times 2, plus 2 times 3, equals 47, to be divided by 10 times 84, equals .056, or 5.6 per cent., index of variation, combining range and frequency into one index. Using this index throughout, we ob taining Table of indexes of variation, comparing our those of others. The indexes given for K1 are for in Table VIII. In these the fractions of a year in mental ages are taken into account. The indexes K2 are for my results when fractions of a mental ages are dropped, the fractions of a

<table>
<thead>
<tr>
<th>TABLE X</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
</tr>
<tr>
<td>2.5</td>
</tr>
<tr>
<td>2.1</td>
</tr>
<tr>
<td>11.8</td>
</tr>
</tbody>
</table>
year in the ages being kept, thus making the indexes more directly comparable with those of Binet and of Bobertag. The indexes for K³ 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 man and Childs probably suffers somewhat because the age was used as the average age in multiplying by the cases to get the total age. But they are more accurate; they would have been if the ages had been used as exact, etc., years, and it makes the results more comparable with of K³. For K³, G., B.-S., and B., the ages were taken as 6, 7, etc., years. This is of course inaccurate for K³, the exact ages being higher, but it makes the results comparable with each other. It is accurate for B.-S., and the ages of their children were exactly 6, 7, etc., years. we may compare the indexes of K¹ with those of T. and K² with those of B.-S., and B.; of K³ with those of G. in judging the validity of these comparisons, and to more detect the factors most responsible for large indexes of var we may add the following summary of the conditions that for and against variations.

<table>
<thead>
<tr>
<th></th>
<th>Against</th>
<th>Rank</th>
<th>For</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>K¹</strong></td>
<td>Exact ages</td>
<td>1</td>
<td>Many, and untrained examiners</td>
</tr>
<tr>
<td></td>
<td>Exact mental ages</td>
<td></td>
<td>Non-selection of children</td>
</tr>
<tr>
<td></td>
<td>Revised scale</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>K³</strong></td>
<td>Exact ages</td>
<td>2</td>
<td>Many, and untrained examiners</td>
</tr>
<tr>
<td></td>
<td>Revised scale</td>
<td></td>
<td>Inexact mental ages</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Non-selected children</td>
</tr>
</tbody>
</table>
Results of Examining Public School 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 4 1908 scale
                       Exact ages
                       Selected children

One trained examiner 3 1908 scale
                     Exact ages
                     Carefully selected children

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

The same is 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 of the average ages. It is noteworthy that Goddard's 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 $K^2$ with B.-S., and B. Cancelling out the factor of "exact age" is common to all three, leaves the influence of the scale for $K^2$ against the influence of "expert examiners," selected children" for B.-S. and B. Since the variations are rank for $K^2$ as compared with the third and fourth for B., and B.-S., the conclusion is suggested that the scale has been more influential in reducing range of error in the mental ages than have the factors of examiners, and the selection of the children with reference to normality. This conclusion is made somewhat more probable 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 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 measure too high rather than too low more than does the 1908. This was seen in comparing the average ages with the average mental ages, above. Hence, dropping the fractions in the results for $K^2$ 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 giving 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, those for $K^1$, $K^2$ and $K^3$. These average indexes of variation decrease at first, reaching their minimum of 5.9 in the ninth year, and then increase again, thus running roughly...
lack of agreement between average ages and averages.

The final conclusion up to this point from this lengthy analysis, with reference to the improvement the scale has made, is simply that the revised scale gives more accurate results on the whole, especially for 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 under which comparisons had to be made. We are now ready to consider the validity of the assumptions made that the majority of children should pass at age, and that the number of retarded equal the number of advanced, and the general usefulness of the distribution curves to show the accuracy of the test and other questions will be taken up in a second article.