# THE STANFORD REVISION OF THE BINET-SIMON SCALE AND SOME RESULTS FROM ITS APPLICATION TO 1000 NON-SELECTED CHILDREN

LEWIS M. TERMAN, GRACE LYMAN, DR. GEORGE ORDAHL, DR. LOUISE ORDAHL, NEVA GALBREATH, AND WILFORD TALBERT[1]

*Stanford University*

## Sources of Data for the Revision

The purpose of this article is to summarize briefly an investigation undertaken for the improvement of the Binet scale, and to suggest some possible contributions of intelligence testing to the psychology of mental development.

The work has extended over a period of five years and has involved the examination of 1700 normal children, 200 defective and superior children, and more than 400 adults. The present discussion will confine itself, however, to an account of tests carried out with 1000 non-selected children.

Tests of 400 children had been made by Childs and Terman in 1910–1911, and of 300 children by Trost, Waddle and Terman in 1911–12. For various reasons, however, the results of these tests did not furnish satisfactory data for a thoroughgoing revision of the scale. Accordingly a new investigation was undertaken, somewhat more extensive than the others and more carefully planned. Its main features may be described as follows:

1. We assembled as nearly as possible all the results which had been secured for each test of the scale by all the workers of all countries. The result was a large sheet of tabulated data for each individual test, including per cents passing the test at various ages, conditions under which the results were secured, method of procedure, etc. After a comparative study of these data, and in the light of results we had ourselves secured, a provisional arrangement of the tests was prepared for try-out.

[1] The tests of 1000 children were made by Miss Lyman, Dr. and Mrs. Ordahl, Miss Galbreath and Mr. Talbert. Terman is responsible for the test series and for the analysis of results. The Stanford Revision will appear shortly as a joint monograph.

2. A plan was then devised for securing subjects who should be as nearly as possible representative of the several ages. The method was to select a school in a community of average social status, a school attended by all or practically all the children in the district where it was located. In order to get clear pictures of age differences the tests were confined to children who were within two months of a birthday. To avoid accidental selection, *all* the children of even-age were tested in whatever grade enrolled. Tests of foreign-born children, however, were eliminated in the treatment of results. There remained tests of approximately 1000 children, of whom 905 were between 5 and 14 years of age.

3. The children's responses were recorded verbatim. This made it possible to re-score the records according to any desired standard and thus to fit a test more perfectly to the age level assigned it.

4. Much attention was given to securing uniformity of procedure. A half year was devoted to training the examiners and another half year to the supervision of the testing. In the further interests of uniformity all the records were scored by one person (Terman).

In working out a revision of the scale based upon these data the guiding principle was to secure an arrangement of the tests and a standard of scoring which would cause the median mental age of the children of each age group to coincide with the median chronological age. If the median mental age at any point in the scale was too high or too low it was only necessary to change the location of certain of the tests, or to change the standard of scoring, until an order of arrangement and a standard of passing were found which would throw the median mental age where it belonged. We had already become convinced, for reasons too involved for presentation here, that no satisfactory revision of the Binet scale was possible on any theoretical considerations as to the per cent. of passes which an individual test ought to show in a given year in order to be considered standard for that year. Such a plan would doubtless be feasible or even preferable with a scale differently founded, but not with that of Binet.

As was to be expected, the first draft of the revision did not prove satisfactory. The scale was still too hard at some points and too easy at others. In fact, three successive revisions were

necessary, involving three separate scorings of the data and as many tabulations of the mental ages, before the desired degree of accuracy was secured.

As finally left, the scale gives a median intelligent quotient closely approximating 100 for our non-selected children of each age from 4 to 14. The revision contains six regular tests and from one to three alternate tests in each year from 3 to 10, eight tests at year 12, six at 14, and six in each of two higher groups which are named, in order, "average adult" and "superior adult."

The tests in the two highest groups were standardized chiefly on the basis of results from 400 adults.[2] The extension of the scale in the upper range is such that ordinarily intelligent adults, little educated, test up to what is called the "average adult" level. Adults whose intelligence is known from other sources to be superior are found to test well up toward the "superior adult" level, and this holds true whether the subjects in question are well educated or practically unschooled.

Enough new tests were included in the trial series to permit the elimination of tests which failed to meet the requirements and still to give the revision some 27 tests more than were in the original scale.

The following method was employed for determining the validity of a test. The children of each age level were divided into three groups according to intelligence quotient, those testing below 90, those between 90 and 109, and those with an intelligence quotient of 110 or above. The percents of passes on each individual test at or near that age level were then ascertained separately for these three groups. If a test fails to show a decidedly higher proportion of passes in the superior I Q group than in the inferior I Q group, it can not be regarded a satisfactory test of intelligence. On the other hand, a test which satisfies this criterion must be accepted as valid or the entire scale must be rejected. Henceforth it stands or falls with the scale as a whole.

When tried out by this method, some of the tests which have been most criticized showed a high degree of reliability; others

[2] Including 30 business men tested by H. E. Knollin and Richard Zeidler, 40 high school students tested by Terman, 150 "migrating unemployed" tested by H. E. Knollin, and 150 delinquents tested by Mr. J. H. Williams.

which have been most praised proved to be so little correlated with intelligence that they had to be discarded.

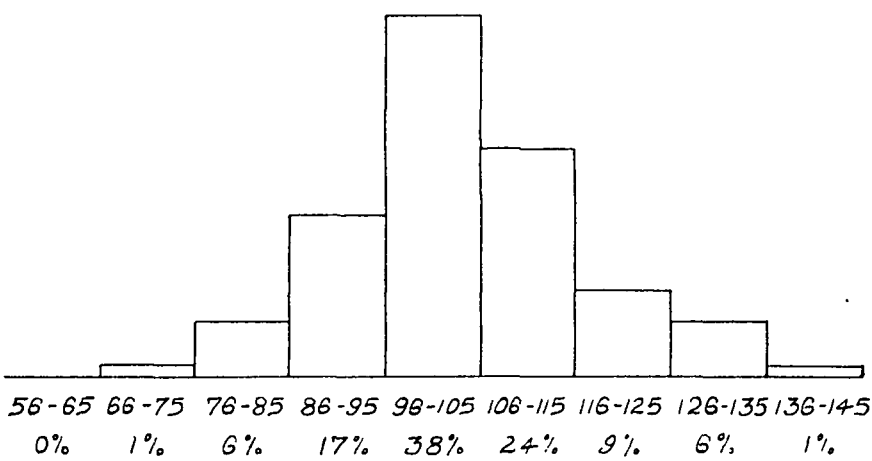
Next a brief summary of some results from the analysis of the data.

## The Distribution of Intelligence

Because of failure to avoid the influence of accidental selection in choosing subjects for testing, most of the Binet studies have thrown little light on the distribution of intelligence. The method of securing subjects for the present study makes our results on this point especially interesting.

The intelligence quotients were calculated for the 1000 children and their distribution was plotted for the ages separately. The distribution was found fairly symmetrical at each age from 5 to 14. At 15 the range is on either side of 90 as a median, and at 16 on either side of 80 as a median, which is of course due to the fact that these left-over retardates are usually below normal in intelligence.

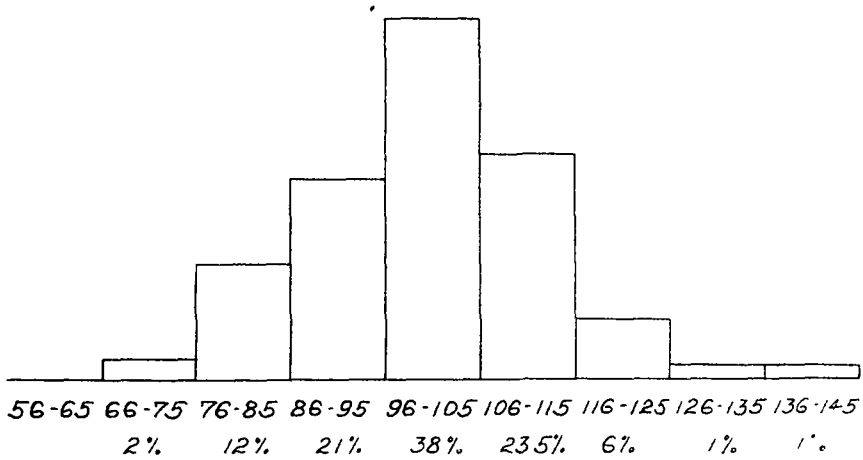
The I Q's were then grouped in ranges of ten. In the middle group were thrown those from 96–105; the ascending groups including in order the I Q's from 106–115, 116–125, etc. Correspondingly with the descending groups. Graphs 1–3 show the results of this grouping for the ages 6, 9, and 13 separately. Graph 4 shows the distribution for the ages 5–14 combined. Graph 5 shows also the distribution for all ages combined, but this time with the I Q's grouped by ranges of 20 points instead of 10 (50–70, 71–90, 91–110, 111–130, 131–150).



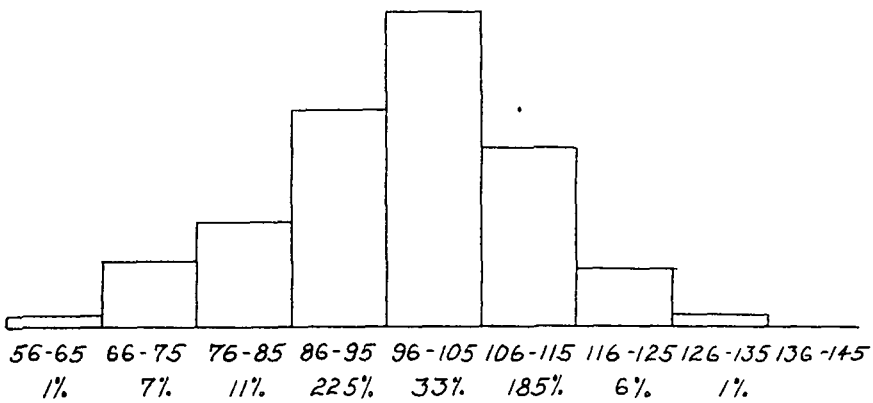

GRAPH 1. Showing distribution of I Q's of 117 non-selected 6-year-olds.

The distributions for the ages combined are seen to be remarkably symmetrical. The symmetry for the separate ages is hardly less marked, considering the fact that only 80–120 children were tested at each age. In fact, the range including the middle 50% of I Q's remains practically constant from 5 to 14 years.

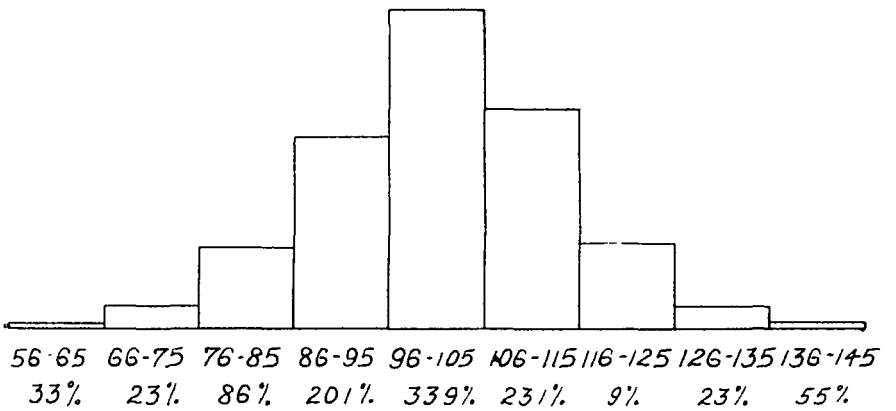
The traditional view that variability in mental traits greatly increases at or near the onset of adolescence is here contradicted as far as intelligence is concerned. Likewise the common opinion that deviations below normal are more frequent and more extensive than deviations above normal. It is found, for example, that for the ages 5–14 combined:



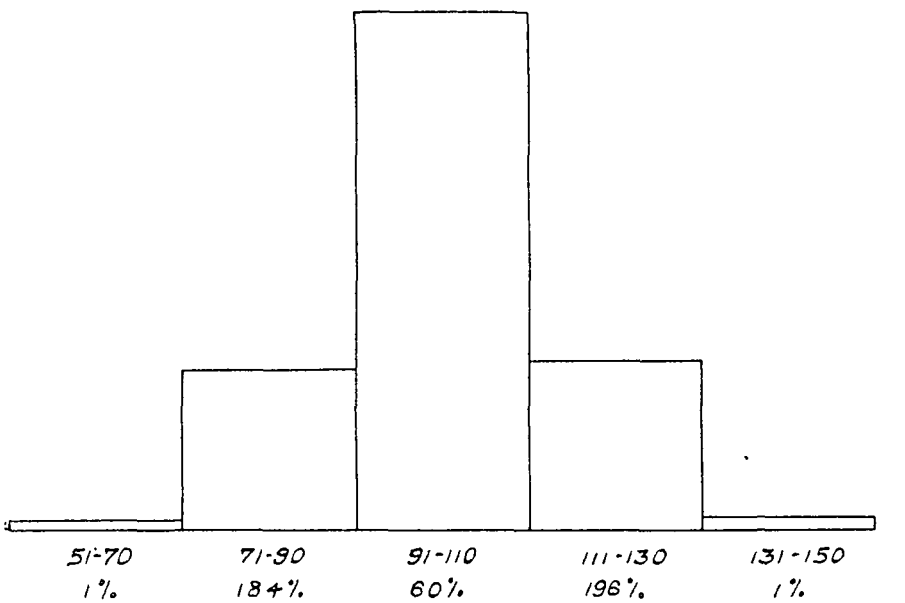

GRAPH 2. Showing distribution of I Q's of 113 non-selected 9-year-olds.




GRAPH 3. Showing distribution of I Q's of 98 non-selected 13-year-olds.




GRAPH 4. Showing distribution of I Q s of 905 non-selected children, ages 5--14 combined.




GRAPH 5. Showing distribution of I Q's of 905 non-selected children ages 5 -- 14 combined.

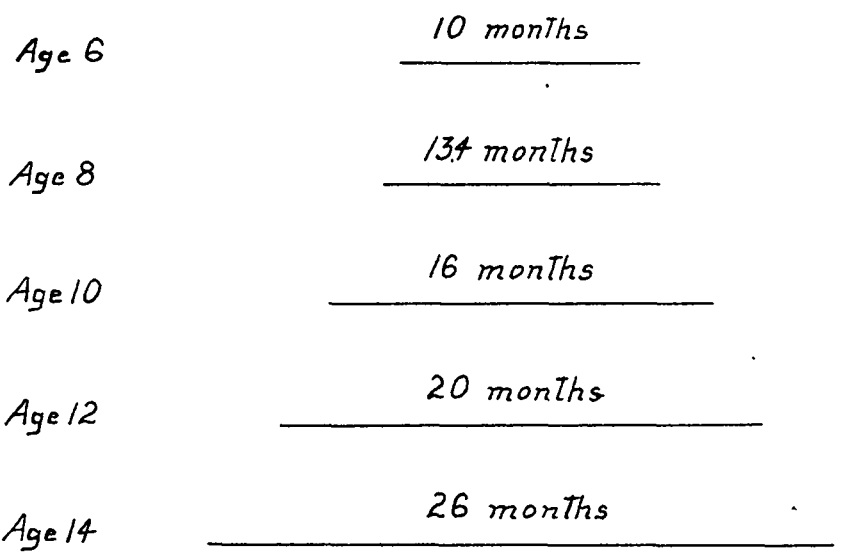
| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| the lowest | 1% | go to | 70 | or below; | the highest | 1% | to | 130 | or above |
| " " | 3% | " " | 76 | " " | " " | 3% | " | 125 | " " |
| " " | 10% | " " | 85 | " " | " " | 10% | " | 116 | " " |
| " " | 20% | " " | 91 | " " | " " | 20% | " | 110 | " " |

The nature of the distribution of I Q's emphasizes the hopelessness of our ever arriving at any universally acceptable definition of feeble-mindedness. The dividing line can be only arbitrary, exactly as would be the case if we attempted to classify all men into the two groups: "normally tall" and "abnormally short."

## The Validity of the I Q

Since the distribution of I Q's remains practically constant at all ages from 5–14, it is evident that the usual method of expressing intelligence status in terms of years of retardation or acceleration is misleading. "A year of retardation" is a unit which has no constant value. A year of deviation at age 6 is exactly equivalent to a deviation of 18 months at age 9, and to 2 years at age 12, etc. This follows necessarily from the uniform nature of the distribution of I Q's in the different ages. The number of 6 year old children testing "at age" is approximately twice as great as the number of 12 year olds testing at age, and 50% greater than in the case of the 9 year olds. The range of months including the middle 50% of the mental ages increases at a fairly constant ratio from 6 to 14, as shown in Graph 6. These facts argue strongly in favor of the validity of the I Q as a means of method of expressing intelligence status.

The crucial experiment, of course, would be to test the same children several times during the period of mental growth in order to find whether the I Q remains constant. Unfortunately, the repeated tests which others have made for this purpose have involved the use of a very imperfect scale, the worst feature of which is that its inaccuracy is unequal at different age levels. We have found with the Binet scale of 1908 that an I Q of 110 at age 6 is no more than equivalent to an I Q of 100 at 9, or to 90 at 12. With a scale equally accurate at the different levels we are now in position to follow the actual development of children from year to year.




GRAPH 6. Showing range of months including the middle 50% of mental ages at various years.

Fortunately 18 children who were tested by Childs and Terman in 1911 were tested some two to four years later by the Stanford revision. When the I Q's resulting from the first test were corrected to accord with the shifting of tests in the revised scale it was found that no I Q of the second test differed more than 8 points from the corresponding I Q of the first test. The median difference was only 4. The superior children of the first test are found superior in the second, the average remain average, the inferior remain inferior, the feeble-minded remain feeble-minded, and always in approximately the same degree. Such facts give little support to the traditional belief that intelligence normally develops by alternate leaps and rests; that men of genius develop out of blockheads, and that genius children usually degenerate intellectually as maturity approaches.

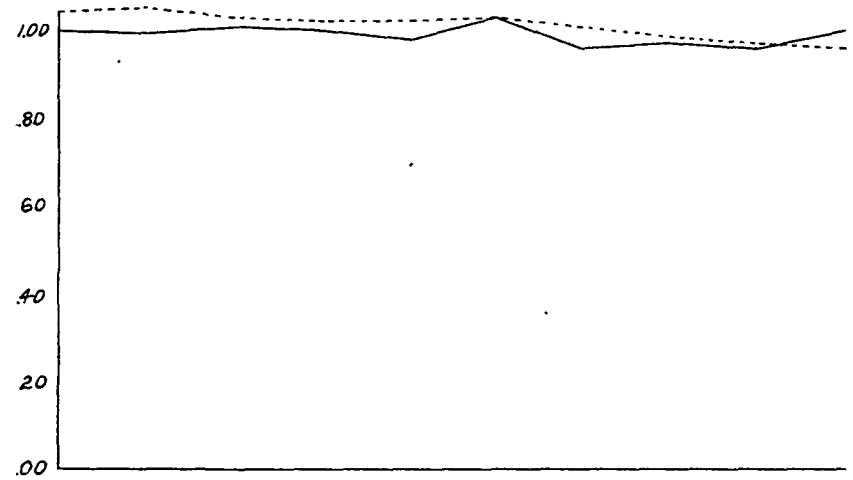
Incidentally it may be noted that if the I Q is a valid expression of intelligence, as it seems to be, then the Binet-Simon "age grade method" becomes transformed automatically into a "point scale method," if one wants to use it that way. As such it

would seem to be greatly superior to the Yerkes-Bridges scale, for it includes a much larger number of tests and its points have definite meaning and equal value.

If future investigations should confirm the validity of the I Q and its necessary corollaries, the practical consequences would be of the greatest importance, for accurate prediction of a child's later development would then become a matter of everyday practice.

## Sex Differences

When the I Q's of the boys and girls are treated separately we find a small but fairly constant superiority of the girls up to age 13. At 14 the median I Q for the girls drops suddenly below that for boys. This is shown in graph 7.




| | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|
| Boys | 1.00 | .99 | 1.01 | 1.00 | .98 | 1.03 | .96 | .97 | .96 | 1.00 |
| Girls | 1.04 | 1.05 | 1.03 | 1.02 | 1.02 | 1.03 | 1.01 | .99 | .97 | .96 |

GRAPH 7. Showing median I Q for boys and girls separately at each age. (Total number of boys 457; total number of girls 448)

Apart from the slight superiority of the girls, the distribution of intelligence in the two sexes is not different. The supposed wider variation of boys is not found. The girls do not group themselves about the median more closely than the boys. The

range including the middle 50% is approximately the same for the two sexes.

The supplementary data, including the teachers' estimates of intelligence on a scale of five, the teachers' judgments in regard to the quality of the school work, and records showing the age-grade distribution of the sexes, were all sifted for evidence as to the genuiness of the apparent superiority of the girls age for age. The results of all these lines of inquiry support the tests in suggesting that the superiority of the girls is probably real even up to and including age 14, the apparent superiority of the boys at this age being fully accounted for by the more frequent elimination of 14 year old girls from the grades by promotion to the high school.

However, sex differences in intelligence are so small (amounting on an average to only 2 to 4 per cent. in terms of I Q) that for practical purposes they would seem negligible. We find no reason to share the opinion expressed by Yerkes and Bridges "that at certain ages serious injustice will be done individuals by evaluating their scores in the light of norms which do not take account of sex differences."

## The Relation of Intelligence to Social Status

In about half our schools it was possible to obtain a classification of the children according to social status. This classification was made by the teachers on a scale of five, "very inferior," "inferior," "average," "superior," and "very superior." The median I Q of the "inferior" group is 93, that of the "superior" group 107; a difference of 14 points. At the age of 7 years this amounts to one year in mental age; at 14 to two years in mental age. By the Pearson method the correlation between intelligence and social status is .40, a result which is fully in harmony with the earlier findings of Binet workers.

The usual assumption has been that such correlation is the artificial product of environmental influences; that the child from a superior home does better because he has had more opportunity to pick up the information needed for success in the tests. A careful sifting of the data has forced upon us the conclusion that the greater part of the difference found is due to an actual average superiority in the endowment of better-class children.

The results of the tests on this point were confirmed by five separate lines of inquiry: (1) The amount of correlation between I Q and social status for children of different ages; (2) a comparison of social status with the teachers' estimates of intelligence; (3) a similar comparison with the teachers' judgments as to the quality of the school work; (4) a comparison of the age-grade status of the children of different social classes; and (5) case studies of exceptionally intelligent and dull children in the same family. The correlation of social status with the teachers' estimates of intelligence was .55, and with the quality of school work .47. The correlation between I Q and social status was .43 for the younger children, .40 for those in the middle years, and only .29 for the older. *In other words, the longer the supposed influence of home environment lasts the more independent of it the I Q becomes.*

These facts and others too involved for presentation here point to the conclusion that the correlation of I Q with social status rests upon actual differences in endowment. In the light of the data available we are unable to agree with the contention of Meumann and of Yerkes that it is unfair to judge the intelligence of any child except in terms of the average intelligence of his own social class. It would seem to us just as logical to insist that it is unfair to the dull or feeble-minded child to judge his intelligence with reference to standard intelligence for normals.

### The Relation of Intelligence to School Success

Three questions are taken up in this connection: (1) The correlation of I Q with the quality of school work as judged by the teacher; (2) The relation of mental age to grade progress; and (3) The correlation of I Q with the teachers' estimates of intelligence. The correlation with school work is .45, that with teachers' estimates of intelligence .48. Between mental age and grade there are many disagreements, most of which are traced to the tendency of the school to promote children by age rather than by ability.

While these correlations are high enough to confirm in a general way the validity of the Binet method, serious disagreements were found in a considerable number of individual cases. For some of the comparisons about one case in ten showed a disagreement of two steps; that is, where the I Q would have stamped a

child as belonging to the "very inferior" group, for example, the grade progress, or the teacher's judgment of the child's school work, or her estimate of intelligence would put the child in the "average" group, two steps removed, or vice versa.

Inasmuch as findings of this kind have led some to question the value of the Binet rating, it seemed worth while to ascertain the causes of such disagreements. Individual study of these cases showed beyond question that practically every two-step disagreement could be fully explained without questioning the validity of the intelligence test. They were due, in the main, to the teacher's neglect to take account of age differences in rating children's intelligence or school work. Time and again, older retarded who were plainly at the borderline of feeble-mindedness were rated "average" or only "slightly inferior" in intelligence and school work. If the child does ordinarily well the school work where he happens to be located the tendency of teachers is to rate his intelligence and school success "average" regardless of age. Conversely, the intelligence and school success of children who are below age for their grade are constantly underrated by the teacher. The study of grade progress in relation to mental age showed that the over-age dull children whose retardation has been so much lamented are as a rule not retarded, but actually accelerated. Under-grade according to chronological age, they are usually over-grade according to mental age. It appears, therefore, that by far the most common retardation is that due to the failure of teachers to recognize and advance the child of superior ability.

In closing, the opinion is ventured that the measurement of intelligence has already progressed far enough to warrant the testing of every child at the very beginning of his school career in order to determine the direction and extent of his deviation from median intelligence. Certainly all will agree that there is no other fact of greater significance for a child's educational guidance.