

DR. AYRES' CRITICISM OF THE BINET AND SIMON
SYSTEM FOR MEASURING THE INTELLIGENCE
OF CHILDREN—A REPLY

BY DR. F. KUHLMANN, FARIBAULT. MINN.

In this article Dr. Ayres wishes to "sound a note of warning against accepting" these tests in their present form as final and satisfactory," and states as his further object to "present considerations which lead the writer to believe that what we must have is a new instrument rather than the re-adjustment of the old; an instrument utilizing what is good in the old, but largely planned on different principles and constructed along different lines." This general conclusion is in direct contradiction to the findings of all who have used these tests extensively and have published their results, and together with his discussion of the more specific criticisms leaves the impression that the system is too defective to be of any great practical service. From my own experience in using these tests in examining about thirteen hundred of the inmates of the Minnesota School for Feeble-Minded I cannot agree with this view taken by Dr. Ayres, and there is much in his more specific criticisms which seems to me erroneous. These criticisms he summarizes as follows:

1. The tests predominantly reflect the child's ability to use words fluently, and only in a small measure his ability to do acts.
2. Five of them depend on the child's recent environmental experience.
3. Seven depend on his ability to read and write.
4. Too great weight is given to tests of ability to repeat words and numerals.

The Binet-Simon Measuring Scale for Intelligence ; Some Criticisms and Suggestions. By Leonard P. Ayres. Psychological Clinic, Vol. V, No. 6. 1911.

5. Too great weight is given to 'puzzle tests.'
6. Unreasonable emphasis is given to tests of ability to define abstract terms."

I shall take them up in order.

1. Perhaps all are willing to grant that the ability merely to use words fluently is no reliable criterion of intelligence, but neither is the ability merely to do acts. Both may become such only under specially arranged conditions. Language is the most direct indication and expression of thought and mental processes, which latter are the reactions we try to invoke in mental tests. We assume only that the language process shall not in itself be so difficult as to prevent its being a fairly adequate means of expression. But this being granted, we are not testing the ability to use words merely because the tests involve language. Dr. Ayres takes exception to this assumption in the case of the Binet-Simon tests, holding that "two-thirds of them are tests of the child's ability to use words," and that they "predominantly reflect the child's ability to use words fluently." Let us consider some of the instances he cites for illustration. He objects to the "Questions of Comprehension" in Test 4 for ten-year-old children. These are divided by the authors into two series, the words of the first series assumed to be easily understood by ten-year-old children, the words of the second series designed to offer some difficulty of understanding, as for example, "What should one do when he has missed the train?" (first series) and "What should one do when he is detained so that he will be late for school?" (second series). I answer the criticism by asking the reader the question, "Does it require a greater fluency of language to comprehend and answer such questions than we can rightly assume a ten-year-old child to possess?" I am left to infer that Dr. Ayres has misunderstood the aim of this test. A logically correct or even grammatically correct answer is not called for. The test is passed if the answers give merely an indication that he comprehends the situation suggested in the question and gives some intelligent reply. The authors give illustrations in each case of good and bad replies. A glance at these should be sufficient to show that the difference in these replies does not lie in

a difference in the ability to use words fluently. A further objection to these questions is found in the fact that they, "Overlook the importance of habit and of the emotions in influencing action;" that the child who would do just the right thing under the circumstances suggested in these questions, "thereby demonstrates a quality and degree of native ability to which few indeed among us may hope to attain." The latter part of this statement may be quite true, but the tests do not put the child under these exciting circumstances and judge his intelligence from his actions. They test his ability to understand what he ought to do if he were in these circumstances. Apparently in this same connection he objects to Test I for eleven-year-old children, in which the child is asked to point out the nonsense in each of a series of statements, on the grounds that the statements are "blood-curdling" and may "constitute a serious nervous shock to some sensitive children while it may not to others." Again, it is probably true that the emotional reaction will be somewhat different for different children. But, as I have stated elsewhere,* my own experience with this test does not reveal any grounds for objection to it on this score. Binet replying to the same criticism made by Whipple says: "Our young Parisians laugh at them." Indeed, it seems difficult to understand how the nervous shock to a child could be so great that he would fail to see the absurdity in the following, for example, if he were able to see it otherwise: "Yesterday they found on the fortifications the body of an unfortunate young girl cut into eighteen pieces. They believe that she killed herself." I turn to Dr. Ayres' second criticism.

2 Tests VII. 8, IX, 2 (naming pieces of money) IX, I (giving date) IX, 2, (naming days of the week) IX, 3, (making

*Binet and Simon's System for Measuring the Intelligence of Children. Journal of Psycho-Asthenics Vol. XV, 1911. Foot-note, P. 84.

Nouvelle Resherches sur la Mesure du Niveau Intelligence chez les Enfants D'Ecole. L'Annee Psychologique, 1911. Foot-note, P. 145.

Manual of Mental and Physical Tests. Baltimore, 1910. Foot-note, P. 509.

change, nine cents out of twenty-five) and XI, 1, (naming months of the year) come under this criticism. These and others have been criticized on the same grounds by other authors. It is said that they test acquired knowledge rather than intelligence, and the former is dependent on more or less accidental circumstances of environment. The answer involves a number of considerations, and views connected with them are not yet all entirely clear. In the first place, there is probably not a single test in the whole system that does not in some manner involve acquired knowledge or acquired ability. For mental functions (discrimination, memory, attention, will, etc.) can express themselves only through such acquisitions. But the tests are based on the assumptions (1) that certain kinds of knowledge and abilities must await the development of the mental functions involved in their acquisition and that these cannot be affected so much by environmental differences; (2) that in other instances the environmental conditions for acquiring certain kinds of knowledge and abilities are so uniform that we may assume that practically every child will acquire them if he has the mental capacity to do so. Binet and Simon discuss these assumptions in connection with some of the tests and admit that in some cases in these tests no conclusion is to be drawn from the result of these particular tests alone, because- of the effect of possibly an unusual environmental condition. If in some tests the child passes, we are to suspend judgment because his success may be due to specially favorable environment. But if he fails in these it shows lack of development of intelligence because environmental conditions are assumed to be uniform and favorable enough for every normal child to acquire the knowledge in question. If in certain other tests he fails, we are to suspend judgment because his failure may be due to specially unfavorable environment. But if he passes in these it shows the development of intelligence, because such development of intelligence is involved in the acquisition of the knowledge in question. Now, it would be difficult to decide off-hand in any given instance whether the intelligence is or is not tested by determining the presence and nature of a certain kind of acquired knowledge. But unless we

are prepared to disprove the underlying assumptions, implied and expressed, we cannot criticize these tests on the grounds merely that they determine acquired knowledge and not intelligence. They aim to test intelligence through acquired knowledge. Without resorting to the empirical facts which show that the system of tests is not so seriously affected by this or other sources of error as to give on the whole very incorrect results, Dr. Ayres' criticism here may be partly answered further by pointing out the unquestionable fact that the low grade feeble-minded fail to pass these simple tests of supposedly merely acquired knowledge that apparently might vary with different environmental conditions, although their greater chronological ages and environmental opportunities have been many times over what any normal child requires to learn to pass them readily.

3. In the 1908 series, which is under discussion, there are two tests (VIII,1, and XI, 5) that involve the child's reading, three, (VII, 3, VIII, 5, and X, 3) in which he has to write, and four, (V, 2, VII, 4, XIII, 1 and 2) in which he has to make simple drawings. The point in question here is only a special instance of what is discussed under the second criticism—tests influenced by acquired knowledge—and the same general answer applies. But it is to be added that some of these are avowedly tests directly of the process involved in the motor acts of writing or drawing, while others involve these motor acts or reading, only incidentally, and are used here as a means of expression of other mental functions that the tests aim at. The latter are used for chronological ages quite beyond the point where the average normal child acquires the ability of these motor acts themselves. This distinction should be considered besides several other matters before applying the general criticism to all the tests of this kind. Test VIII, 1, is on the ability to read a certain passage, and of this the authors say that no conclusion is to be drawn if the child fails, but if he passes it shows his intelligence through the fact that he has been able to learn to read. Test XI, 5, is on the ability to make sentences out of groups of words mixed up, which the child has to read, not on the ability to read. It should need no discussion to make clear that the results are to be used

in the same way here as in Test VIII, 1. Test VII, 3, is on the ability to copy a simple written phrase so as to be legible and thus involves only the initial stages of learning to write. VIII, 5, asks the child to write a simple phrase from dictation and the criticism is valid here in part, if anywhere. In X, 3, the child makes a sentence in which he uses three given words. The authors' directions are that the child write the sentence, but the test is not dependent on this ability to write, since he may be required to give it orally, if he cannot write, without altering the essential nature of the test. Dr. Ayres probably includes the tests involving drawing in his criticism, since he mentions seven while there are only five that involve reading or writing. In V, 2, the child copies a square, in VII, 4, he copies a diamond, and in the other two he draws simple forms the nature of which he has to figure out in the test, which part constitutes the test. In the first two, the processes involved in drawing are themselves directly tested and are so simple that the normal child can pass them at an age before he has had any experience in drawing of any consequence. And that training does not seriously increase the ability to make these drawings when the necessary mental development is not present, is shown by the fact that low grade feeble-minded cannot pass these tests irrespective of their greater chronological ages and more opportunities to learn this. In the other two there can hardly be any question as to the ability to make the drawings if the child is bright enough to give the occasion to give the thirteen-year-old test.

4. The tests in question here are criticized as having only a "remote relation to the ability to cope with the problems of life. The simpler of them can be successfully passed by a gifted parrot; the more difficult ones recently proved beyond the ability of a university professor tested by the writer." The simpler ones referred to are Tests III, 2 and 3; the more difficult ones are XII, 1 and 3. The former are based on the authors' observations that the normal child acquires several things in a quite definite order. The child first learns to understand our gestures and inflections of our voice; next he understands the spoken word; next he acquires the ability to repeat words spoken to him, and last the

ability of spontaneous speech and to name objects. These simpler tests aim to determine whether the child had acquired the ability to repeat things at all, rather than the ability to repeat a certain amount from memory. The question whether certain lower animals even can do as much is besides the point, and the ability to cope with the problems of life hardly comes in for consideration for three-year-old children. The value of the tests depends on whether the authors' observations on which the tests are based are correct or not. The more difficult tests referred to are on the amount that can be remembered on one hearing and on the ability of a brief concentration of attention that is involved. The system of tests rests on the assumption that the mental functions essentially involved in what we call intelligence are fairly well developed by the age of thirteen. If we turn to the results of the numerous memory studies for the different chronological ages we find that this assumption is substantially correct for the particular functions tested in the present kind of memory test. The average child at this age will do approximately as well as the average adult, but with a considerable individual variation for both adults and children, due to causes not yet fully understood. One is therefore not surprised that Dr. Ayres should find a university professor unable to pass the twelve-year-old test. In one way it is in favor of the test rather than against it; it indicates, at least, that the functions tested are not seriously affected by exceptionally long and special training.

5. The tests indicated here are spoken of as having strikingly "little relation to anything the normal person has to do in the ordinary day's work." Two only are mentioned for illustration. Test VIII, 4, counting backwards from twenty to one, and XIII, 2, drawing the figure of two juxtaposed triangles. Of the first, he says that it "is one of the rarest things most people are called on to do", that to teach such a task to children is "educationally vicious", and that to include such a requirement in a test of intellectual ability is at least "questionable". The import of these remarks are not entirely clear to the present writer. For from the previous discussion it is evident that we would not want to follow the rule of making tests of intelligence out of

tasks that people do frequently in everyday work or that are taught to them in the schools. This would bring in directly and constantly the effect of environment and acquired knowledge. In connection with the second, he notes that, "so far he has failed to find anyone able to describe the resulting shape" as is required in the test. I might reply with equal emphasis that I have found no normal adult that could not pass this test. Dr. Ayres surely cannot have given this test a very extensive trial. His result simply shows that there is no perfect correlation between this individual test and intelligence. But such perfection is not claimed for any of them. In what other ways tests of this sort might be objectionable is not pointed out.

6. Two tests, XI, 4, and XIII, 3, of the fifty-three, involve giving definitions of abstract terms. To me this does not seem an unreasonable emphasis on this sort of test. They are objected to further because such definitions are too difficult; even normal adults with the aid of a dictionary cannot find satisfactory definitions of the words chosen in these tests. I think Dr. Ayres has again somewhat misunderstood the aim of these tests, and what the authors intend should be accepted as satisfactory definitions. The tests do not call for definitions logically faultless. They aim to determine whether the child is capable of grasping abstract meanings, and any reply that shows this is accepted. Again, the authors' discussion and illustrations of satisfactory and unsatisfactory definitions should make this matter clear.

Dr. Ayres next takes up the criticism of two sets of facts in favor of the tests. (1) "They have won rapid and wide-spread use and endorsement among hundreds of practical teachers and workers with children." This, he thinks, is due to the fact that the tests grade children according to mental ages, a scale that is at once definite and universally understood. (2) The results with the tests obtained by the authors and by Dr. Goddard on large numbers of normal children show that the mental ages as determined by the tests and the chronological ages coincide or vary only by a year in the vast majority of instances. Dr. Ayres does not agree with Dr. Goddard's conclusion from this result that it is a proof that, on the whole, the tests give an accurate

measurement of mental development, because he finds that of 14,762 public school children passed through the first seven grades, the vast majority did so in seven years, with the numbers that took more or less time to complete these grades decreasing rapidly to zero. He argues that if the former result with the tests is a proof of their accuracy, then this is equally a proof that "the public school systems and courses of study are correctly adjusted to the abilities of their pupils," and that if this were true we would have less need for tests, because the public school system already furnishes the test of the children's intelligence. The reply to this argument and criticism is that the first part of it is entirely correct. On the whole the public school systems and courses of study are correctly adjusted. The practical coincidence of the two kinds of results is a further corroboration of the correctness of Dr. Goddard's conclusion as to the tests, not evidence against it. In the second part of his argument, Dr. Ayres seems to forget that although the tests and public school system might give equally accurate results on the mental status of a child, it takes about an hour with the tests, and several years with the public school system method to decide this in any given case. What the results referred to with the tests do not show is what particular individual tests in the system are poorer; or better than others. Dr. Goddard would have gotten the same kind of frequency distribution curve if all his children tested had been just average children and the system of tests had been sufficiently inaccurate to give occasional errors of a year or two in the mental ages. In this event, also, we would expect that the smaller differences in mental and chronological ages would occur most frequently and the larger differences less frequently.

Dr. Ayres' criticisms are apparently not based on his own experience in any extensive use of the Binet-Simon tests, but, as he says, are the result of "his own attempts to discover, ways in which they may be improved, together with ideas secured through lengthy discussions of their application with Mrs. LOUISE Stevens Bryant, of the Psychological Clinic of the University of Pennsylvania." His general conclusion disagrees with those who have had such experience. They are largely the result of an at-

tempt at an internal analysis of the nature of the tests themselves. Criticisms of this sort have their value as a means of calling attention to problems that may need to be solved, but the literature on the Binet-Simon tests has by this time abundantly shown that for the most part they cannot be accepted on the basis of any mere logical plausibility, without any empirical testing out. To the present writer, Dr. Ayres' criticisms seem to come largely from a misunderstanding as to what the different individual tests aim at, and of the mental processes involved in them. The former might have been largely obviated by a more careful consideration of the authors' original publications, and the latter by a careful and extensive use of the tests themselves. There is, especially, a general impression that the authors meant that the results with each individual test will always come out just right, which impression Dr. Ayres seems to share somewhat. If this degree of perfection were attained, only one test of mental age for each chronological age would be necessary, where the authors use from four to eight, and besides point out in many instances what proportion of correct results the test has been found to give. It is therefore, not a fair criticism to point out that this or that individual test often gives wrong results. Probably not a single test in the whole system is free from such objection. In general, this article reminds one that it is easy to make criticisms and difficult often to clearly disprove them. But even so, the validity of merely possible objections is not thereby established.