A SCALE OF PERFORMANCE TESTS

A SCALE OF PERFORMANCE TESTS

BY

RUDOLF PINTNER

ASSISTANT PROFESSOR OF PSYCHOLOGY, OHIO STATE UNIVERSITY

AND

DONALD G. PATERSON

INSTRUCTOR IN PSYCHOLOGY, UNIVERSITY OF KANSAS

D. APPLETON AND COMPANY

NEW YORK LONDON

1917

Printed in the United States of America

PREFACE

This book is an attempt on our part to contribute to the few scales already in general use another kind of scale for the purpose of testing intelligence. The work grew directly out of the psychological examination of deaf children, for which purpose the ordinary scales for the measurement of intelligence were found to be practically useless. It was therefore decided to assemble a group of tests which did not involve any kind of language response. This work was begun in 1914 with the standardization of a few performance tests and since then has grown to the present dimensions. The work of testing has very largely been done by ourselves. We have, however, decided to incorporate in the scale at least one test that has been standardized by another worker, namely the Seguin Form Board Test as standardized by Sylvester. Miss Margaret M. Anderson, Graduate Assistant in the Department of Psychology, is responsible for the standardization of the Picture Completion Test. Miss Jeannette Reamer, Miss Alice E. Beekman, and Miss Lucille Boylan, Graduate Students, have helped greatly in the accumulation of data for some of the tests.

We take this opportunity to acknowledge the assistance and coöperation on the part of the teachers and principals in the schools of Columbus, in which the tests were conducted. In particular, we wish to thank Mrs. Scatterday, Principal of Northwood School; Miss Gordon, Principal of Ninth Avenue School; Miss Thompson, Principal of Second Avenue School; Miss Neerermer, Principal of Heyl Avenue School, and Mr. Bryant, Principal of Indianola School.

RUDOLF PINTNER.
DONALD G. PATERSON.

Columbus, Ohio. 1917.

CONTENTS

CONTENTS

LIST OF ILLUSTRATIONS

GRAPHS

LIST OF ILLUSTRATIONS

A SCALE OF PERFORMANCE TESTS

A SCALE OF PERFORMANCE TESTS

CHAPTER I

INTRODUCTION

The measurement of intelligence at the present time is a well recognized part of psychology. The growth of this work and the interest shown in it during the last three decades have been truly remarkable. We have witnessed the establishment of innumerable clinics and the appearance of the "mental tester." This growth has been characterized by the practical considerations of clinical examinations. The need for a psychological examination has been recognized and answers to practical situations have been demanded before the psychologist has really had time to formulate his own conceptions of the whole problem. Theoretical considerations have lagged behind the practical application of mental tests. We have been measuring intelligence long before we have decided as to what intelligence really is. Far from being a drawback, as this at first sight would appear to be, it has in fact proved to be of distinct advantage, since the measurement of this something, that we have been making, is leading us slowly but surely to a real knowledge of what can with profit be called "general intelligence." Only after considerable work with mental tests did psychologists arrive at the now generally accepted definition of intelligence, as enunciated by Stern,[1] that "Intelligence is a general capacity of an individual consciously to adjust his thinking to new requirements: it is general mental adaptability to new problems and conditions of life." Although even this definition of intelligence may be modified in the future, it serves, at the present time, as a good working hypothesis for the selection of tests for mental measurement.

Other conceptions of general intelligence are numerous and many are very similar. Binet,[2] for example, says: "It seems to us that in intelligence there is a fundamental faculty, the alteration or the lack of which is of the utmost importance for practical life. This faculty is judgment, otherwise called good sense, initiative, the faculty of adapting one's self to circumstances. To judge well, to comprehend well, to reason well, these are the essential activities of intelligence." Meumann [3] says general

---

[1] Stern, W.: The Psychological Methods of Testing Intelligence, Trs. by G. M. Whipple, *Educational Psychology Monographs,* No. 13, Warwick and York, Baltimore (1914), p. 3.

[2] Binet, A., and Simon, Th.: The Development of Intelligence in Children. Trs. by Kite, Vineland, New Jersey, 1916, p. 42.

[3] Meumann, E.: Experimentelle Pädagogik, Vol. ii, Leipzig (1913), p. 102 *et seq.*

intelligence depends on two qualities, "(1) on the capacity for independent, productive thought (productive, synthetic thinking) . . . (2) the intensity of the whole mental life." Ebbinghaus [4] makes intelligence include abstraction and the ability to compare and contrast. Burt [5] says that the result of his work "strongly suggests that it is one feature or function of attentive consciousness which forms the basis of intelligence, namely, the power of readjustment to relatively novel situations by organizing new psycho-physical coördinations." And lastly Hart and Spearman [6] look upon general intelligence as a "common factor" or "central tendency" not exactly definable, but entering into all sorts of mental activities to a greater or less degree.

In addition to these hypotheses as to the nature of intelligence, theoretical considerations as to the growth of intelligence and the general distribution of the various grades of intelligence are arising as a direct result of the practical work being done. Questions as to the rate of increase in normal mentality are being raised,[7] and the theoretical assump-

---

[4] Ebbinghaus, H.: Grundzüge der Psychologie, Vol. ii, Leipzig (1913), p. 307.

[5] Burt, C.: "Experimental Tests of General Intelligence," *British Journal of Psychology,* Vol. iii (1909-10), pp. 94-177.

[6] Hart, B., and Spearman, C.: "General Ability, Its Existence and Nature," *British Journal of Psychology,* Vol. v (1912-13), pp. 51-84.

[7] Otis, A. S.: "Some Logical Aspects of the Binet Scale," *Psychological Review,* Vol. xxiii, Nos. 2 and 3 (1916), pp. 129-152 and 165-179.

tions, upon which we base our classification of individuals into different groups, are being discussed.[8]

Along with this, the more technical question of the standardization of tests is arising and is becoming more pressing in proportion to the demands for finer and finer differential diagnoses on the part of the practical worker.

The first tests made by psychologists were not intended as measurements of intelligence. We might characterize them as individual tests. They seem to have arisen as a direct result of the individual differences noted in the laboratory by the experimental psychologist. At first these individual differences were a distinct hindrance to the psychologist, but soon he became interested in them for their own sake, and once this occurred we have the birth of the test, which is a measurement of the differences between individuals. The differences between individuals in sensory discrimination led to tests for sensory discrimination, and so on with the other divisions of psychology. These first tests are concerned with the measurement of specific "facul-

---

[8] Pintner, R., and Paterson, D. G.: "A Psychological Basis for the Diagnosis of Feeble-mindedness," *Journal of Criminal Law and Criminology,* Vol. vii, No. 1 (1916), pp. 32-55.

Kohs, S. C.: "The Distribution of the Feeble-minded Arranged by Mental Age (Binet)," *Journal of Delinquency,* Vol. i, No. 2 (1916), pp. 61-71.

Kuhlmann, F.: "Distribution of the Feeble-minded in Society," *Journal of Criminal Law and Criminology,* Vol. vii, No. 2 (1916), pp. 205-218; "What Constitutes Feeble-mindedness?" *Journal of Psycho-Asthenics,* Vol. xix, No. 4 (June, 1915), pp. 214-236.

ties" or capacities or abilities. They are tests of the different mental processes or of the different states of consciousness. There are tests for various motor and sensory capacities, for attention and perception, for association, for learning and memory, for suggestibility, for imagination, and so forth. The work with these individual tests has been very considerable and has thrown a great deal of light upon the mental capacities of individuals. It would be futile to attempt in this book to give any account of the development and scope of the individual test, or of the psychologists who have contributed to this field.[9] It is sufficient for our purpose to mention these facts in order to note that the scale or group of tests for mental measurement has arisen from the individual test. Looked at from this point of view we may say that the mental scale is merely the grouping together of individual tests in order to give a more general picture of the mental make-up of the individual. Strictly speaking, a scale for the measurement of intelligence is more limited in scope than the above description would suggest, since it omits a great many capacities or abilities that are not supposed to be indicative of the mentality of an individual. For example, there are tests for the ability to discriminate two points on the skin, for the ability to discriminate between different shades

[9] For a complete account of individual tests of this nature and for an extended bibliography, see Whipple, G. M.: Manual of Mental and Physical Tests, Two Volumes, Warwick and York, Baltimore (1914-15).

of color, but we do not include such tests in our scales of intelligence, because it is not believed, at the present time, that such tests have diagnostic value for distinguishing between different grades of intelligence. Bagley,[10] for example, found "a general inverse relation between motor and mental ability," although there were numerous individual exceptions.

The idea of using a group of tests for the purpose of estimating the intelligence of an individual originated with one of the best known workers in the field of individual tests, Alfred Binet. He had for a long time been interested in the question of tests for various abilities and we have a long series of articles by him dealing with individual tests.[11] His work gradually led him to a study of individual cases, and in summing up the psychological characteristics of individuals as revealed by mental tests he came upon the idea of using a number of tests as a measure of the individual's capacity. In addition to this his theoretical speculations as to what the tests were testing led him to the conclusion that "attention" and "adaptation" were at bottom the chief factors that distinguished intelligent from unintelligent children.[12] And it is to

---

[10] Bagley, W. C.: "On the Correlation of Mental and Motor Ability in School Children," *American Journal of Psychology*, Vol. xii (1900-01), pp. 193-205.

[11] Binet, A., and Simon, Th.: The Development of Intelligence in Children, Trs. by Kite, Vineland, New Jersey (1916).

[12] Binet, A.: "Attention et Adaptation," *L'année psychologique*, Vol. vi (1900), pp. 248-404.

be noted that his definition of attention is very different from the ones at present customary in psychology and that it approximates very closely to the later definitions of intelligence of Stern and Meumann.

All this work of Binet's led him directly to the problem of the measurement of intelligence. A practical situation presented to him called forth the first actual group of tests for differentiating between intelligent and unintelligent children. This problem was the selection of the most backward children in the schools for the purpose of giving them special instruction. Binet was called upon to discriminate between the normal child and the backward child, and the question was not whether this or that child was better in such a specific thing as memory or imagination and so forth, but whether the child was in general weaker in his intellectual endowment than the average child of his age. Binet, therefore, took the next logical step in advance of the position that he had reached in his work with individual tests. He discarded the specific test for the specific ability and took a group of tests which seemed to cover in general the chief psychological characteristics that go to make up intelligence. And, furthermore, as the norm or standard of intelligence he took what the average child at each age could do.

These two points, the use of a group of tests and the average performance at each age as a standard of measurement, form the basic principles upon which all our measuring scales of intelligence now rest. For this happy combination we have to thank the genius of Binet.

At first the group of tests used by Binet was not arranged according to years, but soon there appeared the Binet-Simon Scale in the form that we now know it.[13] From this time on we can speak of a scale for the measurement of intelligence. The recognition of the value of this scale was immediate and wide-spread. It was used extensively in France and in other countries, and presently we have the appearance of scales adapted to the different countries in which they were used. In America Goddard's[14] Revision appeared early and was and is still extensively used. A translation from the French was made by Town.[15] In Germany the scale was adapted by Bobertag.[16] In England work was done by Johnston,[17] and

---

[13] Binet, A., et Simon, Th.: "Le développement de l'intelligence chez les enfants," *L'année psychologique,* Vol. xiv (1908), pp. 1-94.

[14] Goddard, H. H.: "The Binet-Simon Measuring Scale of Intelligence, Revised," *Training School Bulletin,* Vol. viii (1911), pp. 56-62.

[15] Binet, A., and Simon, Th.: A Method of Measuring the Development of the Intelligence of Young Children, Trs. by Town, Courier (1913).

[16] Bobertag, O.: "Kurze Anleitung zur Ausführung der Intelligenzprüfung nach Binet und Simon," *Institut der Gesellschaft für Psychologische Experimente* (1913), Nr. 8.

[17] Johnston, K. L.: "M. Binet's Method for the Measurement of Intelligence; Some Results," *Journal of Experimental Pedagogy,* Vol. i (1911), pp. 24-31; and also "The Measurement of Intelligence; Binet-Simon Tests," same journal, Vol. i (1911), pp. 148-151.

more recently a new adaptation for English use has been proposed by Winch.[18] Other American adaptations that were proposed are those of Kuhlmann [19] and Wallin.[20]

Within a relatively short time the literature dealing with the Binet-Simon Scale grew to immense proportions,[21] and the uses to which it was put were numerous. It found early and wide-spread use in juvenile courts, in state surveys of feeble-mindedness, in the selection of children for special classes and to some extent in helping to solve other problems of the school. Each one of these varied uses of the scale has a literature of its own and it would lead us too far afield to enter into any one or all of these aspects of the measurement of intelligence.

As was to be expected, the use of the scale and the abuse of it in some quarters aroused a mass of constructive and destructive criticism. The result of this criticism on the constructive side led to a sharper and more definite formulation of the

---

[18] Winch, W. H.: "Binet's Mental Tests; What They Are and What We Can Do with Them," a series of articles in *Child Study,* Vols. vi, vii, and viii.

[19] Kuhlmann, F.: "A Revision of the Binet-Simon System for Measuring the Intelligence of Children," *Journal of Psycho-Asthenics,* Monograph Supplement, No. 1 (1912), p. 41.

[20] Wallin, J. E. W.: Experimental Studies of Mental Defectives: A Critique of the Binet-Simon Tests, Warwick and York, Baltimore (1912).

[21] For the literature up to 1914, see Kohs, S. C.: "The Binet-Simon Measuring Scale: An Annotated Bibliography," *Journal of Educational Psychology,* Vol. v (1914), pp. 215-224, 279-290, 335-346.

principles involved in the work of mental measurement. It is during this period of criticism that we have arrived at a clearer understanding of what we mean by intelligence, and of the requirements that a test must possess in order to be an adequate test of intelligence. We are more generally agreed now as to what intelligence is and we are using this as a criterion for the choice of tests for scales of mental measurement. Certain of the tests in the original Binet Scale have been criticized severely in the light of this newer conception of mental measurement. It is questioned in some quarters as to whether tests of specific pieces of information such as the child may be taught in school or in the home can with justice be included in our scales, since the latter are frankly trying to measure innate or native endowment rather than any particular bit of knowledge acquired by specific training. Binet himself raised this question in his revision of the original scale, and Stern enters into a discussion of the lack of agreement between tests of intelligence and school performance, bringing out clearly the difference between general intelligence and acquired knowledge. Other writers [22] have pointed out how certain tests of the scale depend upon knowledge acquired through experience, while other tests seem to be unaffected by the sub-

[22] Pintner, R., and Paterson, D. G.: "Experience and the Binet-Simon Tests," *Psychological Clinic*, Vol. viii, No. 7 (1914), pp. 197-200; also, by the same writers, "The Factor of Experience in Intelligence Testing," *Psychological Clinic*, Vol. ix, No. 2 (1915), pp. 44-50.

ject's amount of experience. On the other hand, the difficulty of finding anything that is not influenced by education in school is well recognized and we are forced to take for granted the acquisition of such general abilities as reading or writing in children that grow up in the ordinary civilized community.

Again, the demand on the part of the practical worker for more and more accurate diagnoses has raised the whole question of the accurate placing of tests in the scale and the accurate evaluation of the responses made by the child. In general this may be termed the problem of standardization. This question of standardization has led to the two latest revisions of the Binet Scale, namely, the Stanford Revision and Extension of the Binet-Simon Scale by Terman,[23] and the Point Scale by Yerkes, Bridges and Hardwick.[24] The Stanford Revision adheres more closely to the original Binet Scale and makes no departure from the classification of tests according to age. The scale adds certain tests to those originally used by Binet. It makes its chief contribution, however, in the standardization of the tests themselves and in the use of the intelligence quotient as the index of the mentality of the examinee.

---

[23] Terman, Lewis M.: The Measurement of Intelligence, Riverside Text-books in Education, Houghton, Mifflin Company (1916).

[24] Yerkes, R. M., Bridges, J. W., and Hardwick, R. S.: A Point Scale for Measuring Mental Ability, Warwick and York, Baltimore (1915).

The Point Scale, above referred to, while making use of a great many of the original Binet tests, differs considerably in the method adopted to arrive at a measurement of the subject's intelligence. It discards the grouping of tests according to age and adopts the scoring of responses by means of allotting a certain number of points to each test. Like the Stanford Revision, it also rejects the mental age as an adequate statement of the mentality of the case, and proposes the coefficient of mental ability instead. This coefficient is the ratio of the score made to the average score for a child of the age of the individual examined, just as the intelligence quotient is the ratio of the mental age to the chronological age of the child.

These two scales may be taken to represent, for America at least, the result of the constructive work done on the basis of the original Binet tests. Although the Stanford Revision has introduced a great many new tests and the Point Scale a few, yet the general nature of the tests remains much the same as those originally proposed by Binet. While this work has been going on, there has been in addition another phase of criticism of the original Binet Scale that has been directed particularly against the great number of tests in the scale that require language responses. This criticism of the scale was made by Ayres [25] shortly after the scale

[25] Ayres, L. P.: "The Binet-Simon Measuring Scale for Intelligence: Some Criticisms and Suggestions," *Psychological Clinic,* Vol. v (1911), pp. 187-196.

had come into general use. Just how much the ability to handle language is indicative of intelligence is the question at issue. Have we a valid test when ability to pass it depends not merely upon comprehension of language, but also upon the ability to frame an adequate language response? This language difficulty, inherent in the Binet Scale and in all the revisions of it, became very pronounced as soon as the use of the scale spread to workers in various fields of practical work. The clinical psychologist in the large city was face to face with the problem of the foreign child, the speech defective, the deaf child and other children with language difficulties. It was obvious, from the beginning, that the Binet Scale was inadequate for the mental examination of such cases. Other tests not involving language were introduced and this gave rise to the type of test now generally known as the performance test. The essential characteristic of this type of test is that it shall not require any kind of a language response on the part of the child for an adequate performance of the test.

An excellent group of performance tests which had been found of practical value in the diagnosis of cases was described by Healy and Fernald,[26]

Kuhlmann, F.: "A Reply to Dr. L. P. Ayres' Criticism of the Binet and Simon System for Measuring the Intelligence of Children," *Journal of Psycho-Asthenics,* Vol. xvi (1911), pp. 58-67.

[26] Healy, W., and Fernald, G. M.: "Tests for Practical Mental Classification," *Psychological Monographs,* Vol. xiii, No. 2, Whole No. 54 (1911).

and we have incorporated some of these tests in the present scale. It was the problem of the foreign child and the child with language difficulties that forced Healy to have recourse to other tests in addition to those of the regular scales, in order to arrive at a better understanding of the mentality of the children examined. Healy did not attempt to group his tests in the form of a scale, but simply used them as additional aids for diagnostic purposes. Some of these tests have been extensively used by other workers, and partial standardizations of some of them have been made. A discussion of these standardizations will be given later.

Confronted with the problem of testing non-English speaking immigrants at Ellis Island, Knox[27] found it impossible to use scales in which language responses were required, even though the services of an interpreter might be used. He devised a series of performance tests, which he constructed into a kind of scale for the purpose of estimating the mentality of the immigrant. Knox's scale is admittedly rough and lacking in standardization. Many of the tests are excellent and we have included some of them in the present scale. These two groups of tests collected by Knox and by Healy have proved to be very valuable; but

[27] Knox, H. A.: "A Scale, Based on the Work at Ellis Island, for Estimating Mental Defect," *Journal of the American Medical Association,* Vol. lxii (March 7, 1914), pp. 741-747.

there was lacking in both cases an adequate standardization, and this has prevented an understanding of the meaning of any specific performance on the tests.

In the case of the writers [28] themselves, the ordinary scales of intelligence were found absolutely inadequate to test the mentality of deaf children. They were forced to look around and to devise performance tests for this purpose.

These various practical considerations have led us to recognize the necessity of developing a scale of performance tests. Such a scale should prove of distinct advantage in the work of measuring intelligence. It may be used as a supplementary scale in addition to the ordinary scales of intelligence. If it is true that the Binet Scale rates the child with superior language ability too high, then a performance scale used as a supplement to the Binet Scale should serve as a corrective. The Binet Scale and its revisions do not consist entirely of language tests, but the number of tests calling for a language response is very great and it may be, as some workers have felt, that too much credit is given for this type of response. Language ability is not always directly correlated with general intelligence. Healy has called a certain type of cases "verbalist," because this type is characterized by an ability to handle language decidedly

---

[28] Pintner, R., and Paterson, D. G.: "The Binet Scale and the Deaf Child," *Journal of Educational Psychology,* Vol. vi (1915), pp. 201-210.

above its ability along other lines. As Healy[29] says: "On account of their ability to handle language well the members of this group are not properly placed by the ordinary tests of social intercourse. The common method of passing judgment on people is, of course, through conversation. One asks questions and if one gets answers that follow properly, that are consequential and coherent, why then without more ado one infers the answerer to be practically normal. The give-and-take conversational method of the court room may be offered in illustration," and again further on, "One of the weak points of the Binet system is that it so greatly calls for language responses; those who have good language ability easily grade proportionately higher." A better understanding of the mentality of this verbalist type might be arrived at by a performance scale used as a supplement to the ordinary scale of intelligence. Because our scale of performance tests might be used in this supplementary way, we decided not to include in the present scale any tests of the performance type already included in the Binet Scale.

For the testing of non-English speaking children coming from homes where the English language is not customarily spoken, the advantage of a performance scale is, of course, obvious. It is absurd to pretend to measure the mentality of a foreign child by means of our present scales. This has been

[29] Healy, W.: The Individual Delinquent, Little, Brown and Company (1915), p. 473 *et seq.*

done repeatedly in the past by many careless workers, who seem to look upon the Binet Scale as an infallible measuring rod which can be applied with a minimum of critical judgment both as to the method of procedure and the evaluation of results. It is to be hoped that the tentative scale of performance tests here presented will help workers to arrive at a more adequate measurement of the mentality of the foreign child. The difference between the English speaking and the non-English speaking children tested by Yerkes and Bridges was shown to be considerable as judged by their scores on the point scale. Recognizing the number of language tests included in the scale the authors presented norms for both groups, so that future cases might be judged in the light of the group to which they belonged. No one would feel justified in concluding from the scores made by these two groups that the non-English speaking group was lacking in mentality as compared with the English speaking group to the extent suggested by the difference in their scores. Kent,[30] after commenting on the Yerkes racial norms, says: "In testing children of immigrants by groups, it might be possible to make allowance in the final scoring of results for this disadvantage; but if children are to be tested as individual cases this would not be a safe plan to follow, inasmuch as the deficiency in the

---

[30] Kent, G. H.: "A Graded Series of Geometrical Puzzles," *Journal of Experimental Psychology,* Vol. i, No. 1 (February, 1916), p. 41.

language might be so marked as to entirely invalidate the results for the particular child in question."

Closely allied to the problem of the non-English speaking child is the problem of the speech defective. It is doubtless true that a great many speech defectives are mentally backward, but there are many who are not, and who are at present being misjudged by the results of tests made by means of the standard intelligence scales. At any rate, we cannot arrive at an adequate measurement of such cases with our present scales, and the need for a performance scale is obvious.

We feel also that a performance scale will be useful in arriving at a better measurement of the mentality of children coming from different language environments. We mean by this that there are certain types of homes in which the child learns very little in the way of language. There are no books and very little reading is indulged in. On the whole, it may be true that such homes indicate a lower mentality of the people, but this is not true in every case. There are cases, again, where children have never learned to read or write, and this not from inability but from lack of opportunity to learn. It is clear that we cannot arrive at a just measure of the mentality of these cases with the present scales of measurement. Even if we were to take the extreme position and maintain that all such cases are indicative of some slight degree of backwardness in mental development, we would be penalizing them too severely in judging them by means of language tests. They may be backward, but the real extent of their backwardness will be measurable by means of a performance scale rather than by means of a scale including language tests.

Comparing the mentality of dependent children in charitable homes with children in the ordinary public schools, it was found [31] that the dependent children fell below the school children to a much greater extent on tests involving language than they did on tests of mechanical ingenuity, and we venture to suggest that the real difference in the mentality of the two groups of children was expressed by the difference on the performance tests rather than by the difference on the tests involving language. That children in better class schools always test higher on the Yerkes Scale and on the Binet Scale than do children in schools in poorer environments is doubtless due to the better mentality of the former group, but some of the superiority shown by the former group may be due to their superior language environment.[32]

Lastly, a scale of performance tests is a *sine qua*

---

[31] Stenquist, J. L., Thorndike, E. L., and Trabue, M. R.: "The Intellectual Status of Children Who Are Public Charges," *Archives of Psychology,* No. 33 (1915).

[32] Yerkes, R. M., Bridges, J. W., and Hardwick, R. S.: A Point Scale for Measuring Mental Ability (1915), p. 75 *et seq.*

Bridges, J. W., and Coler, L.: "The Relation of Intelligence to Social Status," *Psychological Review,* Vol. xxiv, No. 1 (1917), pp. 1-31.

non in the measurement of the mentality of the deaf. Here we have a group of individuals completely shut off from hearing language and for that reason laboring under a language difficulty that only in rare cases is surmounted to the extent of making them comparable in language ability to ordinary hearing individuals. Any kind of test involving written or spoken language cannot be used as a test of their mentality.[33] If we employ such tests for measuring the mentality of the deaf and use the standardizations obtained from hearing children, we will not be measuring mentality but merely differences in language ability. There may be a greater percentage of feeblemindedness among the deaf than among the hearing, but the fact that a deaf child does not measure up to the language standards of a hearing child is no indication of mental deficiency. For the deaf some kind of performance scale such as we have devised is necessary in order to arrive at a rough measure of their mentality. Our performance scale was specifically devised with the deaf child in mind.[34] Indeed, it has been so constructed that practically no instructions need to be given to

---

[33] Pintner, R., and Paterson, D. G.: "The Ability of Deaf and Hearing Children to Follow Printed Directions," *The Pedagogical Seminary*, Vol. xxiii, No. 4 (1916), pp. 477-497; "A Measurement of the Language Ability of Deaf Children," *Psychological Review*, Vol. xxiii, No. 6 (1916), pp. 413-436.

[34] The writers will publish shortly, in a book entitled "The Psychology of the Deaf," the results of and the methods for the application of this performance scale to deaf children, together with norms for deaf children already tested.

the child with the exception of such as can be made by means of natural gestures.

For all these groups of children—the foreign child, the speech defective, the deaf child and so on—a scale of performance tests is the only adequate means for the measurement of mentality. The language factor must be omitted and our estimate of mentality must be based upon what any of these kinds of children can do as compared with the normal hearing and speaking child.

The Selection of Tests. In selecting performance tests for this scale, the object was to get as many different kinds of tests as possible, so that all the various factors entering into the complex known as intelligence might be brought into play. It was not our aim to make any theoretical analysis of the various capacities or abilities that might be included in such a complex, but rather to choose tests that seemed to call forth different types of response, realizing that the response called forth in any specific instance might be variously described as involving memory or attention or something else, or all of these things together.

In addition to this principle in the selection of tests, there was the other principle which follows from our general definition of intelligence as the capacity of adjusting to relatively new situations, the principle, namely, that each test should present a relatively new situation to the child. A test must not demand the performance of a specific activity that is likely to have been learned by the child. Familiarity or unfamiliarity with the general type of response required by a specific test is always a relative matter. We could not, for example, exclude all picture block tests simply because some children are more familiar with this type of toy than other children. What was done, however, was not to include as a test any well-known article which was already in common use as a child's toy or plaything. From this point of view a recently devised test might be criticized as being a toy in very common use among children.[35]

A third criterion in the selection of tests was that no verbal instructions should be necessary in order to give the tests. All of the tests in our scale, with a few minor exceptions, can be called self-explanatory. The situation itself calls for some response without the necessity for any verbal instructions on the part of the examiner. A sign to go ahead quickly is all that is necessary. Naturally in giving the test to hearing children the examiner will say something, but what he says is not essential for the understanding of the test. If the examiner in testing a hearing child were not to say anything, he would introduce an embarrassing and abnormal element into the situation. It is for this reason that in the directions for the tests verbal instructions for the examiner are given. These are not necessary and in testing deaf or for-

[35] Kelley, T. L.: "A Constructive Ability Test," *Journal of Educational Psychology*, Vol. vii (1916), pp. 1-16.

eign children can be omitted without changing in any way the nature of the tests. In some of the tests, however, in testing deaf or foreign children, preliminary practice is necessary, as in the Substitution Test, and this preliminary procedure has been standardized and will be published in our book on the psychology of the deaf.

While the work of standardization was in progress some tests, which we had originally chosen, were rejected since they did not seem to be giving satisfactory norms. One of these tests was Knox's Imbecile Form Board. Another was a type of Seguin Form Board constructed by ourselves. This differed somewhat from the one in common use and was abandoned because it showed no superiority over the standard Seguin Board and because our data would, therefore, be incomparable to the data already gathered by other workers. We have contented ourselves with using the data gathered by Sylvester with the Seguin Form Board.

The tests chosen which we believe fulfill the above requirements are as follows:

1. The Mare and Foal Picture Board. A modification of the original as designed by Healy.
2. The Seguin Form Board. Twitmeyer's adaptation of the Goddard Board or the Goddard Board itself.
3. The Five Figure Board, devised by Paterson.
4. The Two Figure Board, devised by Pintner.

5. The Casuist Form Board, a copy of the original board devised by Knox.[36]
6. The Triangle Test, devised by Gwyn.
7. The Diagonal Test, devised by Kempf.
8. Healy Construction Puzzle A, devised by Healy.
9. The Manikin Test, devised by Pintner.
10. The Feature Profile Test, devised by Knox and Kempf.[37]
11. The Ship Test, devised by Glueck.
12. The Picture Completion Test, devised by Healy.
13. The Substitution Test, devised by Woodworth and Wells.
14. The Adaptation Board, devised by Goddard.
15. The Cube Test, devised by Knox and modified by Pintner.

Note. Tests Nos. 3, 4, 5, 9, 10, and 11 can be obtained from A. P. Freund, Mechanic, Ohio State University, Columbus. All the other tests can be obtained from C. H. Stoelting Co., Chicago, Illinois. Test 15, as supplied by Stoelting, will not be suitable for Pintner's modification, which merely requires five cubes of the same size and color. The writers have generally made use of the Binet cubes, but any similar blocks, all of the same color and size, will do.

---

[36] This board was made by the writers before the Stoelting Company supplied Knox's tests. Our board is larger than the one supplied by Stoelting and differs in some details. *Our norms cannot be said to apply to the board manufactured by Stoelting.*

[37] This test also was made by the writers before the Stoelting Company supplied Knox's tests. It differs somewhat from the one supplied by Stoelting, and our norms can only with certainty apply to our form of the test.

CHAPTER II

THE TESTS

THIS chapter contains a description of the tests used, and they are described in the order in which they were generally presented to the children. The sequence here given is recommended for other workers. Similar tests have been grouped together. The first test (Mare and Foal) is one of the easiest and is of the picture form board variety. The nature of the performance required is understood by almost all children without verbal instructions. A glance at the board with the pieces out is enough to call forth the response of filling in the pieces. After this follow Tests 2 to 8, which are all of the form board character. They require the insertion of blocks in appropriate spaces and, increasing in difficulty as they do, the child is led naturally on from one to the other with a minimum of instructions. Tests 9 and 10 can hardly be called form board tests, but the nature of the performance is similar. This time the child sees that he must fit things together, but without the help of spaces into which the parts must fit. Test 11 demands the construction of a picture, continuing the idea of making up something, the parts of which are before the child at the beginning of the test. Test 12 demands the fitting in of blocks, but this time there must be the selection of appropriate blocks from a large number of others. Test 13 is radically different and requires new instructions. It is at this point that the material deviates radically from the form board type. The last two tests (14 and 15) are likewise totally different from the others, but by this time the child is well adjusted to the examination.

Test 1. The Mare and Foal Picture Board

(a) *Description.* This test is a slight modification of the one devised by Healy.[1] It is a board measuring 29 by 24.5 centimeters and 1 centimeter thick, upon which a colored picture is pasted. The picture represents a mare and foal in a field with two sheep lying down and three chickens in the foreground. In the background two houses are seen in the distance. Eleven pieces have been cut out of the picture and the pieces are of different shapes. They represent certain parts of the animals or of the scene. The modification of the original board as made by us is the omission of the four geometrical pieces at the top of the picture. After some preliminary experimentation these four pieces were

[1] Healy, W., and Fernald, G. M.: "Tests for Practical Mental Classification," *Psychological Monographs,* Vol. xiii, No. 2. A picture is shown in Figure 3, differing from the test as supplied by the makers. No reference is made to the test in the text.

glued into place and not used for testing purposes. Two of these pieces are triangles and two are somewhat in the shape of a diamond. On Figure 1, giving a picture of the test, these four pieces can be discerned. Our reasons for the omission of these

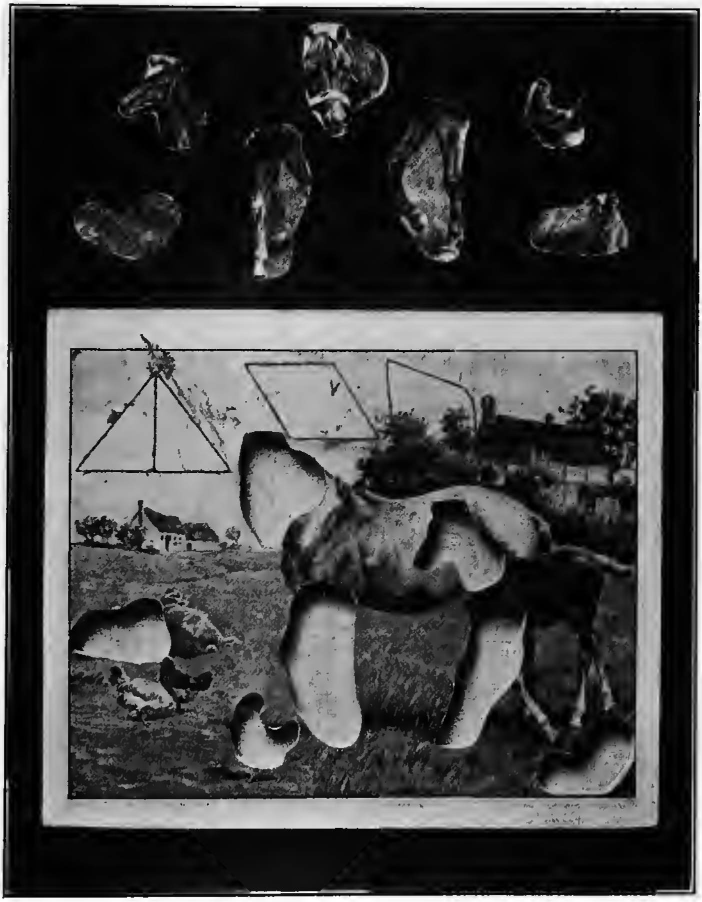


Fig. 1.—The Mare and Foal Test.

four pieces from the test were two. In the first place, they differ radically in nature from the other pieces and are decidedly more difficult. The rest of the pieces are not simple geometrical forms, but are more or less shaped according to the part of the animal or scene which they represent. These four pieces are simple geometrical forms and the shape does not in any way correspond to the picture pasted upon it. We found that younger children had great difficulty with these four pieces and that the insertion of them presented an entirely different problem from the insertion of the other pieces. In the one case the child may be guided by the picture on the cut-out as well as by the shape. In the case of these four pieces it is practically shape alone that is the determining factor in placing them correctly.

In the second place Test 6 (Triangle) presents the problem of inserting two pieces together to make a triangle, like the two pieces in the Mare and Foal Test. Having this other test (the Triangle Test), it would be useless to demand the same performance twice. It seemed wiser to us to demand this kind of performance as a separate test, in view of the fact that the type of performance required is radically different in nature from the insertion of differently shaped cutouts.

This test was suggested first by Healy,[2] and

[2] Healy, W. and Fernald, G. M.: *Op. cit.*, pp. 12-13.

without modification it has been used by Schmitt,[3] who gives results for 132 cases. These results are not comparable with ours, since our modification of the test has made it very much easier. Healy [4] has used these cases of Schmitt as tentative norms for the test, and his norms are also not comparable with ours.

(b) *Method.* The method of giving the test is simple. The board is placed in front of the child with the seven ~~nine~~ pieces scattered at the top, as in Figure 1 (page 27). The instructions are: "Put these pieces in the right places as quickly as you can, without making any mistakes." The stop watch is started and the time for the complete performance is taken. During the performance the examiner counts the number of errors. An error is any attempt on the part of the child to place a piece in a wrong space. If the child holds a piece over a space hesitatingly without bringing it down to touch the board, we have not counted this as an error. The child is allowed to work at the test for 5 minutes. If he fails to complete the test within 5 minutes the examiner should proceed to the next test. This limit of 5 minutes is put upon almost all of our tests for practical reasons. Otherwise, with some children the ex-

---

[3] Schmitt, C.: "Standardization of Tests for Defective Children," *Psychological Monographs,* Vol. xix, No. 3, Whole No. 83 (1915), p. 86 *et seq.* The plate showing this test corresponds to the board used by us.

[4] Healy, W.: The Individual Delinquent, Little, Brown and Co. (1915), Plate I, opp. p. 86, and p. 106.

amination would stretch over too long a period.

(c) *Record.* The record shows the time for the complete performance and the number of errors. D.N.C. (Did Not Complete) is recorded if the child fails to finish the test within the 5 minute limit.

Test 2. The Seguin Form Board

(a) *Description.* Although the writers gathered considerable data with a Seguin Form Board of their own construction, it was decided to abandon this in place of the standard Seguin Form Board in view of the large amount of data already gathered by other workers. It was also deemed advisable to incorporate into our group of performance tests a test that has already been standardized and that is already familiar to, and in the possession of, a large number of workers.

Sylvester's [5] standardization of the Seguin Form Board is the one chosen by the writers for inclusion in their group of tests, because the author has published his data in such form as to make possible the necessary calculations for their purposes.

Sylvester's [6] description of the board is as follows: "The ten geometrical figures, as nearly uniform in size as their variety of form will allow, are cut through an oak board 20 x 14 x 3/8 inches. This oak board is glued to a soft wood board of

[5] Sylvester, R. H.: "The Form Board Test," *Psychological Monographs,* Vol. xv, No. 4, Whole No. 65 (1913).

[6] Sylvester: *Op. cit.,* p. 1.

the same length and breadth, 5/8 inch thick. The result is a thick board of moderate weight with a hard oak surface in which the ten forms appear as shallow holes or recesses. About the edge is placed an oak strip, 1¼ x ¼ inches, fitting flush with the soft wood back and forming a ¼ inch raised edge about the oak surface. Corresponding to the ten recesses are ten walnut blocks, 7/8 inch in thickness, each of which fits loosely into its corresponding recess. The thickness being more than twice the depth of the recesses, the blocks can be easily grasped and removed. The board and the blocks are finished in their natural oak and walnut colors and the recesses are painted black. The whole is carefully finished in order to give it an attractive appearance—an important feature in a mental testing device. This description applies to what may be called the standard form board—the type now in most general use." Although this description of the form board used by Sylvester differs slightly from that of the Goddard Form Board, as manufactured by Stoelting, the writers are inclined to believe that the two boards are sufficiently alike to warrant the use of Sylvester's norms for Goddard's Board.[7] Of course, we cannot be cer-

[7] For other descriptions of the form board and for work done with it see:

Wallin, J. E. W.: "Age Norms of Psycho-Motor Capacity," *Journal of Educational Psychology,* Vol. vii, No. 1 (1916), pp. 17-24.

Norsworthy, N.: "The Psychology of Mentally Deficient Children," *Archives of Psychology,* No. 1 (1906), The Science tain that norms obtained by means of the Goddard Board would be identical with those obtained by Sylvester. The boards may vary slightly in ease or difficulty. We do not believe that any such variation, if it exists, can be very great in view of the great similarity in the age averages obtained by Goddard and by Sylvester.[8] The difference between any two averages for ages six to twelve inclusive is never greater than 3 seconds.

No detailed description of the form board, beyond what we have quoted from Sylvester, is necessary.

(b) *Method.* Sylvester's method of procedure in giving the test is to be recommended since we are using his norms. To quote:[9]

"The form board lies horizontally on a table, its lower edge even with the edge of the table

---

Press, pp. 25-26. In this work it is called the Block Test.

Young, H. A.: "The Witmer Form Board," *Psychological Clinic*, Vol. x, No. 4 (1916), pp. 93-111.

Goddard, H. H.: "The Form Board as a Measure of Intellectual Development in Children," *Training School Bulletin*, Vol. 9 (1912), pp. 49-52.

Pintner, R., and Paterson, D. G.: "The Form Board Ability of Young Deaf and Hearing Children," *Psychological Clinic*, Vol. ix, No. 8 (1916), pp. 234-237.

Wallin, J. E. W.: "Experimental Studies of Mental Defectives," *Educational Psychology Monographs*, No. 7, Warwick and York, Baltimore (1912).

Whipple, G. M.: Manual of Mental and Physical Tests, Vol. i, Warwick and York (1914).

[8] Sylvester: *Op. cit.*, p. 46, footnote 4.

[9] *Idem*, p. 44.

next to which the child stands. The table must be low enough to allow him to lean well over the board and to look down upon its center. The blocks are placed in three piles on the table next to the upper edge of the board, no block in the pile nearest its recess, the lozenge and the elongated hexagon not in the same layer, and the star in the lower layer. This is the arrangement at the beginning of each of three trials. The child is introduced to the test with no introduction concerning it except, 'Let us see how quickly you can put the blocks into place.' His first reactions and his behavior until he succeeds in getting the blocks into place or fails are carefully studied. After this first trial he is given any instruction necessary to make him understand where the blocks belong and that he is to replace them as quickly as possible. Then he is given a second and third trial, in which he is encouraged and urged in every way to make the best record of which he is capable. These last two trials are timed with a stop watch and the shortest of the two records is taken as the child's form board index."

In actual practice the writers have always taken a record of the three trials, and the shortest of the three trials has been used as the child's form board index.

(c) *Record.* A record of the time of the three trials is kept. For practical testing it does not seem necessary to keep a record of the number of errors. The time limit is 5 minutes.

Test 3. The Five Figure Board

(a) *Description.* The Five Figure Board, devised by Paterson, is a form board 1.2 centimeter thick, measuring 57.4 x 20.3 centimeters, with five cut-outs. The length of the cut-outs varies from about 7 to 14 centimeters. It was devised with the idea of making a form board somewhat more complex than the Seguin Form Board. For this reason each one of the cut-outs is divided into two pieces, with the exception of one which is divided into three pieces, whereas in the Seguin Board each one of the cut-outs is one entire piece. The cut-outs in the Five Figure Board are an oval, a circle, a square, a hexagon and a cross. The cross is divided into three pieces. This is an original test and no previous work has been done with it. Our results on this test seem to show that it has answered admirably the purpose for which it was designed, namely, to serve as a more difficult form board of the Seguin type.

(b) *Method.* The board is placed in front of the subject, as shown in Figure 2 (page 35). The square is at the subject's left and the oval at his right. The pieces are scattered around at the top of the board somewhat as in the figure. The two parts of the oval are at the farthest ends, next to which, proceeding inwards, are the two parts of the hexagon, then the two parts of the square, and in the center are the three parts of the cross and the two parts of the circle.

The experimenter says to the child: "Put this together as quickly as possible."

(c) *Record.* A record of the time and number of errors is kept. An error is any attempt on the part of the child to put a piece into a wrong hole, or to put a piece in a wrong position in the right hole. If the child hesitates, holding a wrong piece

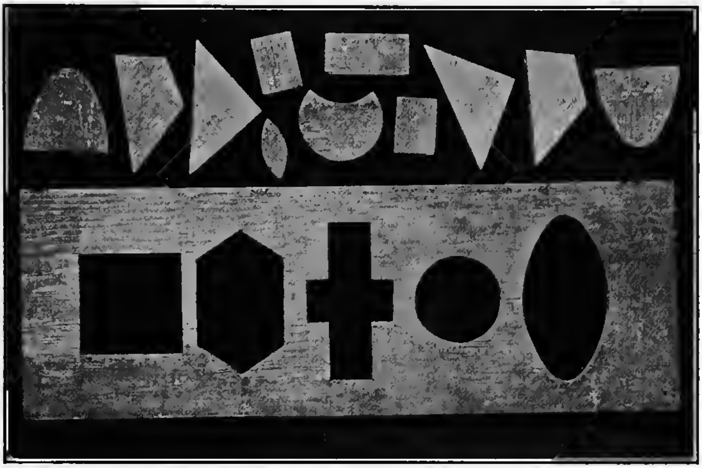


FIG. 2.—The Five Figure Form Board.

above a hole without touching the hole, an error is not counted. The time limit is 5 minutes. If not completed within this time limit D.N.C. is recorded.

TEST 4. THE TWO FIGURE BOARD

(a) *Description.* The Two Figure Board was devised by Pintner to be a more difficult board than either the Seguin or the Five Figure Board. As a matter of fact, the results seem to indicate that it is slightly easier than the Five Figure Board, inasmuch as it is ordinarily done in a somewhat shorter time, and also as fewer children fail to

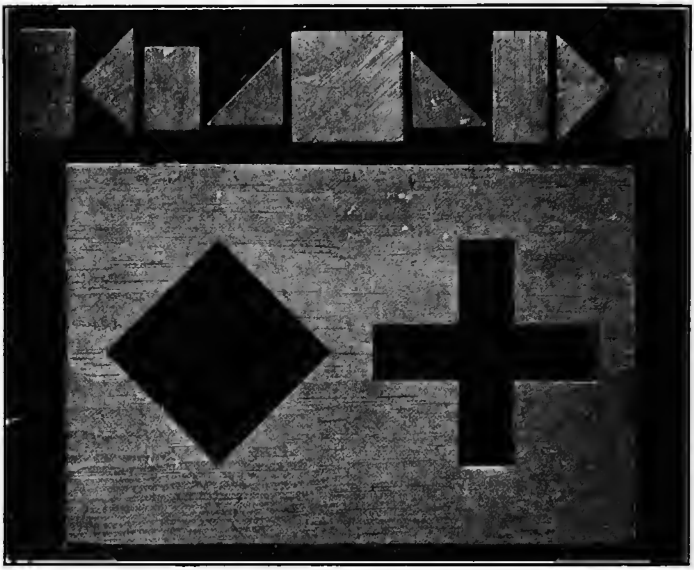


FIG. 3.—The Two Figure Form Board.

complete the test. It is a board 1.4 centimeter thick, measuring 38.3 x 25.4 centimeters, and having two cut-outs, a square and a cross. The cross consists of four pieces, two measuring 7.5 x 2.7 centimeters and two 5.6 x 3.7 centimeters. The square is filled by another square measuring 7.5 x 7.5 centimeters, and it fits into the larger square, with room for the remaining pieces only when put in so that the sides of the smaller square form diagonal lines from the middle points of the sides of the larger square. The remaining sections are filled by four small triangles measuring 7.2 x 5 x 5 centimeters. This again is an original board on which no previous work has been reported.

(b) *Method.* The board is placed before the child, as in Figure 3 (page 36), with the cross at the right-hand side, and the pieces scattered at the top. The arrangement of the pieces is as in the picture, the square in the center with the little triangular pieces separating the small rectangular pieces. The examiner says to the child: "Put this together as quickly as you can."

(c) *Record.* A record of the time and number of moves is kept. A move is any attempt, right or wrong, to place a block in a space. The fewest number of moves possible is nine. The time limit is 5 minutes.

TEST 5. THE CASUIST FORM BOARD

(a) *Description.* This test was devised by Knox,[10] and we copied the test from his description before Knox's tests were sold by the dealers, and it differs from the one sold by them. Our board measures 50 x 25.7 centimeters and is 1.5 centi-

---

[10] Knox, H. A.: "A Scale, Based on the Work at Ellis Island, for Estimating Mental Defect," *Journal of the American Medical Association,* Vol. lxii (March 7, 1914), pp. 741-747.

meter thick. There are three circles of varying sizes, having diameters of about 13, 11 and 7.7 centimeters respectively, and a fourth aperture in the shape of an elongated oval with the sides parallel part of the way. The two larger circles are each

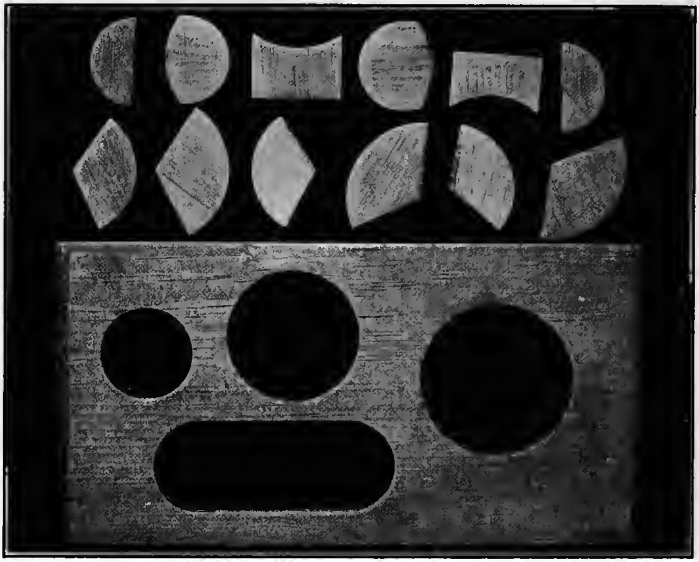


FIG. 4.—The Casuist Form Board.

cut up into three equal segments, while the smaller circle is cut into two equal segments. The elongated oval is cut into four pieces, two more or less circular end pieces, and two middle pieces. These are difficult to describe and can best be understood from Figure 4.

Knox places this test among his twelve-year-old tests in his year scale of tests. To quote: "Do

Casuist Test in 5 minutes with sensible mistakes, if any." No description of "sensible mistakes" is given. This is obviously a very rough standardization.

As far as the writers are aware no other mention of this test has been made in the literature of mental testing. We have not been able to find any other norms or attempted standardizations. As we shall see below, in the chapter on standardization, the median time for our twelve-year-olds to complete the test is only sixty-six seconds. None of the twelve-year-olds fail to do the test within 5 minutes. The median number of errors is four. We did not consider it practicable to make a distinction between sensible and not sensible errors. Further inspection of our results for this test (see Tables 8 and 9, and Graphs 13 and 14, pages 112-114) would seem to show that the Casuist Test performed within five minutes, with the allowance of a few sensible errors according to Knox's description, is a very easy twelve-year-old test. We are inclined to believe that, used in this way, it might be passed by much younger children. Seventy-five per cent of our seven-year-olds complete the test within five minutes, although the average number of errors for the seventy-five per cent is thirty, which would probably not fulfill Knox's requirement of "sensible mistakes."

(b) *Method.* The board is placed before the child as in Figure 4, with the pieces scattered around in more or less definite order at the top of the board, as shown in the figure. The three segments of the large circle are placed alternately with the three segments of the smaller circle in a row at the top of the board. In the row above these the remaining pieces are placed with the two halves of the small circle at each end and the four remaining pieces between, as shown in the figure.

The examiner says to the child: "Put these pieces together as quickly as possible."

(c) *Record.* A record of the time and number of errors is kept. The time limit is 5 minutes.

Test 6. The Triangle Test

(a) *Description.* This is a test devised by Gwyn and described by Knox.[11] Our board is the standard one furnished by the dealers. It measures 17 x 12.8 x 1 centimeters. The size of the rectangle at the top is 4.9 x 6 centimeters. The triangle is about 6 centimeters high, with a base measurement of about 9.5 centimeters. The rectangle is cut diagonally into two pieces, and the triangle is cut into two by a vertical section from the apex to the middle point of the base line. This results in four triangular pieces of exactly the same size.

Knox does not pretend to have arrived at a standardization of this test. He merely classifies this among his "Make-up Tests for Adults," and his requirements are: "Put the four pieces into

[11] Knox: *Op. cit.*

Gwyn's triangle three times in forty-five seconds." We have not given the test three times to our subjects, so that comparison with this requirement of Knox is impossible. No results of work done with this test have come to our notice, so that we have nothing with which we can compare our norms.

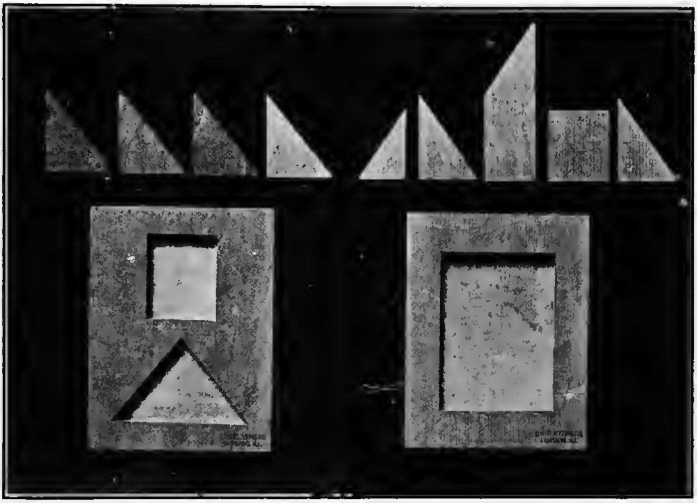


FIG. 5.—The Triangle Test (left). The Diagonal Test (right).

(b) *Method.* The test is placed before the subject as in Figure 5 (page 41), with the four triangles at the top, the right angle at the left and all the triangles pointing the same way.

The experimenter says to the child: "Put this together as quickly as possible."

(c) *Record.* A record is kept of the time and number of errors. The time limit is 5 minutes.

Test 7. The Diagonal Test

(a) *Description.* The Diagonal Test was devised by Kempf and is described by Knox,[12] who places this, along with the previous test, among his "Make-up Tests for Adults." His requirements are: "Put the pieces into Kempf's diagonal inside of three minutes." We know of no other mention of this test and therefore can make no comparison with the results obtained by us.

Our board is the one supplied by the dealers. The outside measurements of the frame are 16.5 x 12.7 centimeters, the frame being one centimeter thick. The inside measurements of the frame into which the pieces fit are 11 x 8 centimeters. The cut-outs can be seen on Figure 5 (page 41). They may be described as two larger right-angle triangles, one small right-angle triangle, one rectangle 4.3 x 5 centimeters and one large quadrilateral, from the top part of which the smaller triangle has been cut. There are two or three possible ways in which these pieces may be fitted into the frame, although the small triangle always goes with the quadrilateral, filling in one-half of the space, and the square and two triangles the other. The fact that there are different ways of fitting in the pieces, while making the test more difficult, introduces an element of chance. The different ways of fitting the pieces in do not seem to be all equally difficult; therefore, if a child happens by chance to start with

[12] Knox: *Op. cit.*

one of the easier ways, he has an advantage over the child who by chance makes a different, though equally rational, first move. This kind of test we have designated as belonging to the puzzle type of test.

(b) *Method.* The pieces are scattered at the top of the test in the order shown in Figure 5. We doubt whether the arrangement of the pieces in presenting the test is of great significance, so long as no two pieces that belong together in the test are placed in juxtaposition. Our arrangement, however, has generally been, beginning at the left-hand side facing the test: first, the small triangle, then one of the larger triangles, then the long rectangular piece, then the small rectangle, and lastly the second of the larger triangles. It is well for the experimenter to get in the habit of placing the blocks in a certain order. It saves time and thought. After arranging the test, the instructions to the child are: "Put these together as quickly as possible."

(c) *Record.* A record of the time and number of errors is kept. The time limit is 5 minutes. An error is counted when a piece is placed in such a position that would make the filling in of the rest of the pieces impossible. Owing to the number of different ways in which the pieces may be arranged, errors are relatively infrequent among the first few moves. It is well for the experimenter to study all possibilities of this test before taking permanent records.

TEST 8. THE HEALY PUZZLE "A"

(a) *Description.* This is a test first described by Healy and Fernald in 1911.[13] A picture of the test is given in Figure 6 (page 45). Our test is the one sold by the dealers. The outside measurements of the frame are 15.3 x 12.7 centimeters; the inside measurements of the frame 10.3 x 7.9 centimeters. There are five rectangular pieces to be fitted into the frame. The measurements of our pieces are as follows: the largest 7.5 x 3 centimeters, the next largest 7.2 x 2.5 centimeters, the next 5.1 x 3.4 centimeters, and the two small pieces, both the same size, 3.8 x 2.5 centimeters.

Healy gives Freeman credit for making the first sketch of this test. Healy and Fernald, however, seem to have altered Freeman's original sketch and devised the test as we now know it. Their description of the psychological character of the test is as follows: "This test brings out perception of relationship of form and also the individual's method of mental procedure for the given task—particularly his ability to profit by the experience of repeated trials, in contradistinction to the peculiar repetition of impossibilities characteristic of the subnormal and feeble-minded groups." The method of scoring recommended by Healy and Fernald consists of noting (1) the time; (2) number of moves; (3)

[13] Healy, W., and Fernald, G. M.: "Tests for Practical Mental Classification," *Psychological Monographs,* Vol. xiii, No. 2, Whole No. 54 (1911), pp. 14-15.

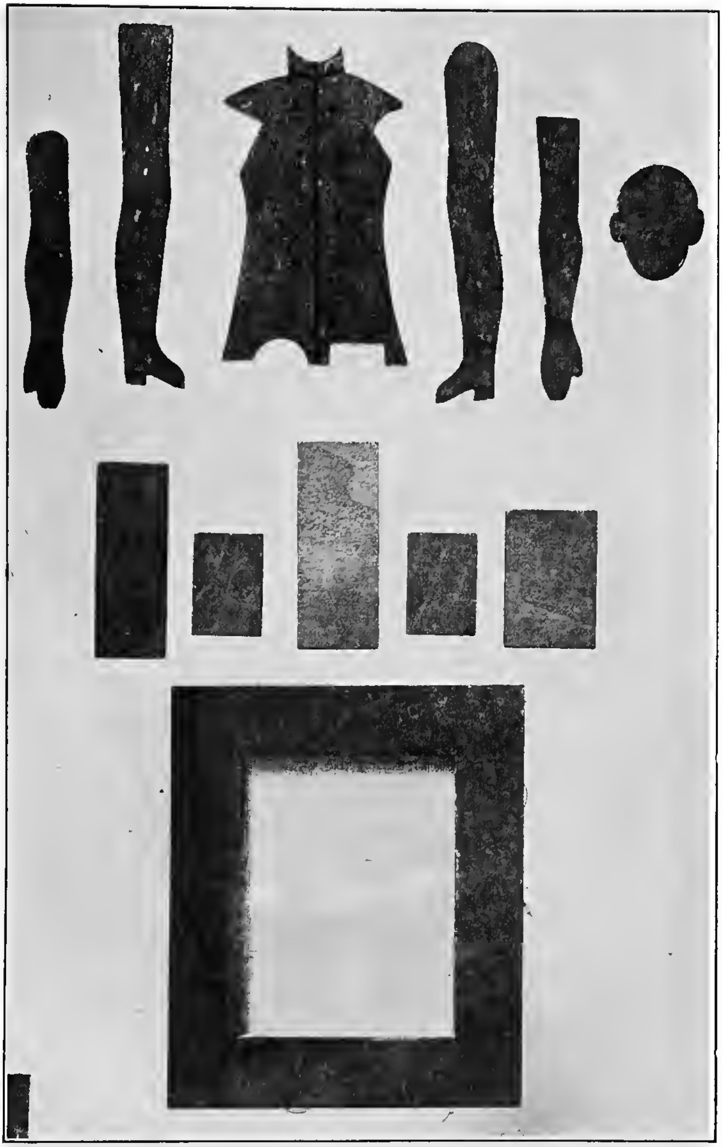


FIG. 6.—The Manikin Test (top). Healy Puzzle "A" (bottom).

number of impossible moves; (4) repetition of such obvious impossibilities. A record of one case is given.

No further standardization either of procedure or of results was given by the authors at that time. Since then, however, we have at least four studies dealing wholly or in part with this test, in addition to the norms given by Healy in a later work.

Schmitt's[14] standardization, which appeared in 1915, shows the results for 154 children. Her time limit is ten minutes. She gives the average time and number of errors of the cases distributed according to grade, and also a division of the results into planned, trial and error, and chance methods distributed according to grade and also according to age. More emphasis is placed upon the method of doing the test than upon the time. No explanation of what is meant by the three methods (planned, trial and error, and chance) is given. Although the names of the methods are more or less self-explanatory, yet anyone who has had some experience with this test will recognize at once that many performances would be very difficult to classify. Most performances, we venture to suggest, are a mixture of trial and error and planning. Many children start out with a good move by chance and complete the performance by trial and error or by planning. The determination by the

---

[14] Schmitt, C.: "Standardization of Tests for Defective Children," *Psychological Monographs,* Vol. xix, No. 3, Whole No. 83 (1915), pp. 93-96.

examiner of the method of a specific performance is far too subjective a procedure to give any reliable measurement. It may be that with practice and great care an examiner might reach a high degree of uniformity in his classification into the three methods named above; but even then, owing to the subjective nature of this kind of evaluation, his results would be absolutely worthless for purposes of comparison with the results of other workers. Furthermore, we may say that, on the whole, the planned method will be done more quickly than the trial and error method and will be done in fewer moves; and further, that, with some exceptions, the chance method will take the longest time and require the greatest number of moves. This is borne out by Schmitt's results, for the average time of the kindergarten children is 3 minutes and 10 seconds, the longest time period for any one of the groups, and it is among this group that the largest percentage of cases belonging to the chance method occurs. It seems best, therefore, to take the time and the number of moves as measures of the performance. These are perfectly objective and measurable values. In so doing we may occasionally overestimate a performance that is due to chance, and such cases undoubtedly occur. To that extent the test is of the puzzle variety, and to that extent it is open to criticism.

Healy [15] in a later statement of norms says that:

[15] Healy, W.: The Individual Delinquent, Little, Brown and Co. (1915), p. 107.

"No normal person over 8 or 9 years should fail to do it in 5 minutes," but we find this doubtful in view of the fact that 11 out of 117 of our ten-year-olds, 4 out of 105 eleven-year-olds, 4 out of 88 twelve-year-olds and 2 out of 44 fourteen-year-olds failed. Failure at these ages is certainly very poor. All of these cases lie below the 10 percentile for their age and might give rise to a suspicion of mental defect, but this would have to be corroborated by the use of many other tests. Healy, like Schmitt, also lays emphasis upon the method, but leaves evaluation of the performance according to method entirely a subjective matter.

Hall's [16] work with this test gives the results with 180 cases, ranging from age seven to age twelve. She shows for each age the per cent successful, the average time and the average number of moves. A comparison of these average times with our medians has little value, since Hall's average time at each age is merely for those completing the test successfully. Our medians, of course, take into account those who failed to complete the test within the 5-minute time limit. We give on page 49 our medians and Hall's average time for ages seven to twelve inclusive.

The surprising thing about this comparison is that our medians are generally lower than Hall's averages, notwithstanding the fact that our medians

[16] Hall, G.: "Eleven Mental Tests Standardized," *Eugenics and Social Welfare Bulletin No. V,* State Board of Charities, New York (1915), pp. 26-32.

Age	Hall's Average	Our Median
7	126.8	131
8	133.5	117
9	95.9	86
10	75.5	70
11	48.5	54
12	41.8	46

are influenced by the number of failures at each age. It may be that Hall's children were, on the whole, somewhat below the normal child, if our cases can be said to be sufficiently numerous to approach the normal. Or it may be that our children did somewhat better because of previous practice with the other tests of the same nature that preceded this test in our series. We doubt whether this would make a great difference.

The per cent successful in completing the test at each age for our cases and for Hall's cases is as follows:

Age	Hall's Cases	Ours
7	50.0	58.6
8	53.3	69.4
9	96.6	83.6
10	90.0	90.8
11	90.0	96.3
12	93.3	95.5

It will be noted that at all ages, with the exception of age nine, our percentages are higher than those of Hall, showing a somewhat better performance of our cases, which is what would be expected from the comparison of the time. Hall's nine-year-olds, as noted in other tests, seem to be better than average nine-year-olds, because they frequently make better records than her ten-year-olds.

Another study of this test was made by Bruckner and King.[17] They give the results for 90 eight-year-olds and for 59 ten-year-olds. The authors lay stress upon the value of the test as a learning test and give three trials. It is the first trial only that interests us here. The median time for eight- and ten-year-olds offers an opportunity for comparison with our medians. Nothing is said about failures and none are recorded in the tables. The median for eight-year-olds is 140 seconds, while ours is 117; the median for ten-year-olds is 69 seconds, while ours is 70 seconds. The eight-year-old median differs considerably from ours, and we are unable to explain this difference. Bruckner and King's eight-year-old median is worse than our seven-year-old median, though better than our six-year-old. The only suggestion we have to offer as an explanation or partial explanation of this difference is what we have said above as to the possibility of practice with similar material in tests which preceded this test in our series. But if this is really

[17] Bruckner, L., and King, I.: "A Study of the Fernald Form Board," *Psychological Clinic*, Vol. ix, No. 9 (1916), pp. 249-257.

an influential factor, then it ought to be shown in the ten-year-old results, and this is not the case.

Bronner [18] reports some results with this test, appearing as a sort of criticism of the work referred to above, in which she questions the value of this test as an age test, although her reasons for this are not clear. She notes the fact that Terman has placed it in his scale. In regard to the test she says: "We believe that it is better adapted to throw light upon ability along certain lines regardless of age." No indication is given as to what kind of ability is meant. Even if this is so, we can at the same time standardize this ability for each age. It is her point of view in regard to standardization, however, to which we would take most objection. She says: "When, however, the standardization of a test is based upon data gathered in schools the mentality of the children cannot be determined and there might conceivably be included feeble-minded and greatly retarded as well as the normal. A few such extreme cases would alter averages considerably." [19] It is true that they would alter averages considerably and that is why it is better to use the median as the standard. But, even so, the inclusion of these cases in a large enough group is greatly to be desired if a complete standardization is to be attained. Why is there such a horror of including a

[18] Bronner, A. F.: "'Construction Test A' of the Healy-Fernald Series," *Psychological Clinic,* Vol. x, No. 2 (1916), pp. 40-44.

[19] Bronner: *Op. cit.,* p. 43.

feeble-minded case now and then, and no horror of including the abnormally bright child? The latter will play havoc with averages and medians (if it is regarded as havoc) to just the same extent as will the feeble-minded.

Then again, why should we know the mentality of the children we are testing? What we want to arrive at is the ability of children of a specific age on a certain test, and we ought not to be influenced by other estimates of their ability. Our sole endeavor must be to get a fair sampling of cases at each age. If we standardize according to mental age arrived at by any scale, we are moving in a circle and presupposing that our determination of mentality is accurate and final. This question we have discussed more fully in Chapter VIII.

Bronner's medians are given for boys and girls separately and her description of the cases as normal children "of good innate ability" would lead one to suppose that they are above normal for any age as a whole. Her age groups range from eleven to seventeen inclusive. The comparison of the medians with ours for ages eleven to fourteen is as follows:

	Bronner		Ours
Age	Boys	Girls	
11	45	61	54
12	41	38	46
13	54	47	38
14	30	24	55

Our fourteen-year-old median is based upon only 44 cases and we do not feel that it is very reliable, since the sampling of fourteen-year-olds is scarcely a fair sampling of fourteen-year-olds in general.

(b) *Method.* The test is placed before the child, as in Figure 6 (page 45), the three large pieces being separated from each other by the two small pieces of equal size. The examiner says to the child: "Put this together as quickly as you can."

(c) *Record.* A record of the time and the number of moves is kept. The time limit is 5 minutes.

Test 9. The Manikin Test

(a) *Description.* This test was devised by Pintner and is described here for the first time. It was designed as a test for young children. It demands the same kind of ability as the Feature Profile Test. The scattered fragments suggest some kind of a complete whole and the child has to synthesize these scattered impressions and plan to reach a definite end.

The test represents the conventional figure of a man; such as is often drawn by children, i.e., a body, two arms, two legs and a head. The pieces are shown in Figure 6 (page 45). The figure is cut out of wood which is about 7 millimeters thick. The sizes of the pieces are: body, 11.6 x 6.7 centimeters at the widest points; legs about 13.2 centimeters long; arms about 10.4 centimeters long; head about 4.2 centimeters long. The wood is varnished on one side and a few lines are painted to represent eyes, nose, mouth, cuffs, shoes, coat, collar and buttons.

The places where the arms and legs fit into the body are not the same shape either for both arms or for both legs, one being rectangular and the other circular in each case. This device was adopted to add to the difficulty of an exact performance. It has proved to be an excellent test, for it permits of a definitely objective method of scoring.

No results on this test have been published up to the present time.

(b) *Method.* The test is placed before the child as in Figure 6. It is to be noted that the leg with the rectangular end is at the opposite side of the body from the place where it fits. Similarly with the other leg and the two arms. In other words, to make a complete performance the child has to bring the leg and arm at the right over to the left side, and the leg and arm at the left over to the right side.

The experimenter says to the child: "Put this together as quickly as you can." *Do not say anything about its being a man.*

(c) *Record.* Although the time of performance was taken for this test, we have not seen fit to make it a measure of ability for the test. Instead, we have devised a method of scoring for the end result attained by the child. This system of scoring is as follows:

A complete performance, absolutely accurate—5 points.

One or both arms up or out, i.e., not exactly fitting in the joints—4 points.

One reversal, i.e., right arm for left arm and vice versa, or right leg for left leg—3 points.

Two reversals, i.e., both arms and both legs reversed—2 points.

Legs or arms interchanged or arms at sides, or any other result which looks like a man—1 point.

Failure to see that it is a man—0 points.

The experimenter either makes a note of the position or merely notes the score, if he is familiar enough with the method of scoring.

The time limit is five minutes.

Test 10. The Feature Profile Test

(a) *Description.* This test was devised by Knox and Kempf and has been described by Knox.[20] The author says: "It is our highest and most difficult performance test and yet it is eminently fair, because everyone has seen a human head; the subjects are told 'This is a head.'" Our instructions to the subject are somewhat different, as will be seen

[20] Knox, H. A.: "A Scale, Based on the Work at Ellis Island, for Estimating Mental Defect," *Journal of the American Medical Association,* Vol. lxii (March 7, 1914), pp. 741-747.

below. Our subjects are not told that it is a head.

Our test was a copy from Knox's description. It is made of wood about 1 centimeter thick, and measures about 21 x 17 centimeters at its greatest measurements. Like the Manikin Test described above, it demands that synthetic ability of seeing the parts of a whole and of putting these together, a kind of ability which seems to be one of the essential factors in general intelligence.

Knox places this test among the group of tests headed "At from Thirteen Years Onward," and his time limit is 10 minutes. Our results show that with a time limit of only 5 minutes, 16 out of 68, or 24 per cent, of the thirteen-year-olds fail. That means that 76 per cent of the thirteen-year-olds pass the test and, therefore, it may be called a thirteen-year-old test, with a time limit of 5 minutes. A time limit of 10 minutes would make the test a very easy thirteen-year-old test.

Some results of this test with feeble-minded children have already been published by us.[21]

(b) *Method.* The test is placed before the subject as in Figure 7 (page 57). The three pieces forming the face or profile are separated from each other by the four pieces forming the ear. These are placed at the top of the head, which is placed in correct position in front of the subject. The experimenter says: "Put this together as quickly

[21] Pintner, R., and Paterson, D. G.: "The Factor of Experience in Intelligence Testing," *Psychological Clinic,* Vol. ix, No. 2 (1915), pp. 44-50.

as you can." If the subject changes the position of the head during the manipulation of the pieces, the examiner is not permitted to place it in its correct position again, or to help the subject in any

FIG. 7.—The Feature Profile Test.

way. *The subject is not told what the test represents.*

(c) *Record.* A record of the time alone is kept. The time limit is 5 minutes.

Test 11. The Ship Test

(a) *Description.* This test was devised by Glück and mentioned by Knox.[22] He places this test among the group headed "At from Thirteen Years Onward." His time limit again is ten minutes. No results are given for this test by Knox. Our results show about 60 per cent of the thirteen-year-olds making a perfect performance within a time limit of 5 minutes. It is doubtful whether a longer time limit would lead to better results. If such is not the case, then Knox's standard would appear to be rather difficult for thirteen-year-olds. A perfect performance is not made by 75 per cent of the cases in any of our age groups.

Our test is the one supplied by the dealers. It consists of a frame 1 centimeter thick, of which the outside measurements are 25 x 16.2 centimeters and the inside 21.4 x 12.5 centimeters. Into this frame there can be fitted 10 pieces, each measuring 21 x 6.2 centimeters, which when properly fitted together form a ship.

This test differs from the Mare and Foal, the Manikin or the Feature Profile, inasmuch as all the pieces are of the same size and shape. The size

[22] Knox: *Op. cit.*

and shape of the pieces give no help in determining the correct position. The subject must be guided solely by the picture he is trying to make.

The ultimate picture constructed by the subject seems to the writers to be the most significant feature of the test. Obviously this will vary all the way from a correct picture to an impossible one. To grade the various possible combinations of the test, the following scheme was adopted: A score of 20 is allowed for a perfect performance, i.e., 2 points for each piece in its correct position. A score of one is allowed for each of the lower or upper pieces, if placed in the lower or upper portion of the frame, i.e., the "water" pieces at the bottom and the "sky" pieces at the top. If a child puts the five "water" pieces at the bottom and the five "sky" pieces at the top, he receives a score of 10. In addition to this, a score of one is given to any piece that is in correct position in relation to any other piece, i.e., any two or more adjoining pieces correct each receive a score of one. The maximum score is 20, i.e., ten for "sky" and "water" pieces, and 10 for each piece being next to its correct adjoining piece.

(b) *Method.* The test is placed before the subject as in Figure 8 (page 60). The ten pieces are always placed in the same position, in order to eliminate any possible advantage that might accrue to some children owing to a helpful arrangement that might now and then result from a chance arrangement of the pieces. To aid the examiner in quickly arranging the pieces, they are numbered consecutively on the backs and upper edges

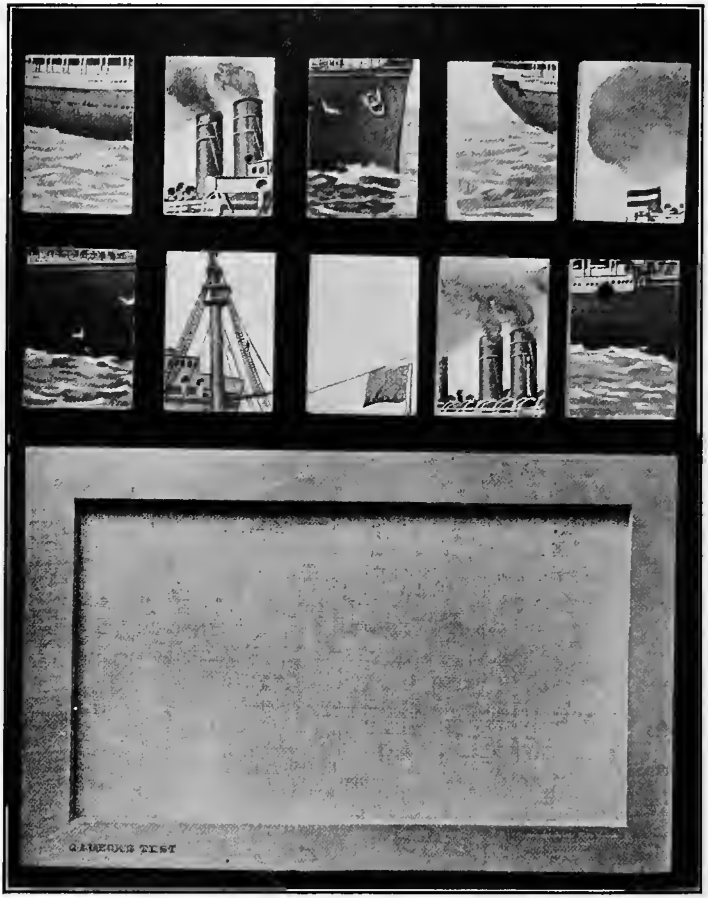


Fig. 8.—The Ship Test.

of the pieces, i.e., so that he can see the numbers when sitting opposite the subject. The numbers begin with the block at the upper left hand corner on the picture and continue consecutively along to the right hand block, and similarly with the second row. The examiner says to the child: "Put this together as quickly as you can."

(c) *Record.* A record of the time and score was kept, but we have only used the score in the evaluation of the test. There is no definite time limit. The child is urged to continue as long as he shows any signs of completing the test. As soon as he indicates in any way that he has finished, the test is taken away and scored. The examiner must not show in any way his disapproval of any performance, but rather his approval. The child must not be given an opportunity to make any changes, if the examiner has suggested by his attitude that the performance is incorrect. During the actual performance of the test the child may, of course, make as many changes as he wishes.

Test 12. The Picture Completion Test

(a) *Description.* A full description of the test with a critical survey of work already done with the test, together with norms for each age, is given by Pintner and Anderson [23] and need not be repeated here.

(b) *Method.* The test is placed before the child

[23] Pintner, R., and Anderson, M. M.: "The Picture Completion Test," *Educational Psychology Monographs;* Warwick and York, Baltimore.

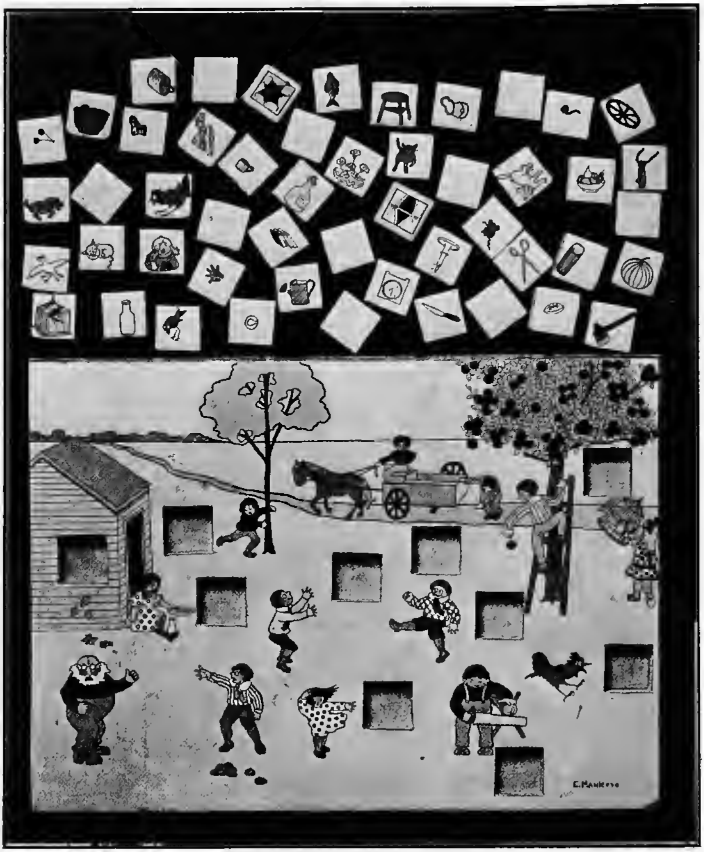


FIG. 9.—The Picture Completion Test.

as in Figure 9 above. For the method of procedure see Pintner and Anderson's monograph, referred to on page 61.

(c) *Record.* The method of scoring devised by Pintner and Anderson has been followed.[24] They

[24] Pintner, R., and Anderson, M. M.: *Op. cit.*

do not prescribe any time limit, but suggest that a time limit of 10 minutes be adhered to, for practical reasons, when a subject has a whole series of tests to perform. Very few children will require more than 5 or 6 minutes.

Test 13. The Substitution Test

(a) *Description.* This test was reported by Woodworth and Wells.[25] They give results for eleven adults, showing the average times for the first half, the second half and the whole blank. The average time for the first half for this group is 79.6 seconds, which is, as would be expected, lower than the average time for any of our groups of children.

A picture of the test is shown in Figure 10 (page 64).

We have used only the upper half of the test sheet, because of the practical necessity for limiting the length of time taken with each test in a series of tests.

(b) *Method.* The sheet is placed before the child and his attention is called to the blank key at the top. In Figure 10 the key has been filled in. The examiner says: "I will put a number in each one of these little figures and I want you to copy the same number in the same figure. What

[25] Woodworth, R. S., and Wells, F. L.: "Association Tests," *Psychological Monographs,* Vol. xiii, No. 5, Whole No. 57 (1911), pp. 53-55.

number will you put in this figure?" (pointing to one of the figures on the test blank). If the child answers, "The same as in that figure," and indicates the right one in the key, the examiner then repeats the question with other figures. If not, the examiner tells the child and continues until he is reasonably sure that the child understands.

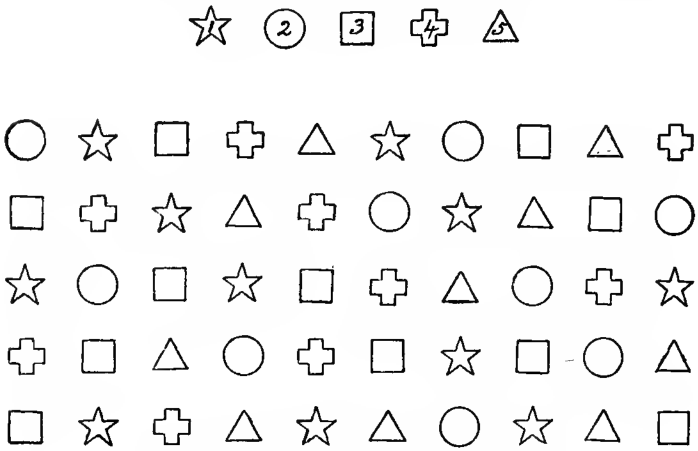


FIG. 10.—The Substitution Test.

The examiner then fills in the key without allowing the child to see what numbers he puts in the key, and says: "Now fill these numbers in the right figures beginning here and working along each one of the lines."

In giving this test to children who do not understand English or to deaf children, a short practice sheet with a key in which letters are inserted instead of figures is used. By pointing and by showing the child what to fill in, we believe we attain the same results by way of explanation as in the case of oral instructions.[26]

(c) *Record.* A record of the time is taken. The blank is then corrected and the number of errors recorded. The score for the performance is arrived at by adding to the time an additional penalty for each error. This penalty is determined by the time taken for the whole test; each error is counted 1/50 of the total time for the test. The theory is that, if the child were given an opportunity to correct his errors, the actual time for correcting (not finding) them would be about the time taken to fill in one figure. If we divide the total time by 50 (i.e., the number of figures to be filled in), we arrive at the average time for filling in one figure. This is then multiplied by the number of errors and the resulting value is added to the total time. The penalty for an error varies with the time taken for the test. A high score is poor and a low score is good.

Test 14. The Adaptation Board

(a) *Description.* This test was devised by Goddard and was described by him,[27] and norms for

[26] Pintner, R., and Paterson, D. G.: The Psychology of the Deaf. To appear shortly.

[27] Goddard, H. H.: "The Adaptation Board," *Proceedings of the Washington Meeting of the Psychological Associations, Psychological Bulletin,* Vol. ix (1912), p. 79.

certain moves of the test for normal and feeble-minded children have been published.[28] Goddard's Board measures 22 x 28 centimeters, with holes measuring in diameter 6.3 and 6.5 centimeters. Our board measures 22 x 25 centimeters and is 0.5 centimeter thick and has four holes in it, three of them being 6.8 centimeters in diameter and the fourth 7 centimeters in diameter. A wooden block with a handle fits the large hole exactly.

(b) *Method.* The examiner takes the board in his left hand and, with the right hand holding the block, shows the child that it will fit into the larger hole but not into any of the other holes. The board is so held that the large hole is at the examiner's upper right hand corner. The child is then given the block and the examiner says: "Put it into the right hole." If the child fails, he is shown how to do it. When this has been done, the examiner says: "Watch closely." The board is now turned over in such a way that the large hole at the upper right hand corner approaches the child in turning and rests at the examiner's upper left hand corner. As before, the child is told to put the block in the right place. Again, if the child fails, he is shown where the right hole now is. The examiner then turns the board over towards the child so that the large hole occupies the position at the examiner's lower left hand corner. The child reacts as before. The

[28] Goddard, H. H.: "The Adaptation Board as a Measure of Intelligence," *Training School Bulletin,* Vol. xi, 10 (1915), pp. 182-188.

next move is to turn the board so that the large hole occupies the lower right hand corner. For the last move, the examiner holds the board at the top right hand corner with the right hand and the bottom left hand corner with the left, and turns the board toward the child diagonally so that the large hole rests finally at the upper left hand corner.

Each move of the board takes about ½ second. It is a steady movement and not a hurried procedure.

(c) *Record.* The number of moves correctly made is recorded.

TEST 15. THE CUBE TEST

(a) *Description.* This test was devised by Knox [29] and first described by him. As used by him there were five different movements or lines, which were placed at various ages in his scale.

Pintner's [30] modification and expansion of the Cube Test is the one used here. The lines devised by Pintner are as follows:

A	1234	C	1432	G	13124
X	12343	D	1423	H	143124
Y	12342	E	13243	I	132413
B	1324	F	14324	J	142341

[29] Knox: *Op. cit.*

[30] Pintner, R.: "The Standardization of Knox's Cube Test," *Psychological Review*, Vol. xxii, No. 5 (1915), pp. 377-401.

Many of the original data collected by Pintner, together with additions, have been used in arriving at the norms established by us.

The material required is five blocks of the same color and size. We have, in general, made use of the Binet black cubes. Since the work on this test was started, Knox has devised different material, namely, four cubes of different colors mounted on a baseboard, and this is the material supplied by the dealers under the name of Knox Cube Test. Needless to say, this difference in material may lead to a radical difference in results and should not be used in this scale of performance tests, if our norms are to be used.

(b) *Method.* The four cubes are placed on the table in front of the subject at a distance of about two inches apart. "The examiner holds the fifth cube in his hand. He says to the subject: 'Watch carefully, and then do as I do.' He then taps the blocks with the fifth cube in a certain definite order and at a certain definite rate (about one tap per second), always beginning with the cube at the child's left or the examiner's right, if he is facing the child. He then lays the fifth cube down in front of the child equidistant between the third and fourth cube, but nearer to the child, and says: 'Do that.' . . . If we number the blocks the different combinations will be readily understood, and the following diagram should make absolutely clear their position with regard to the subject and the examiner (if he is facing the subject)."

THE TESTS

SUBJECT

4 3 2 1

EXAMINER [31]

(c) *Record.* A record of the number of lines passed or failed is kept. The examiner continues as far as possible with the child, always continuing with at least three lines after the child fails, and in many cases with more than three lines if there seems to be any possible chance of the child's succeeding in additional lines.

---

[31] Pintner, R.: *Op. cit.,* pp. 377-378.

CHAPTER III

STANDARDIZATION OF THE TESTS

In the history of mental tests the problem of standardization is one that is being emphasized more and more. The specific manner in which this problem of standardization has developed is due to the growth of scales of intelligence. It was owing to the fact that the scales were not proving themselves as accurate measures as critical workers demanded, that the question of standardization came to the front. It was the connection with scales of intelligence that made the question of standardization center around the correct placing of tests at specific ages, since the first scales of intelligence were age scales. But the problem of standardization has now advanced far beyond this specific question of the right placing of a test at a certain age for the use of this test in an age scale.

In general the question of standardization divides itself into two parts: (1) the standardization of procedure, and (2) the standardization of response, or the accumulation of sufficient results so that a specific response may be interpreted in the light of previous results, with a tolerable degree of certainty that the results are sufficiently numerous to warrant a generalization for the age or group of individuals in question.

The first part, i.e., the standardization of procedure, is so obvious as to make extended discussion unnecessary. Our standard method of procedure in regard to all of the tests discussed in the present volume has been laid down in Chapter II. Needless to say, our results must be interpreted in the light of that method of procedure, and results obtained by workers who do not follow strictly the procedure there laid down cannot be directly comparable with ours. This principle has been emphasized again and again by the most careful writers on mental tests,[1] and further insistence on it seems to verge upon pedantry. It ought by this time to be taken for granted in any work with mental tests in which the results of different workers are compared. A further point concerning this same aspect of standardization is the importance of using exactly the same test material. This is of particular importance with performance tests such as those described here. We have noted in Chapter I, in the enumeration of the tests used, the cases in which our test material differs from that commonly supplied by the dealers. Ordinarily we have chosen to work with the material which

[1] Cf. Woodworth, R. S., and Wells, L. F.: "Association Tests," *Psychological Monographs,* Vol. xiii, No. 5, Whole No. 57 (1911). This work deals largely with the question of arriving at the best methods for the presentation of the tests therein described.

can be readily obtained, but in some cases (as noted above) work had been begun with tests made by ourselves before they were obtainable through the usual channels, and later comparison of the materials showed no advantage over that made by us.

The other division of the problem of standardization deals with the establishment of norms. Here the question of supreme importance relates to the number of cases necessary before we can be certain of reliable norms. No dogmatic answer is possible to this question, and indeed few writers have discussed it. In some quarters, however, decided faith is placed in large numbers. Without any real reason large numbers and large numbers alone are deemed necessary for the group used in standardizing. The general argument runs somewhat as follows: If I wish to find out what a normal eight-year-old performance on a test is, I will get a fair norm if I test 300 eight-year-olds, I will get a better norm if I test 600 cases, and a still more reliable norm if I test 1,000 or 2,000 cases. The argument is seldom stated so bluntly, but the evident delight of some workers in mere numbers really amounts to the same thing.[2] As a matter of fact, the accumulation of an additional thousand cases to the first thousand, or an additional hundred to the first hundred, may be simply a waste of time. The question resolves itself into a consideration of the group of individuals tested, the variation of the

[2] Young, H. H.: "The Witmer Form Board," *Psychological Clinic,* Vol. x, No. 4 (1916), pp. 93-111.

norm with the addition of each group of results and the type of standardization required.

Differences in Social Status. The group of individuals tested will give results typical of that group and of no other group. Whether it is permissible to draw inferences from one group of individuals tested in regard to a different group of individuals who have not been tested is very debatable. If children in the best schools are selected, we shall obtain norms for children of good social status only, and we do not seem warranted in drawing conclusions as to what children of medium or poor social standing will be able to do from the norms obtained in such a way. That distinct differences in the performance of mental tests exist among children of different social status has been pointed out by a few workers.[3] Our norms may be perfectly reliable, but their reliability will extend only to the specific group tested. If we are seeking norms for the general population at large, a fair sampling of the general population at large would be the ideal method. Theoretically we ought to include individuals of all classes and of all degrees

---

[3] Yerkes, R. M., Bridges, J. W., and Hardwick, R. S.: A Point Scale for Measuring Mental Ability, Warwick and York (1915).

Bridges, J. W., and Coler, L.: "The Relation of Intelligence to Social Status," *Psychological Review,* Vol. xxiv, No. 1 (1917), pp. 1-31.

Strong, A. C.: "Three Hundred and Fifty White and Colored Children Measured by the Binet-Simon Measuring Scale of Intelligence: A Comparative Study," *Pedagogical Seminary,* Volume xx (1913), pp. 485-515.

of intellect. Our curve of distribution of the performances will only be a true curve if we include a sampling of all grades of intelligence from the very lowest to the very highest, but the sampling must include all the different grades in the same proportion as they exist in the community at large, i.e., assuming these proportions to be known. Such random selection of cases seems to be the only method of securing a true normal curve, and to call this method one of standardizing "on the basis of normal and abnormal material" [4] is not only to misstate the case, but to ignore the sound principle upon which it is based. If the lowest grades of intelligence are called abnormal and are to be omitted, we must be consistent and call the highest grades abnormal and omit them likewise, or else our norm will be shifted slightly too high. In establishing a norm for height we would not permit the rejection of those individuals who happen to be taller than some preconceived notion of height by which we had decided that all people above a certain height should be called abnormal or pathological cases.

This sampling made up of the right proportion of cases of all kinds is frankly ideal and up to the present has not been attainable in the standardization of mental tests. Various methods have of necessity been employed to arrive at reliable norms

[4] Wallin, J. E. W.: "Mentality Tests: A Symposium," *Journal of Educational Psychology,* Vol. vii, No. 6 (1916), p. 356.

for the population at large. Simpson[5] took two groups of individuals, a good and a poor group, and argued that the median performance of these groups would give a fair estimate of the ability of the normal or average in the population at large. The careful selection of individuals made by Simpson would seem to lend color to this claim. Similarly, Young[6] took two schools, one in a better class and one in a poor environment, upon which to standardize the Witmer Form Board. The method employed in these two instances ought, theoretically, to result in fairly reliable norms.

In the present work another method has been adopted, namely, the use of schools attended by children of the middle classes. One school might be said to represent the lower middle class or working population, and the other the upper middle class, made up of smaller tradesmen and some of the professional classes. The combination of these two groups of children, it was felt, would be very representative of the middle class of the population at large and would include a fair sampling of all grades of intelligence. It was felt that the medians at any rate would be very reliable, although the upper and lower end of the distribution might be somewhat lacking. It is doubtful, however, whether the distribution at the upper or lower

---

[5] Simpson, B. R.: "Correlations of Mental Abilities," Teachers College, Columbia University, *Contributions to Education,* No. 53 (1912), p. 122.

[6] Young, H. H.: *Op. cit.*

end would have been markedly affected by the inclusion of a school in the best neighborhood and a school in the worst neighborhood of the city. Until the collection of data can be extended over whole cities, some such method similar to these mentioned will have to be used.

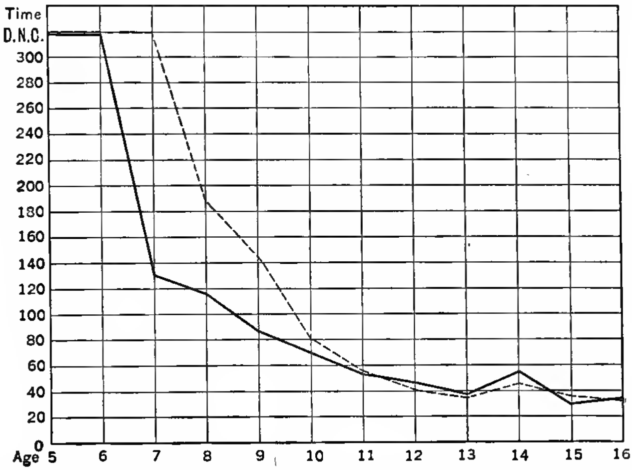
The Stability of the Norm. The question as to the number of cases adequate for a reliable norm for any age group can be determined only by a study of the fluctuation of the norm from time to time. Having decided upon the type of individual to be tested, an indication of the adequacy of the number is obtained when the addition of more cases fails to alter the norm materially. The ideal method would be to work out the values we require for each group tested at stated intervals throughout the work, watching what change occurs with the addition of each new set of results. Such a method was suggested and adopted by one of us[7] in a previous study. It seems at the present time to be the only way of answering this question. In the tentative standardization of the tests for the performance scale here described this method was only partially employed with a few tests because the number of tests used was so great as to preclude the collection of a great many cases at each age. It was felt that at no age have we arrived at the stage of having more than enough cases to establish a reliable norm, although at many ages we feel that there

---

[7] Pintner, R.: "The Standardization of Knox's Cube Test," *Psychological Review,* Vol. xxii, No. 5 (1915), p. 382 *et seq.*

are sufficient numbers to give fairly reliable norms.

With four of the tests, however, a partial employment of the method advised by Pintner was used. The use of the method is partial, since the norms were only computed twice and not at stated intervals. These results are best shown by means

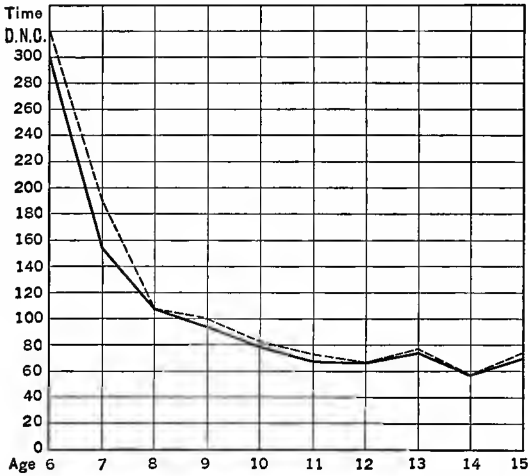


GRAPH 1.—Healy Puzzle "A." Time. Broken line=1915, 341 cases. Solid line=1916, 1,000 cases.

of graphs. In the graphs the broken line represents the first group of cases and the solid line the final standardization arrived at by the addition of a great many more cases.

Graph 1 shows the results for Healy Puzzle "A" (Time). The broken line shows the first 341 cases, and the solid line the total of 1,000 cases, which of course includes the first 341 cases. The greatest shift of the median takes place at ages 7, 8 and 9, otherwise the curves remain relatively the same. The shift at the ages mentioned leads to a better median performance at those ages.

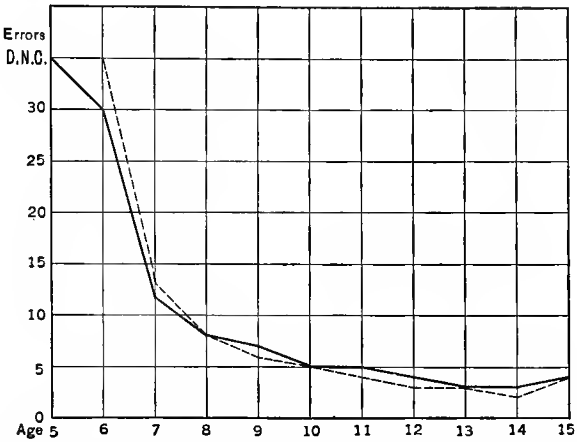
Graph 2 shows the results for the Casuist Form Board (Time). The two curves are practically the same. The addition of 477 cases to the first group



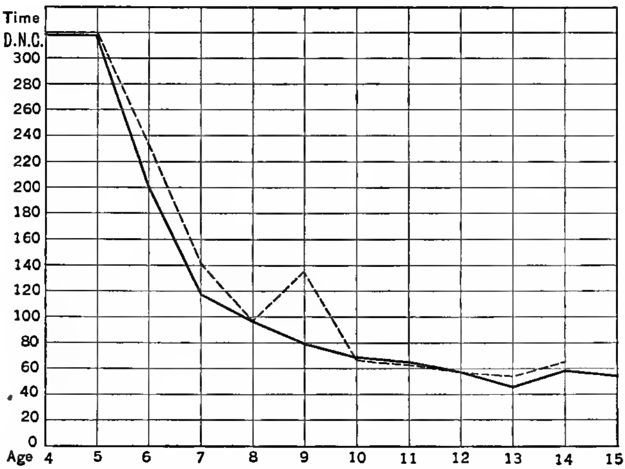
GRAPH 2.—The Casuist Form Board. Time. Broken line=1915, 428 cases. Solid line=1916, 905 cases.

of 428 has not altered the shape of the original curve. This means that our additional 477 cases were practically useless, as far as the medians are concerned. What influence they may have had upon a percentile distribution we cannot tell. However, the results on this test show that mere increase in numbers in and for itself is no criterion of added reliability of the norms.

Graph 3 shows the results for the Casuist Form



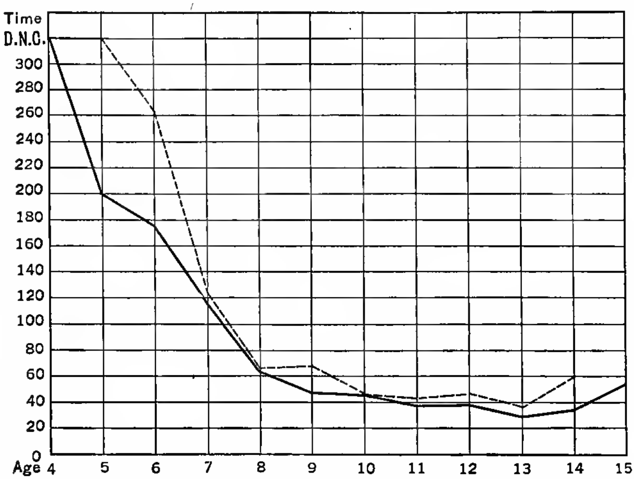
GRAPH 3.—The Casuist Form Board. Errors. Broken line=1915, 428 cases. Solid line=1916, 905 cases.



GRAPH 4.—The Five Figure Form Board. Time. Broken line=1915, 295 cases. Solid line=1916, 963 cases.

Board (Errors). The curves show practically the same facts as those in the preceding graph.

Graph 4 gives the results for the Five Figure Board (Time). With the exception of age nine, the medians show little change with the addition of 668 cases to the first 295 cases. This graph,



GRAPH 5.—The Two Figure Board. Time. Broken line=1915, 309 cases. Solid line=1916, 978 cases.

however, shows the advantage of additional cases where we have obvious abnormalities in the curve, as at age nine. The first group of nine-year-olds was not representative of nine-year-olds in general. The addition of more cases has smoothed the curve.

Graph 5 shows the results for the Two Figure Board (Time). The addition of 669 cases to the first 309 cases has smoothed the original curve somewhat and has lowered the medians perceptibly at ages five and six.

We offer these results as indicative of the method which must be followed in the determination of an adequate number of cases for purposes of standardization.

Various Types of Standardization. The number of cases necessary is also conditioned by the type of standardization we are attempting to make. We may at the present time distinguish between three types of standardization.

The simplest type aims at the establishment of median or average performances. If we are satisfied with this, a relatively small number of cases in each age group may be sufficient. Our interest does not center so much upon the whole of the curve of distribution as upon the middle part of it. Naturally enough, the median will be influenced by the distribution as a whole, but obvious discrepancies at the upper or lower ends need not be disastrous to the median. This type of standardization is familiar to us in much of the work dealing with the standardization of single tests.

A somewhat more complex type of standardization is presented in the attempt to place a test at a specific age in an age scale. The question at issue here is as to the percentage of cases that must pass a test in order to make the test a valid test for the age in question. Seventy-five per cent has been generally adopted as the standard, although various other suggestions have been made.

This type of standardization became important in dealing with the problem of placing tests in the Binet Scale. Binet himself nowhere states specifically the percentage of passes necessary to place a test. From his actual work we would infer that he did not keep to a rigid standard, but fluctuated between 60 and 90. In the same way Terman and Childs seem to vary in their standard from 60 per cent upwards. Pintner,[8] in his standardization of the Cube Test for age scale purposes, laid most stress upon a sharp rise in the curve anywhere above 60 per cent, emphasizing the point that above this percentage the greatest differentiation between any two ages would indicate the most suitable age for a test.

Bobertag, Goddard and Kuhlmann adhere much more closely to the 75 per cent basis. The justification for this method seems to be based on the normal curve of distribution. We may assume that at each age 50 per cent of the individuals are normal and 25 per cent above and 25 per cent below normal. If a test is suited to the normal ability for children at a specific age, then it will be passed by the 50 per cent normal individuals and also by the 25 per cent above normal, i.e., by 75 per cent of the children altogether. Only the lowest 25 per cent will fail. In the arrangement of these performance tests into a year scale (see Chapter V) it will be noted that we have kept to the rigid 75 per cent standard. This is in part due to the nature

[8] Pintner, R.: *Op. cit.*

of the data with which we are dealing. The tests for the most part are not scored as mere passes or failures. The quality of the performance is based, for the most part, upon the time and upon the number of moves or errors made. It seemed most appropriate, therefore, to mark off the point above which 75 per cent of the cases lay and to consider any score or time value better than this as normal for the age in question. A more extended description of the application of this method to our tests is given in Chapter V. It is the application of the 75 per cent method to tests involving time or a wide range of scoring.

The third type of standardization is the percentile method. Here the whole range of distribution is divided up into as many percentile groups as is deemed feasible. These percentiles, if sufficiently numerous, give a fairly reliable picture of the distribution of the cases. In general practice the division into percentiles has not gone beyond 10 percentiles. This is, indeed, as fine a differentiation as we require at the present stage of standardization of tests.

This percentile method of standardization has been made use of in some recent studies of mental tests by Woolley.[9] In our standardization the 10

[9] Woolley, H. T., and Fischer, C. R.: "Mental and Physical Measurements of Working Children," *Psychological Monographs,* Vol. xviii, No. 1 (1914); and also, Woolley, H. T.: "A New Scale of Mental and Physical Measurements for Adolescents and Some of Its Uses," *Journal of Educational Psychology,* Vol. vi, No. 9 (1915), pp. 521-550.

percentiles have been used. The advantage of this type of standardization is due to the fact that it allows a comparison of a particular child's performance with the performance of other children of the same age. It is much more desirable to be able to compare a specific child with other children of the same age than with children of differing ages. We can then express the child's ability as being equal to that of a 10 or 20 or 60 or 80 percentile child of his own age. It is obvious at once, however, that this type of standardization will require a greater number of cases in order to give reliable norms than either of the other two types. Indeed, we must be tolerably certain that we have included a fair sampling of all grades of intelligence at the age in question.

We may arrive at fairly reliable medians by a judicious selection of children, we shall require more cases to fix the 75 per cent point, and we shall need the greatest number of cases to fix with any degree of accuracy the 10 percentile points from zero to 100.

Standardized Tests. The actual work accomplished in the standardization of mental tests for the estimation of intelligence may be divided into: (1) the standardization of scales, and (2) the standardization of individual tests.

Scales. The work on the standardization of scales may be said to have begun with Binet himself. Binet's 1908 scale may be called the first standardized scale, and the 1911 revision may be looked upon as another standardization of the same scale. This was followed in America by the restandardization of the scale for American children by Goddard.[10] From this time on we have the standardizations of Bobertag, Terman and Childs, and Winch, culminating in the greatly modified scales of Terman (the Stanford Revision) and in the Point Scale by Yerkes and Bridges. With the exception of the latter, all these standardizations discuss the question of the appropriate placing of tests at specific ages. And the chief point in this discussion is, as we have mentioned above, the per cent of passes necessary to place a test at a specific age. It will be needless for us to enter into a detailed discussion of these standardizations of the Binet tests. The history of this aspect of the subject is marked by an increasing accuracy in standardization and a growing discussion of the theoretical assumptions underlying the whole basis of standardization. Stern,[11] Kuhlmann[12] and Ter-

---

[10] Goddard, H. H.: "Two Thousand Normal Children Measured by the Binet-Simon Measuring Scale of Intelligence," *Pedagogical Seminary*, Vol. xviii (1911), pp. 232-259.

[11] Stern, W.: "The Psychological Methods of Testing Intelligence," Trs. by Whipple, *Educational Psychology Monographs*, No. 13, Warwick and York (1914).

[12] Kuhlmann, F.: "Some Results of Examining a Thousand Public School Children with a Revision of the Binet-Simon Tests of Intelligence by Untrained Examiners," *Journal of Psycho-Asthenics*, Vol. xviii, No. 3 (March, 1914), pp. 150-179; and No. 4 (June, 1914), pp. 233-269; "The Present Status of the Binet and Simon Tests of the Intelligence of Children," *Journal of Psycho-Asthenics*, Vol. xvi, No. 3 (March, 1912), pp. 113-139.

man [13] have added much of value to this phase of the question.

A different kind of standardization has been attempted by Trèves and Saffiotti.[14] The tests of Binet and Simon are arranged in order of difficulty for each age and then grouped into three classes according as they are adapted to dull (*faibles*), average (*moyens*) or bright (*forts*) children in each age group. This classification of dull, average and bright is determined by the percentage of children passing the different tests. All tests passed by 60 per cent or more of the children are called tests for the dull group; tests passed by from 40 to 60 per cent are called tests for the normal group, and tests passed by less than 40 per cent are for the bright group. There seems to be no principle underlying this division into groups. Just why these particular percentages are chosen we are not told. That a middle 30 per cent of the children should be chosen as representing the average seems strange. Forty per cent are relegated to the dull group and 30 per cent to the bright group.

---

[13] Terman, L. M.: "The Binet-Simon Scale for Measuring Intelligence; Impressions Gained by Its Application on Four Hundred Non-selected Children," *Psychological Clinic,* Vol. v (1911), pp. 199-206, 239-244.

Terman and Childs: "A Tentative Revision and Extension of the Binet-Simon Measuring Scale of Intelligence," *Journal of Educational Psychology,* Vol. iii (1912), pp. 61-74, 133-143, 198-208, 277-289.

[14] Saffiotti, U.: "L'échelle métrique de l'intelligence de Binet-Simon modifiée selon la méthode Trèves-Saffiotti," *L'année psychologique,* Vol. xviii (1912), pp. 327-340.

Under each of these three groups 3 smaller subgroups are given. It seems strange that the authors did not assume the normal curve of distribution and use it as the basis of their classification.[15] As it is, the classification is loose and arbitrary in the extreme. Historically it would seem to foreshadow the percentile method. The grouping of children into percentiles and the use of the percentile as a description of the child's mental status are a distinct advance in clearness of thinking over the arbitrary grouping proposed by Trèves and Saffiotti.

Apart from the Binet Scale, there have been very few other scales established or standardized. De Sanctis' Scale,[16] appeared about 1906 with scarcely any attempt at standardization. It was the aim of de Sanctis primarily to try to group different grades of known mental deficiency, or at most to pick out the feeble-minded among normal children. In his own words: "Je puis conclure en général que la série, avec les modifications que j'y ai portées jusqu' à ce jour, donne certainement d'excellents résultats pour les enfants et adolescents *faibles* de sept à seize ans."[17] No statistical presentation of the material such as we are now familiar with in works on standardization was

---

[15] Pintner, R., and Paterson, D. G.: "A Psychological Basis for the Diagnosis of Feeble-mindedness," *Journal of Criminal Law and Criminology*, Vol. vii, No. 1 (1916), pp. 32-55.

[16] de Sanctis, S.: "Types et degrés d'insuffisance mentale," *L'année psychologique*, Vol. xii (1906), pp. 70-83.

[17] *Idem*, p. 80.

given at that time. Recently a standardization and modification of this scale for American children has been made by Martin.[18]

The scale of performance tests arranged by Knox [19] seems to be the only other attempt at the standardization of a whole scale. The standardization of Knox's Scale is obviously inadequate, as the author realizes. The scale was constructed to fill immediate and urgent practical needs in the work of detecting mentally defective immigrants. The tests making up the rough year scale devised by Knox are largely of the performance type. Some have been devised by Knox himself, together with borrowings and adaptations of tests of Binet, Healy and others. Interesting norms of performance have been obtained by Knox for children of different nationalities at different ages.

Lastly, a scale of tests for adolescents has been proposed by Woolley.[20] It is the outcome of measurements on from 600 to 800 adolescents of ages fourteen and fifteen. It is of interest as being a very distinct departure in every way from the

---

[18] Martin, L.: "A Contribution to the Standardization of the de Sanctis Tests," *Training School Bulletin*, Vol. xiii, No. 4 (1916), pp. 93-110.

[19] Knox, H. A.: "A Scale, Based on the Work at Ellis Island, for Estimating Mental Defect," *Journal of the American Medical Association*, Vol. lxii (March 7, 1914), pp. 741-747.

[20] Woolley, H. T.: "A New Scale of Mental and Physical Measurements for Adolescents, and Some of Its Uses," *Journal of Educational Psychology*, Vol. vi, No. 9 (1915), pp. 521-550.

Binet type of scale. None of the tests belongs to the original Binet group of tests and the method of evaluating the results is the percentile method. We shall refer in more detail to the percentile method in Chapter VIII. It is sufficient to remark here that this is the first attempt known to the writers to evaluate performance in reference to percentile points for each age. Constant use of the percentile method would very soon lead us to attach very definite meanings to such terms as 10 percentile ability or 70 percentile ability and so forth.

In regard to the tests used in Woolley's Scale, it is to be noted that none of them is taken from the Binet Scale. They cover a wide range, including physical tests, tests of motor ability, as well as purely mental tests. The radical difference between Woolley's Scale and the one presented here is the inclusion in the former of many tests involving language. The drawback of Woolley's Scale at the present time is its limited scope, since it has only been standardized for ages fourteen and fifteen. It must be recognized, however, that the standardization for these ages is very thorough. It is much more complete than the standardization of any group of tests made up to the present time.

Individual Tests. In addition to the standardization of the scales referred to above, we have also the standardization of individual mental tests. These tests vary all the way from very inadequate and incomplete standardizations to very accurate and thorough ones. These standardizations are interesting from the fact that they have broadened considerably the discussion of our problem and indicated types of standardization in addition to the age scale type.

Healy's[21] first description of his performance tests was not accompanied by anything in the way of an adequate standardization. His emphasis in this work was laid upon the tentative nature of his results, to quote: "but it is to be distinctly understood that we ourselves still regard our tests and methods as strictly tentative."[22] There is no direct reference to the question of standardization, and no attempt made at it. Sample performances of a few cases on each test are given, but these are of course useless to guide any other worker in regard to what sort of a performance he may expect at any stage of intelligence. Doubtless, constant use of a test will give the worker some idea of what a child can be expected to do, but this is always uncertain and of no help to other workers.

That a standardization of these tests was felt to be desirable is evidenced by the appearance of the work of Schmitt,[23] "done by the author while psychologist at the Chicago Psychopathic Insti-

---

[21] Healy, W., and Fernald, G. M.: "Tests for Practical Mental Classification," *Psychological Monographs,* Vol. xiii, No. 2, Whole No. 54 (March, 1911).

[22] *Idem,* p. 3.

[23] Schmitt, C.: "Standardization of Tests for Defective Children," *Psychological Monographs,* Vol. xix, No. 3, Whole No. 83 (1915).

tute." This work is unfortunate, however, in two respects: first, in the type of children selected, and secondly, in the small number of cases at each age. The children tested were from "the kindergarten and first six grades of a private school in Chicago. . . . They were the children of people of the professional class mainly."[24] A further description of these children and the school they attended makes it clear that they are distinctly above the average child. We are safe in concluding that the norms established on such a group will be too high. Secondly, the number of children at each age is seldom much above twenty, and at some ages is considerably below. It is doubtful whether a valid median or average performance can be obtained from such a small number. If the group were very homogeneous, as the author claims, such might be the case; but the homogeneity of the subjects is not so apparent wherever we can guess at the distribution from the presentation of the results as given.

In Healy's[25] later description of these tests he devotes a paragraph to norms, making use of Schmitt's work, to quote: "Some of her results are embodied in our statement of norms."[26] In some cases the norms of Healy seem to have been taken directly from Schmitt's work, and in other cases to

---

[24] Healy, W.: *Op. cit.,* p. 2.

[25] Healy, W.: The Individual Delinquent, Little, Brown and Co. (1915).

[26] *Idem,* p. 106.

have been supplied from his own experience with the tests. They are for the most part decidedly vague and confusing, and can scarcely lay claim to adequate standardization.

Another standardization of a group of tests, including four of the Healy tests, has appeared in the work of Hall.[27] It does not lay any claim to completeness in any respect, and the conditions imposed upon the investigators did not allow them to test more than 180 (30 each at ages seven, eight, nine, ten, eleven and twelve) unselected public school children. The rest of the children tested were inmates of feeble-minded institutions or of orphan asylums. The data accumulated from this group include over a thousand cases and are interesting, although of doubtful value for general standardization purposes. Furthermore, the value of establishing norms on the basis of the mental age as determined by the Binet or any other scale is a questionable procedure. It is based upon the presumption that the Binet Scale is the final and only court of appeal for establishing the mental age of a child. This is rather a bold assumption to make at this time. Tests so standardized could never be used as correctives and complements of the Binet Scale, and it is the feeling of the writers that such correctives and complements of the Binet Scale are at present required.

[27] Hall, G.: "Eleven Mental Tests Standardized," *Eugenics and Social Welfare Bulletin,* No. V, State Board of Charities, Albany, New York (1915).

The results of the 180 public school children seem too few to make a satisfactory standardization, yet they have formed a beginning in the standardization of these tests. Unfortunately no tables of distribution are presented so that they cannot be incorporated into the work of others who are accumulating data on these tests. It would seem to the writers very desirable that each worker should publish the results in such form that they might be added to the data of other workers.

The standardization of the Form Board by Sylvester [28] marks a distinct advance to a more minute analysis and a more accurate standardization of a particular test. In all 1,537 children were tested, at ages ranging from five to fourteen. There were from 80 to 221 children at each age, and we have the data presented in such a form as to be readily accessible to other workers, so much so that we have incorporated this test as standardized by Sylvester into the present scale. The children were an almost unselected group of ordinary school children, so that we may take it for granted that the norms will be fairly representative for children of the ages tested. The table of distribution of the cases has given us the opportunity of making use of the data in the various ways in which our scale of performance tests is presented.

No advance upon this work seems to have been

[28] Sylvester, R. H.: "The Form Board Test," *Psychological Monographs,* Vol. xv, No. 4, Whole No. 65 (1913).

made by Young,[29] who presents another standardization of the form board. The form board used in this standardization is unfortunately of a different kind from the one used by Sylvester or the one used by Goddard. It is so different as to make invalid a comparison of the two standardizations or a combination of the two sets of data. A comparison and a combination of the data obtained by Young and Sylvester would have been interesting and valuable, if such could have been made. A comparison would have shown us how much, if at all, the distribution of the results differed, and if they differed, the combination of all the results would have led to a still more accurate standardization. The amount of difference in the resulting norms from those previously obtained would have served as an indication of the reliability of the ultimate standardization.

Pintner,[30] in his work with the Cube Test, has presented a standardization based on 867 cases. He has suggested there the method of watching the fluctuation of the results with the addition of more data, as a criterion for a sufficient number of cases for a satisfactory standardization. His adaptation of the original test has broadened the scope of the test.

Pintner and Anderson,[31] in their standardization

[29] Young, H. A.: "The Witmer Form Board," *Psychological Clinic,* Vol. x, No. 4 (1916), pp. 93-111.

[30] Pintner, R.: "The Standardization of Knox's Cube Test," *Psychological Review,* Vol. xxii, No. 5 (1915), pp. 377-401.

[31] Pintner, R., and Anderson, M. M.: "The Picture Completion Test." *Educational Psychology Monographs.* Warwick and York, Baltimore (1917).

of the Picture Completion Test, have given a minute analysis of the test and an exact standardization of procedure and interpretation of results. The method of scoring adopted aims to set an objective, in place of a subjective, evaluation of the result of the performance. These two last standardizations have been made use of in the present scale.

A standardization of the Healy Construction Puzzle "A" has been made by Bruckner and King.[32] The study is very incomplete, since it deals only with eight- and ten-year-old children. Ninety eight-year-olds and 59 ten-year-olds were tested. As far as these norms go, they appear to be very good, and reference has been made to them in the previous chapter in the description of this test. This study falls in line with the others that take up a minute analysis of a particular test.

A form board called the "Arrow Board" has recently been described and partly standardized by Dunham.[33] He has reported results for 184 high school pupils aged fifteen, sixteen, seventeen and eighteen. The number of subjects tested and the nature of the selection of subjects make the test of little value in practical clinical work at the present time.

---

[32] Bruckner, L., and King, I.: "A Study of the Fernald Form Board," *Psychological Clinic,* Vol. ix, No. 9 (1916), pp. 249-258.

[33] Dunham, F. L.: "The Arrow-Board, an Adult 'Form Board Test,'" *Pedagogical Seminary,* Vol. xxiii, No. 2 (June, 1916).

In summing up this brief résumé of the most important work on standardization, we may note that the general trend at the present time seems to be toward a more minute analysis of each test and toward the accumulation of a greater number of cases at each age. We have left out of consideration a great many mental tests that are not at present used for the diagnosis of intelligence.[34] Our aim has been to deal with those tests bearing more directly upon our special problem. Our criticism of much of the standardization has been from the point of view of its inadequacy, both in regard to the presentation of the data and in regard to the number of cases examined. With regard to the number of cases examined we are well aware of the limitations of our own data, but in regard to the presentation we hope that we have given it in such a form that it may be useful for other workers, so that it may be added to in the future and manipulated in any way that future methods of standardizations may require.

---

[34] For a complete account of such tests see Whipple, G. M.: A Manual of Mental and Physical Tests, Two Volumes, Warwick and York (1914-15).

CHAPTER IV

THE PRESENTATION OF THE DATA

Tables of Distribution. The data on the score sheets have all been arranged in tables of distribution (see Tables 1 to 21). It has been our aim to make the distribution as extensive as possible. Wherever feasible small steps were used. The smallest is that of one second, as in Table 3, showing the results for the Seguin Form Board.[1] In the other cases of time distribution such a small step would have been impossible owing to the large number of steps that would have been required. In these cases a compromise was resorted to and relatively short steps were taken for the shorter times where the majority of cases was likely to fall, and relatively longer steps at the upper end for the longer time periods. In general, steps of 10 seconds were made from 0 to 100, and from there up to 300 (the time limit) steps of 25 seconds were used. This is the arrangement in Tables 12, 10, 6, 4 and 8. In Tables 1 and 14 the shorter steps of 10 seconds were extended beyond the 100 second

[1] This table has been copied from Sylvester, R. H.: "The Form Board Test," *Psychological Monographs*, Vol. xv, No. 4, Whole No. 65 (1913).

limit. In Table 17 a slight modification was introduced in order to allow of a grouping of cases according to the number of minutes taken to pass the test.

In regard to the tables showing the number of moves or errors, or the score made, the same policy was adhered to. Tables 16, 18, 20 and 21 give each possible step of the score to be obtained on those tests. Table 19 is given in steps of 10 up to 250, and from there on in steps of 50 and 100 because of the large scores obtainable by the method of scoring adopted. Table 2 goes up by steps of 1 to 14, since no case made more than 14 errors in a completed performance. Tables 13 and 11 are arranged according to steps of 1 up to 20 and by steps of 5 beyond. Table 7 is arranged in steps of 1 up to 25, with steps of 5 beyond. Tables 5 and 9 are arranged in steps of 1 up to 15 and steps of 2 beyond. Table 15 goes by steps of 5 up to 100.

The distribution in all the tables is given for ages four to sixteen inclusive. In most tests there are very few cases at ages four and sixteen, and we do not pretend that reliable norms have been obtained at these ages. Ages fourteen and fifteen have relatively few cases and our standardization for these ages is very uncertain. We have, however, decided to include all the data that we were able to obtain so that it might be added to in the future.

The last line in the tables of distribution just above the line giving the total number of cases at each age is marked D.N.C. (Did Not Complete). This shows the number of cases which did not complete the test within the time limit. All cases which did not complete within the time limit were so marked both for time and for errors or moves, since in such cases the number of moves or errors at the end of the time limit was not comparable with the number of moves or errors made by an individual completing the test. At the bottom of each of the tables of distribution are given the median, the 75 percentile and 25 percentile, and finally the quartile, which serves as a measure of the range of distribution.

The graphs for each of the tests show the median (solid line) and the 75 and 25 percentiles (dotted lines). The space between the two percentile curves represents the amount of variation among the middle 50 per cent.

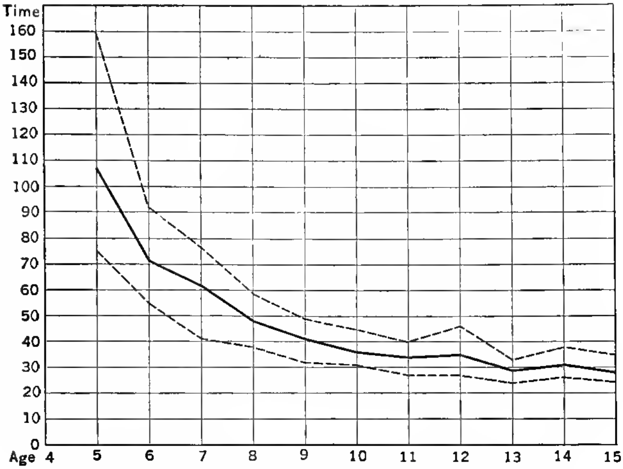
THE MARE AND FOAL TEST

Time. (Table 1 and Graph 6.) The distribution shows relatively little scattering. It is obviously a test where ability to deal with the situation increases fairly rapidly from age five to age ten at least. Only 5 children fail to complete the test and these are all aged eight or below. No child completes it in less than 10 seconds. The curve for the medians shows a steady and uniform decrease to age eleven, from which age onwards no

TABLE 1. THE MARE AND FOAL TEST. TIME.

Age	4	5	6	7	8	9	10	11	12	13	14	15	16	
Time														
10	..	..	..	..	..	..	..	..	2	3	1	..	..	..
20	..	..	..	2	10	14	18	27	21	22	11	9	..	..
30	..	..	3	18	23	19	31	34	23	12	11	6	2	..
40	..	..	11	16	22	23	18	13	11	5	3	..	..	..
50	..	2	6	4	20	7	6	5	7	..	1	1	..	..
60	..	4	13	18	7	4	..	1	3	..	..	..	..	..
70	..	4	10	7	6	4	2	1	3	..	..	..	..	..
80	..	1	7	4	3	..	1	..	..	..	..	..	..	..
90	1	2	5	6	1	..	..	..	..	..	..	..	..	..
100	..	3	2	2	1	..	..	..	..	..	..	..	..	..
110	..	3	2	1	2	..	..	..	..	..	..	..	..	..
120	..	..	1	1	1	1	1	..	..	..	..	..	..	..
130	..	2	3	1	..	..	..	..	..	..	..	..	..	..
140	..	..	..	..	..	..	..	..	..	..	..	..	..	..
150	..	1	2	..	..	..	..	..	..	..	..	..	..	..
160	..	1	..	1	..	..	..	..	..	..	..	..	..	..
170	..	..	1	..	..	..	..	..	..	..	..	..	..	..
180	..	1	..	..	..	..	..	..	..	..	..	..	..	..
190	..	2	..	..	..	..	..	..	..	..	..	..	..	..
200	..	..	..	3	..	..	..	..	..	..	..	..	..	..
225	..	..	..	..	..	..	..	..	..	..	..	..	..	..
250	..	1	..	..	..	..	..	..	..	..	..	..	..	..
275	..	..	..	..	..	..	..	..	..	..	..	..	..	..
300 +	..	..	..	..	1	..	..	..	..	..	..	..	..	..
D. N. C.	..	2	1	1	1	..	..	..	..	..	..	..	..	..
Total	1	29	67	85	98	72	77	81	70	42	27	16	2	667
75% ile	..	75	55	41	38	32	31	27	27	24	26	24	..	..
Median	..	107	71	62	48	41	36	34	35	29	31	28	..	..
25% ile	..	160	92	77	59	49	45	40	46	33	38	35	..	..
Quartile	..	42.5	18.5	18	10.5	8.5	7.0	6.5	9.5	4.5	6.0	5.5	..	..

marked increase in rapidity in solving the test is shown. The variation in performance among the younger children is greater than the variation among the older children. On the whole, the variation at any age (with the exception of age five) is not great.

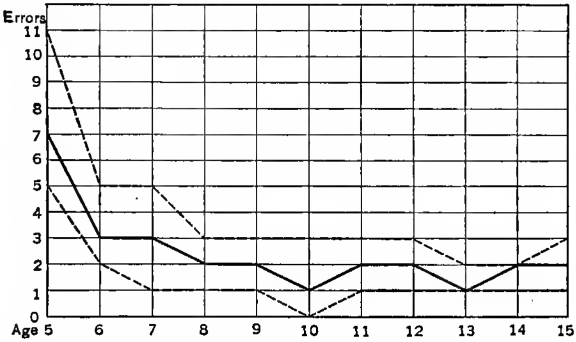


GRAPH 6.—The Mare and Foal Test. Time.

TABLE 2. THE MARE AND FOAL TEST. ERRORS.

Age	4	5	6	7	8	9	10	11	12	13	14	15	16	
Errors														
0	..	..	5	7	13	12	22	15	17	11	5	2	..	..
1	..	..	7	17	28	22	19	23	15	11	8	3	1	..
2	..	1	12	16	19	10	15	19	17	11	9	6	..	..
3	..	..	12	14	16	12	16	15	14	4	1	2	..	..
4	..	4	10	7	10	4	7	7	2	4	1	2	..	..
5	..	4	7	10	5	6	2	..	2	1	2	1	1	..
6	..	3	5	4	2	2	..	1	2	..	1	..	..	..
7	..	6	2	3	..	2	..	1	..	..	..	..	..	..
8	..	1	3	1	3	2	1	..	1	..	..	..	..	..
9	1	1	1	..	1	..	..	..	..	..	..	..	..	..
10	..	2	1	2	..	..	..	..	..	..	..	..	..	..
11	..	1	1	..	..	..	..	..	..	..	..	..	..	..
12	..	2	..	1	..	..	1	..	..	..	..	..	..	..
13	..	1	..	1	..	..	..	..	..	..	..	..	..	..
14	..	1	..	1	..	..	..	..	..	..	..	..	..	..
D. N. C..........	..	2	1	1	1	..	..	..	..	..	..	..	..	..
Total............	1	29	67	85	98	72	77	81	70	42	27	16	2	667
75% ile..........	..	5	2	1	1	1	0	1	1	1	1	1	..	..
Median..........	..	7	3	3	2	2	1	2	2	1	2	2	..	..
25% ile..........	..	11	5	5	3	3	3	3	3	2	2	3	..	..
Quartile..........	..	3.0	1.5	2.0	1.0	1.0	1.5	1.0	1.0	0.5	0.5	1.0	..	..

Errors. (Table 2 and Graph 7.) The distribution table shows much the same appearance as the one for the time. There are a great many cases completing the test with no errors. The curve showing the median number of errors at each age is not very uniform. From age eight upwards it fluctuates continually between 1 and 2 errors.



GRAPH 7.—The Mare and Foal Test. Errors.

The drop at age ten shows a remarkably good performance for the ten-year-olds. The percentiles show a fairly narrow range. At no age is the quartile greater than 2 errors (with the exception of age five).

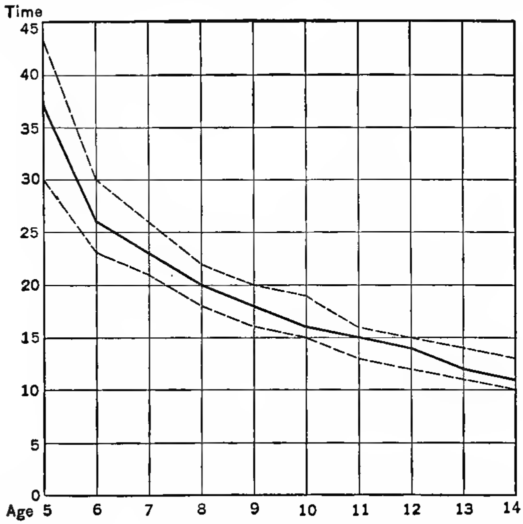
THE SEGUIN FORM BOARD

Time. (Table 3 and Graph 8.) This distribution taken from Sylvester is exceptionally good. Some of it is due to the fact, as we have noted elsewhere, that exceptionally dull or nervous children were excluded. This accounts for the lack of scat-

TABLE 3. THE SEGUIN FORM BOARD. TIME.

Age	5	6	7	8	9	10	11	12	13	14	
Time											
9	..	..	..	..	..	..	2	..	4	8	..
10	..	..	..	..	..	..	2	6	3	16	..
11	..	..	..	..	..	..	4	17	15	16	..
12	..	..	..	..	..	13	17	21	23	16	..
13	..	..	..	..	4	17	19	20	11	9	..
14	..	..	..	2	9	24	29	31	11	6	..
15	..	..	1	6	30	38	39	22	6	4	..
16	..	..	5	16	26	28	26	8	5	2	..
17	..	..	5	20	29	28	17	9	2	1	..
18	..	6	10	13	31	16	7	2	..	..	..
19	..	6	8	19	19	22	4	1	..	..	..
20	..	11	13	39	19	11	2	2	..	..	..
21	..	12	24	28	8	9	1	1	..	..	..
22	1	7	18	16	10	2	2	..	..	..	..
23	2	6	13	12	6	4	..	1	..	..	..
24	2	21	17	4	3	6	1	..	..	..	..
25	1	14	7	6	4	..	..	..	..	..	..
26	5	11	16	6	4	2	..	..	..	..	..
27	1	9	13	2	2	1	..	..	..	..	..
28	3	11	5	5	2	..	..	..	..	..	..
29	4	5	6	2	3	..	..	..	..	..	..
30	2	10	4	3	1	.		..	..	..	..
31	..	11	2	1	3	..	..			..	..
32	3	9	..	2	..		..	..	..	..	..
33	3	3	3	..	..	..	..	..	..	..	..
34	5	6	1	..	1	.	..	..	..	..	..
35	6	3	..	..	..	..	..	..	..	..	..
36	1	1	1	..	..	..	..	..	..	..	..
37	2	3	..	..	..	..	..	..	..	..	..
38	6	1	1	..	..	..	..	..	..	..	..
39	3	1	..	..	..	..	..	..	..	..	..
40	2	1	..	..	..	..	..	..	..	..	..
41	2	2	..	..	..	..	..	..	..	..	..
42	5	..	..	..	..	..	..	..	..	..	..
43	4	..	..	..	..	..	..	..	..	..	..
44	2	..	..	..	..	..	..	..	..	..	..
45	2	..	..	..	..	..	..	..	..	..	..
46	1	..	..	..	..	..	..	..	..	..	..
47	1	..	..	..	..	..	..	..	..	..	..
48	1	..	..	..	..	..	..	..	..	..	..
49	1	..	..	..	..	..	..	..	..	..	..
50	2	..	..	..	..	..	..	..	..	..	..
51	..	..	..	..	..	..	..	..	..	..	..
52	..	..	..	..	..	..	..	..	..	..	..
53	1	..	..	..	..	..	..	..	..	..	..
54	..	..	..	..	..	..	..	..	..	..	..
55	1	..	..	..	..	..	..	..	..	..	..
56	3	..	..	..	..	..	..	..	..	..	..
57	..	..	..	..	..	..	..	..	..	..	..
58	1	..	..	..	..	..	..	..	..	..	..
75	1	..	..	..	..	..	..	..	..	..	..
Total....	80	170	173	206	214	221	172	141	80	80	1537
75% ile..	30	23	21	18	16	15	13	12	11	10	..
Median..	37	26	23	20	18	16	15	14	12	11	..
25% ile..	43	30	26	22	20	19	16	15	14	13	..
Quartile..	6.5	3.5	2.5	2.0	2.0	2.0	1.5	1.5	1.5	1.5	..

tering cases on the table—a phenomenon that is generally observed on all the other tables of distribution. This scattering, however, would not spoil the general upward trend of the table, which is reflected by the constant and steady decrease of the medians for each age. The percentiles keep



GRAPH 8.—The Seguin Form Board. Time.

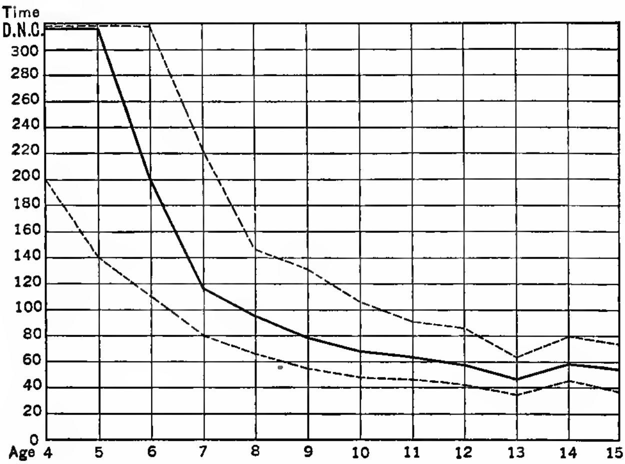
fairly close to the median all the way along, and the very narrow range of variation of the middle 50 per cent is indicative of the value of this test for the purpose of differentiating grades of intelligence. At no age is the quartile greater than 6.5 seconds and at most ages it varies from 1.5 to 2.0 seconds. The longest time record, made by a five-year-old child, is 75 seconds, while the shortest, made by 14 individuals, is 9 seconds.

The Five Figure Form Board

Time. (Table 4 and Graph 9.) The distribution here is not so compact as in the previous test. This may be due partly to our policy of absolute nonselection of cases, and partly to the fact that the puzzle board idea, which introduces a slight element of chance, may influence the performance of the test to a very slight extent. There are cases of inability to complete the test at all ages up to

Table 4. The Five Figure Form Board. Time.

Age	4	5	6	7	8	9	10	11	12	13	14	15	16	
Time														
20	..	..	..	..	..	..	1	3	6	6	5	3	..	..
30	..	..	1	1	2	7	9	12	15	17	2	2	..	..
40	..	1	1	3	5	7	18	23	18	7	6	3	..	..
50	..	..	2	8	14	16	14	16	17	10	6	2	1	..
60	..	..	5	8	20	10	13	18	11	6	6	4	2	..
70	..	1	4	8	10	7	8	13	7	2	3	1	..	..
80	..	1	7	13	12	8	7	6	7	3	1	2	..	..
90	..	1	6	6	12	6	7	8	5	2	1	..	..	.
100	..	7	6	14	10	7	12	8	8	1	3	2	..	..
125	1	3	12	7	19	6	9	2	1	..	..	..	..	..
150	1	1	6	7	3	2	4	2	2	1	..	..	..	..
175	..	5	7	6	7	1	2	4	1	..	1	..	..	..
200	1	2	7	3	2	1	1	4	3	..	1	..	..	..
225	..	1	5	2	5	2	..	1	..	..	2	..	..	..
250	..	..	1	1	1	3	..	1	2	..	..	..	..	..
275	..	2	2	1	5	..	..	..	..	..	..	..	..	..
300	1	1	1	..	..	..	..	..	..	..	..	..	..	..
D. N. C.	7	28	41	24	10	10	1	1	..	..	..	..	..	..
Total......	11	54	114	112	137	93	106	122	103	55	37	19	3	966
75% ile....	200	139	112	80	67	56	49	48	43	35	45	38	..	..
Median....	D. N. C.	D. N. C.	200	117	97	79	69	64	58	47	59	55	..	..
25% ile....	D. N. C.	D. N. C.	D. N. C.	225	146	132	107	91	85	63	80	75	..	..
Quartile....	..	..	..	72.5	39.5	38.0	29.0	21.5	21.0	14.0	17.5	18.5	..	..



GRAPH 9.—The Five Figure Form Board. Time.

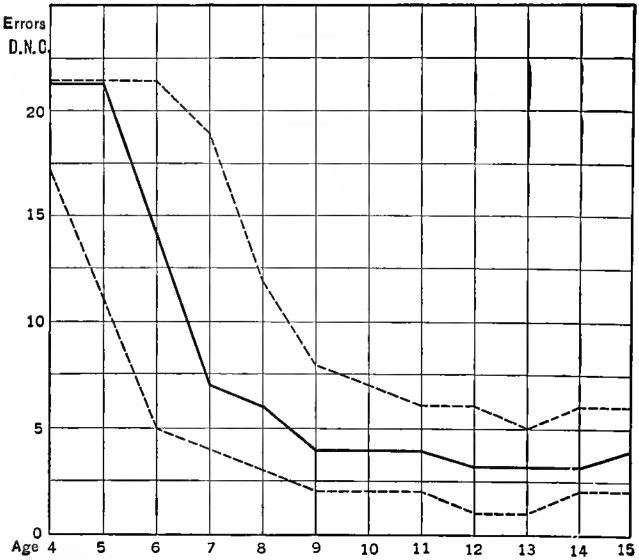
eleven. At ages four and five more than half of the children are unable to complete the test. It is evidently a distinctly harder test than the Seguin Form Board. The shortest time to complete the test is between 20 and 30 seconds. No children under ten are able to complete it in this shortest time period. The graph shows a very marked increase in ability to solve the test up to about age twelve, and the amount of variation decreases constantly to that age.

Errors. (Table 5 and Graph 10.) The distribution of the errors shows a great scattering, although the median as shown on the graph indicates a constant decrease in the number of errors up to age twelve. Three or four errors is the general expectation for children above age eight. At this age and below the number increases rapidly from

TABLE 5. THE FIVE FIGURE FORM BOARD. ERRORS.

Age	4	5	6	7	8	9	10	11	12	13	14	15	16	
Errors														
0	..	..	2	4	4	5	9	8	11	6	..	1	1	..
1	..	1	3	5	10	11	13	18	19	11	6	3	..	..
2	..	..	6	3	11	10	16	19	15	8	5	4	..	..
3	..	1	6	14	22	10	13	11	15	7	11	1	..	..
4	..	2	6	10	6	18	14	16	6	6	3	1	2	..
5	..	1	7	7	14	3	7	10	10	10	3	1	..	..
6	..	3	5	9	11	6	5	12	10	1	1	5	..	..
7	..	4	3	12	11	6	6	6	5	2	2	..	..	..
8	..	1	5	3	5	2	6	5	..	..	1	..	..	..
9	..	..	3	3	2	2	5	2	..	..	..	1	..	..
10	..	..	4	2	3	6	2	3	2	1	1	..	..	..
11	..	2	1	3	2	2	3	2	2	1	1	1	..	..
12	..	..	1	2	3	2	1	1	3	..	..	..	..	..
13	..	..	4	..	1	..	..	3	..	..	1	..	..	..
14	..	..	2	5	4	..	3	1	2	1	..	..	..	..
15	1	3	8	..	5	..	2	1	1	..	..	..	..	..
17	1	4	3	2	5	..	..	..	..	..	1	1	..	..
19	..	2	..	2	1	..	..	1	1	..	1	..	..	..
21	..	1	2	1	3	..	..	2	..	1	..	..	..	..
23	..	1	1	1	1	..	..	..	..	..	..	..	..	..
25	..	..	1	..	1	1	..	..	..	..	..	..	..	..
27	..	..	1	..	..	..	..	..	..	..	..	..	..	..
29	..	..	..	..	1	..	..	..	1	..	..	..	..	..
31	1	..	1	..	1	..	..	..	..	..	..	..	..	..
D. N. C.	8	28	39	24	10	9	1	1	..	..	..	..	..	..
Total......	11	54	114	112	137	93	106	122	103	55	37	19	3	966
75% ile....	..	11	5	4	3	2	2	2	1	1	2	2	..	..
Median....	D. N. C.	D. N. C.	14	7	6	4	4	4	3	3	3	4	..	..
25% ile....	..	D. N. C.	D. N. C.	19	12	8	7	6	6	5	6	6	..	..
Quartile....	..	..	..	7.5	4.5	3.0	2.5	2.0	2.5	2.0	2.0	2.0	..	..

six to inability to complete the test. At all ages from six upwards there are children who complete the test without making any errors.



GRAPH 10.—The Five Figure Form Board. Errors.

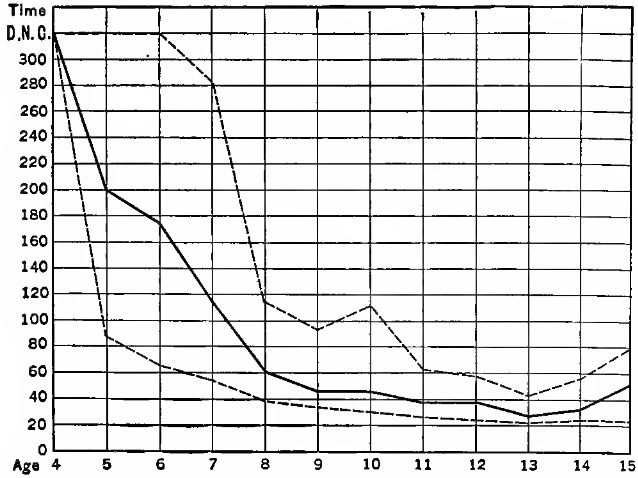
THE TWO FIGURE FORM BOARD

Time. (Table 6 and Graph 11.) The distribution again suggests a slight element of chance due to the puzzle nature of the board. The scattering at some ages is very great. The median, however, shows a very marked decrease at every age from four to nine, from which point onward it drops much more slowly to age thirteen. The variation of the middle 50 per cent as shown by the percentile curves is fairly great at the lower ages and the tendency for it to decrease with age is interrupted by the rise of the 25 percentile at age ten. There are cases of inability to complete this test at all ages from four to eleven inclusive and in addition an isolated case at age fourteen. The shortest time taken to complete the test lies between 10 and 20 seconds and there are children at all ages from eight to fifteen who are able to complete the test within this shortest time period.

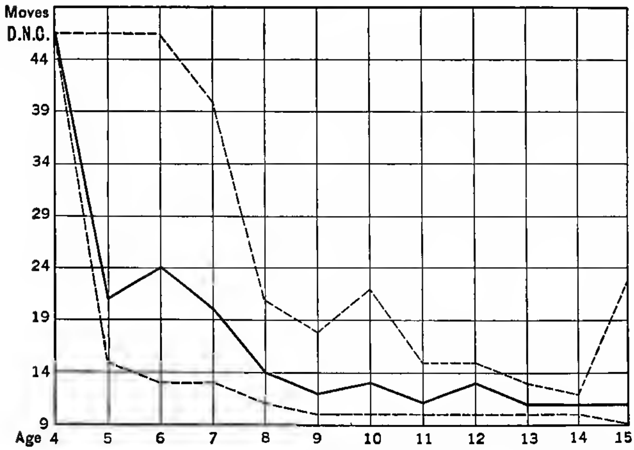
Moves. (Table 7 and Graph 12.) What has been said in regard to the time applies equally well to the number of moves taken to complete the test.

TABLE 6. THE TWO FIGURE FORM BOARD. TIME.

Age	4	5	6	7	8	9	10	11	12	13	14	15	16	
Time														
10	..	..	..	..	1	2	2	4	2	9	4	2	..	
20	..	..	3	4	15	15	22	43	27	22	13	7	2	..
30	..	1	5	6	22	21	18	23	24	9	6	3	..	..
40	..	1	10	12	19	11	15	15	15	7	2	..	..	..
50	..	2	8	13	12	8	6	5	11	2	6	1	..	..
60	..	4	8	7	10	7	4	7	10	..	1	..	..	..
70	..	1	5	6	8	3	2	4	1	1	..	2	..	..
80	..	4	1	2	10	3	3	7	3	1	1	..	..	..
90	..	1	5	2	2	5	2	..	2	3	..	2	..	..
100	..	6	2	9	12	4	14	3	5	..	3	1	..	..
125	..	4	6	6	3	2	4	1	1	..	2	1	..	..
150	..	2	6	7	4	4	3	2	1	1	..	..	..	..
175	..	1	4	6	4	..	1	1	..	..	..	..	..	..
200	..	2	1	1	3	1	2	2	..	..	..	..	..	..
225	..	1	3	2	1	1	..	1	..	..	..	..	..	..
250	..	..	2	3	2	1	1	1	..	..	..	..	..	..
275	..	3	2	3	1	..	..	1	..	..	..	..	..	..
300	..	..	2	1	1	..	..	..	1	..	..	..	..	..
D. N. C.	10	21	45	25	12	7	7	3	..	..	1	..	..	..
Total......	10	54	118	115	142	95	106	123	103	55	30	19	2	980
75% ile....	D. N. C.	89	65	56	39	34	31	27	26	23	26	25	..	..
Median....	D. N. C.	200	175	116	62	47	47	38	39	29	35	54	..	..
25% ile....	D. N. C.	D. N. C.	D. N. C.	284	116	94	112	64	59	43	58	80	..	..
Quartile....	..	..	..	114.0	38.5	30.0	40.5	18.5	16.5	10.0	16.0	27.5	..	..



GRAPH 11.—The Two Figure Form Board. Time.



GRAPH 12.—The Two Figure Form Board. Moves.

TABLE 7. THE TWO FIGURE FORM BOARD. MOVES.

Age	4	5	6	7	8	9	10	11	12	13	14	15	16	
Moves														
9	..	..	9	4	16	11	12	23	14	9	8	5	..	..
10	..	1	6	7	12	14	15	20	13	13	8	4	2	..
11	..	3	9	8	18	17	17	21	17	11	6	2	..	..
12	..	2	5	8	8	9	5	11	13	5	7	2	..	..
13	..	2	6	9	13	9	9	7	7	7	1	..	..	..
14	..	2	4	5	6	4	3	4	8	..	1	..	..	..
15	..	5	1	3	6	3	5	7	7	2	..	..	..	..
16	..	3	3	5	7	..	2	3	6	..	1	..	..	..
17	..	2	1	3	4	4	..	2	3	1	..	..	..	..
18	..	2	4	1	3	2	3	1	3	1	1	..	..	..
19	..	1	1	3	5	2	3	3	3	1	..	..	..	..
20	..	..	..	4	4	2	2	1	..	..	..	..	..	..
21	..	2	1	4	6	1	1	3	1	1	1	1	..	..
22	..	1	4	2	1	1	2	3	..	1	..	..	..	..
23	..	..	2	1	3	1	3	1	1	..	..	2	..	..
24	..	..	..	2	3	1	3	..	1	..	..	..	..	..
25	..	3	9	4	5	4	6	3	4	..	1	3	..	..
30	..	..	1	4	4	1	2	2	..	1	1	..	..	..
35	..	2	2	6	3	..	4	1	1	1	1	..	..	..
40	..	..	3	3	1	2	1	3	..	..	..	..	..	..
45	..	2	2	4	2	..	1	1	1	1	..	..	..	..
D. N. C.	10	21	45	25	12	7	7	3	..	..	1	..	..	..
Total......	10	54	118	115	142	95	106	123	103	55	38	19	2	980
75% ile....	D. N. C.	15	13	13	11	10	10	10	10	10	10	9	..	..
Median....	D. N. C.	21	24	20	14	12	13	11	13	11	11	11	..	..
25% ile....	D. N. C.	D. N. C.	D. N. C.	40	21	18	22	15	15	13	12	23	..	..
Quartile....	..	..	..	13.5	5.0	4.0	6.0	2.5	2.5	1.5	1.0	7.0	..	..

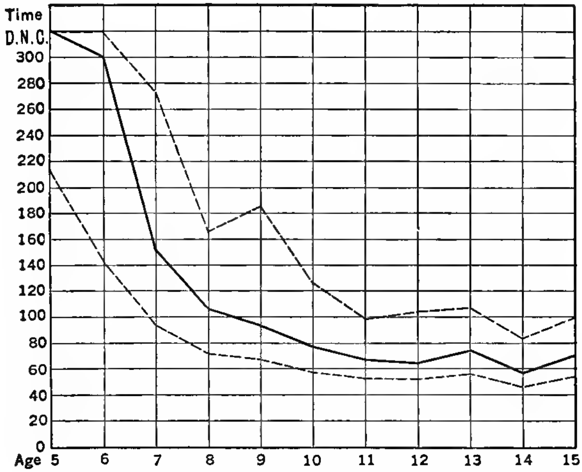
The scattering is fairly great and the median performance does not improve with regularity from age to age. The percentiles indicate a wide variation at the lower ages. There are cases of children completing the test with the shortest number of moves possible (i.e., 9) at all ages from six upwards.

The Casuist Form Board

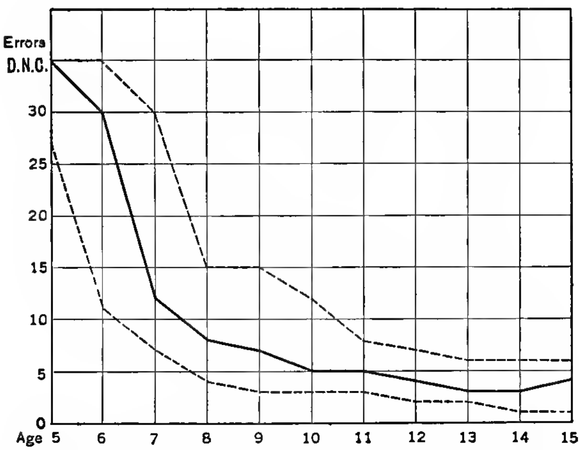
Time. (Table 8 and Graph 13.) Although a tendency to scattering is noticeable, the distribution, on the whole, is fairly good, and this is reflected in the graph showing the medians and percentiles. The median drops rapidly down to age eleven or twelve, from which point there is practically no increase in rapidity in completing the

Table 8. The Casuist Form Board. Time.

Age	4	5	6	7	8	9	10	11	12	13	14	15	
Time													
20	..	..	..	..	..	..	1	1	..	1	1	..	..
30	..	..	..	..	..	2	4	10	2	2	1	..	..
40	..	..	..	3	7	13	9	14	9	2	9	2	..
50	..	..	4	6	17	10	21	23	8	7	9	2	..
60	..	..	2	7	10	14	18	20	12	7	3	2	..
70	..	..	..	6	11	13	18	8	7	6	2	2	..
80	..	1	2	8	11	7	10	7	1	3	3	3	..
90	..	..	3	3	13	7	4	13	2	3	1	1	..
100	..	3	8	18	16	11	15	9	6	6	1	2	..
125	..	2	10	10	15	7	8	4	1	3	1	..	..
150	..	..	7	9	13	5	10	4	1	2	1	..	..
175	..	1	4	9	5	5	2	6	1	..	2	1	..
200	..	2	2	6	4	3	3	..	1	..	..	..	..
225	..	..	7	4	3	3	2	1	2	..	..	..	..
250	..	2	3	4	1	2	1	2	1	..	..	..	..
275	..	2	1	2	1	2	3	..	1	1	..	..	..
300	..	..	2	1	1	2	1	2	1	1	1	..	..
D. N. C.	1	16	51	27	16	15	4	2	..	..	..	1	..
Total...........	1	29	106	123	144	121	134	126	56	44	35	16	935
75% ile.........	..	212	143	93	72	69	59	53	53	57	47	55	..
Median.........	..	D. N. C.	300	154	106	93	78	68	66	75	58	70	..
25% ile.........	..	D. N. C.	D. N. C.	274	165	185	128	99	104	108	84	100	..
Quartile........	..	..	..	90	46	58	34	23	25	25	18	22	..



GRAPH 13.—The Casuist Form Board. Time.



GRAPH 14.—The Casuist Form Board. Errors.

TABLE 9. THE CASUIST FORM BOARD. ERRORS.

Age	4	5	6	7	8	9	10	11	12	13	14	15	
Errors													
0	..	..	1	..	2	5	5	7	3	2	2	2	..
1	..	..	1	1	7	10	10	13	5	5	11	2	..
2	..	..	4	5	11	11	17	8	9	8	5	2	..
3	..	..	1	5	10	14	18	21	9	8	4	..	..
4	..	..	3	8	15	6	11	12	8	3	5	4	..
5	..	1	1	5	8	6	14	12	6	4	..	..	..
6	..	..	2	5	8	8	4	14	2	5	1	3	..
7	..	..	2	5	6	5	10	4	2	2	..	..	..
8	..	1	1	10	8	3	5	9	2	1	2	1	..
9	..	..	6	10	8	2	1	1	1	..	..	..	..
10	..	2	4	4	2	7	4	5	1	2	..	..	..
11	..	..	3	4	4	9	1	3	..	..	..	..	..
12	..	..	1	4	7	2	5	3	1	..	3	..	..
13	..	..	4	8	7	..	2	5	..	..	..	..	..
14	..	..	3	3	3	2	4	..	1	2	1	1	..
15	..	1	1	3	8	4	5	2	1	..	..	..	..
17	..	2	4	5	5	2	3	1	1	..	1	..	..
19	..	..	..	..	5	1	2	..	1	..	..	..	..
21	..	..	4	2	..	2	1	1	1	..	..	..	..
23	..	..	4	2	1	1	4	..	..	..	..	..	..
25	..	1	2	2	2	..	..	1	..	..	..	..	..
27	..	1	..	1	1	1	1	..	1	..	..	..	..
29	..	2	2	2	1	2	1	..	..	1	..	..	..
31	..	2	1	2	1	3	2	1	1	..	..	..	..
D. N. C	1	16	51	27	16	15	4	3	..	..	..	1	..
Total...........	1	29	106	123	144	121	134	126	56	43	35	16	934
75% ile.........	..	27	11	7	4	3	3	3	2	2	1	1	..
Median.........	..	D. N. C.	30	12	8	7	5	5	4	3	3	4	..
25% ile.........	..	D. N. C.	D. N. C.	30	15	15	12	8	7	6	6	8	..
Quartile........	..	..	..	11.5	5.5	6.0	4.5	2.5	2.5	2.0	2.5	2.5	..

test. The decrease in the range of variation, as indicated by the percentiles, is fairly constant at all ages, with the exception of age nine, which shows an increase over the preceding age. At all ages from four to eleven there are children who are unable to complete the test within the time limit, while at age fifteen there is one isolated case. The shortest time taken to complete the test lies between 20 and 30 seconds. No child below age ten completes the test within this shortest time period.

Errors. (Table 9 and Graph 14.) The decrease of the curve for the median showing the number of errors is fairly constant and uniform down to age thirteen, and the percentile curves follow the same general tendency. The largest number of errors made, while completing the test within the time limit imposed, is about 30. At almost all ages there are cases of children completing the test without error. The table of distribution shows a fair amount of scattering at all ages.

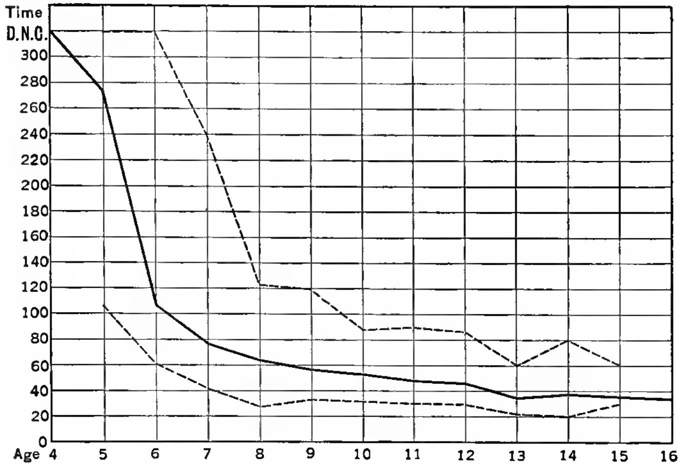
The Triangle Test

Time. (Table 10 and Graph 15.) As in the previous test the table shows a fair amount of scattering at all ages. The median decreases constantly but rather slowly after age nine. The amount of variation as shown by the percentiles tends, on the whole, to decrease with increasing age, although there are the usual irregularities at the upper ages. There are cases of inability to complete the test at all ages from four to eleven. The shortest time record is less than 10 seconds, made by two twelve-year-olds.

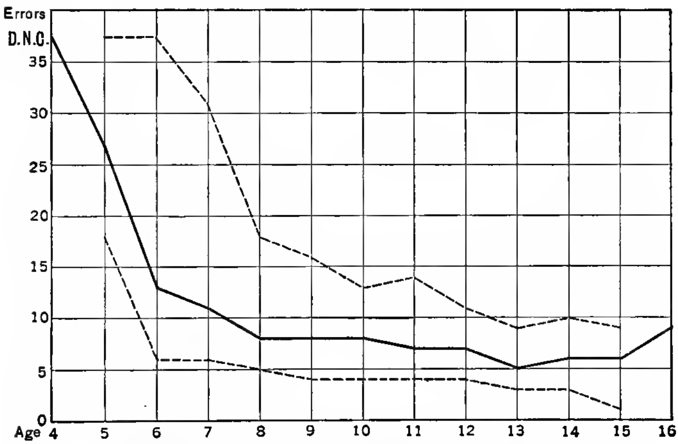
Errors. (Table 11 and Graph 16.) The same amount of scattering is shown in the table of errors as in the table of time, and the curve for the median, on the whole, presents much the same appearance as the curve for the median time. The greatest number of errors made is about 35 and there are isolated cases at many ages completing the test without any errors. These cases seem to

TABLE 10. THE TRIANGLE TEST. TIME.

Age	4	5	6	7	8	9	10	11	12	13	14	15	16	
Time														
—10	..	..	..	..	..	..	..	..	2	..	..	..	..	..
10	..	..	2	3	4	3	10	8	5	9	6	3	..	..
20	..	..	4	6	11	7	6	12	10	6	2	..	..	..
30	..	1	2	10	10	14	10	10	11	8	5	7	1	..
40	..	..	3	11	13	7	8	11	8	7	3	1	1	..
50	..	..	5	6	8	5	8	5	7	1	2	1	..	..
60	..	1	4	4	8	3	1	7	4	1	..	..	..	..
70	..	2	3	3	5	1	8	5	3	2	1	1	..	..
80	..	1	5	4	7	2	8	2	4	3	1	..	..	..
90	..	..	3	3	1	4	3	1	2	..	..	..	..	..
100	..	3	3	3	7	8	5	11	5	1	3	3	..	..
125	..	..	3	5	1	6	4	2	1	2	2	..	..	..
150	..	..	1	3	4	3	2	2	3	1	..	..	..	..
175	..	2	2	..	7	1	..	1	..	1	..	..	..	.
200	..	2	3	1	1	2	..	..	1	..	..	..	..	.
225	..	1	..	2	1	..	1	2	2	..	..	..	..	..
250	..	1	..	6	1	1	..	..	2	..	1	..	..	..
275	..	2	1	1	1	..	..	..	..	..	..	..	..	..
300	..	..	1	..	..	..	..	1	..	..	..	..	..	..
D. N. C.	1	13	22	14	8	4	3	1	..	..	..	..		..
Total......	1	29	67	85	98	71	77	81	70	42	26	16	2	665
75% ile....	..	106	61	42	29	35	33	30	30	22	20	31		..
Median....	D. N. C.	275	108	77	64	58	55	49	48	37	39	37	35	..
25% ile....	..	D. N. C.	D. N. C.	240	123	120	89	90	88	60	80	60	..	..
Quartile....	..	..	..	99.0	47.0	42.5	28.0	30.0	28.0	19.0	30.0	14.5	..	..



GRAPH 15.—The Triangle Test. Time.



GRAPH 16.—The Triangle Test. Errors.

TABLE 11. THE TRIANGLE TEST. ERRORS.

Age	4	5	6	7	8	9	10	11	12	13	14	15	16	
Errors														
0	..	..	1	..	1	..	1	1	1	..	..	3	..	..
1	..	..	..	3	4	2	2	2	5	4	2	1	..	..
2	..	..	2	..	5	1	4	10	3	4	3	..	..	..
3	..	..	2	11	7	9	8	6	6	3	2	..		
4	..	..	5	2	7	8	6	6	5	7	3	1	..	.
5	..	1	4	2	11	4	5	5	6	5	2	3		..
6	..	..	5	8	7	4	5	2	5	2	2	1	..	..
7	..	..	2	6	5	2	3	8	5	3	3	3	..	.
8	..	..	..	4	5	5	9	4	7	1	1	..	1	..
9	..	..	4	2	6	4	8	4	5	3	..	1	.	.
10	..	2	..	4	1	1	4	1	4	1	2	1	1	..
11	..	2	3	3	..	3	1	4	2	..	1	..	..	..
12	..	..	4	2	4	1	1	5	..	2	2	..	..	..
13	..	..	1	5	3	3	3	2	1	..	..	..		..
14	..	1	..	3	2	1	3	3	1	1	..	..		..
15	..	..	1	..	..	4	3	2	2	..	1	..	..	..
16	..	..	2	..	1	3	..	3	1	1	1	..		..
17	..	..	3	..	1	4	1	2	1	2	..	1	..	..
18	..	1	1	3	3	..	..	1	..	..	..	..	..	..
19	..	2	1	1	2	1	1	3	3	..	..	..	..	..
20	..	4	1	4	4	4	3	4	3	2	..	1	..	
25	..	3	1	2	3	1	3	..	2	1	..	..	..	.
30	..	..	1	3	3	1	..	1	1	..	..	..	..	
35	..	..	1	3	5	1	..	1	1	..	..	..	..	..
D. N. C.	1	13	22	14	8	4	3	1	..	..	..	..	..	..
Total......	1	29	67	85	98	71	77	81	70	42	25	16	2	664
75% ile....	..	18	6	6	5	4	4	4	4	3	3	1	..	
Median....	D. N. C.	27	13	11	8	8	8	7	7	5	6	6	9	..
25% ile....	..	D. N. C.	D. N. C.	31	18	16	13	14	11	9	10	9	.	..
Quartile....	..	..	..	12.5	6.5	6.0	4.5	5.0	3.5	3.0	3.5	4.0	.	..

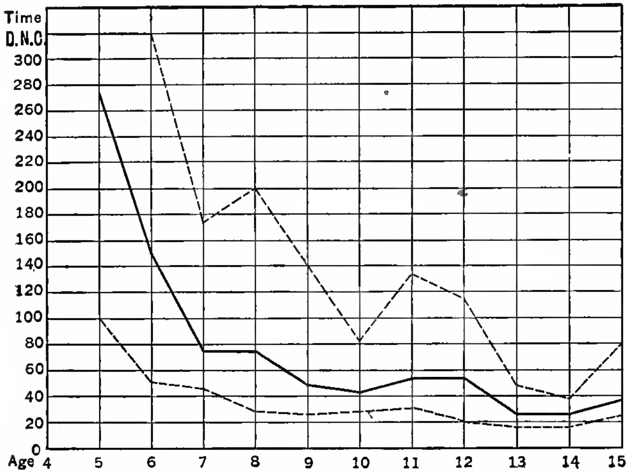
be due to the puzzle nature of the test, which allows for the entrance of a chance solution every now and then.

The Diagonal Test

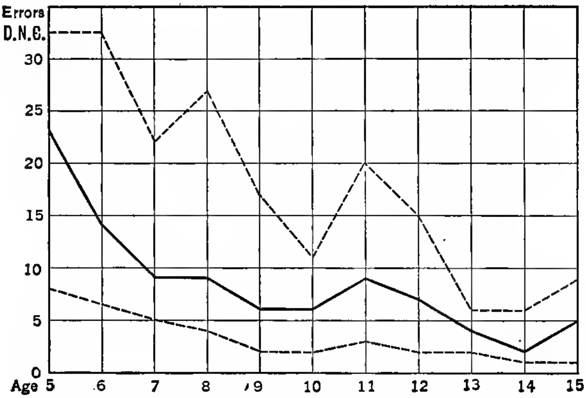
Time. (Table 12 and Graph 17.) The irregularity of the curve for the medians and percentiles bears out what was obvious to the authors while making the tests, i.e., the element of chance entering into this test. This is also shown by the scattering in the table of distribution. The puzzle

Table 12. The Diagonal Test. Time.

Age	4	5	6	7	8	9	10	11	12	13	14	15	16	
Time														
—10	..	..	..	..	..	..	1	..	1	1	..	1	..	..
10	..	1	4	7	11	13	12	15	17	16	11	1	..	..
20	..	..	6	6	14	10	13	9	5	8	6	2	..	..
30	..	1	3	4	11	8	11	7	9	5	4	5	..	..
40	..	1	3	8	3	5	9	8	1	2	1	1	1	..
50	..	1	2	7	3	1	5	4	5	3	2	1	..	..
60	..	..	3	9	2	3	3	1	4	2	1	1	..	..
70	..	1	4	6	6	3	3	4	4	1	..	..	1	..
80	1	..	3	3	3	4	4	1	3	..	..	1	..	..
90	..	2	..	3	4	2	..	4	2	1	..	..	..	..
100	..	2	2	5	4	3	4	5	3	..	1	1	..	..
125	..	2	3	4	4	3	4	7	4	1	..	1	..	..
150	..	2	3	2	2	3	..	2	3	2	..	..	..	..
175	..	1	4	1	3	2	1	4	2	..	..	..	..	..
200	..	2	3	1	3	2	1	2	4	..	1	1	..	..
225	..	..	2	2	5	..	3	2	..	..	..	..	..	..
250	..	..	..	1	3	..	..	1	..	..	..	..	..	..
275	..	1	2	1	2	1	1	..	..	..	..	..	..	..
300	..	..	1	..	..	2	..	..	..	..	..	..	..	..
D. N. C.	..	12	19	15	14	6	2	5	3	..	..	..	..	..
Total......	1	29	67	85	97	71	77	81	70	42	27	16	2	665
75% ile....	..	100	51	45	29	25	27	31	20	17	17	25	..	..
Median....	..	275	150	75	76	49	42	54	54	25	25	38	..	..
25% ile....	..	D. N. C.	D. N. C.	174	200	141	83	135	116	49	39	80	..	..
Quartile....	..	..	..	64.5	85.5	58.0	28.0	52.0	48.0	16.0	11.0	27.5	..	..



GRAPH 17.—The Diagonal Test. Time.



GRAPH 18.—The Diagonal Test. Errors.

TABLE 13. THE DIAGONAL TEST. ERRORS.

Age	4	5	6	7	8	9	10	11	12	13	14	15	16	
Errors														
0	..	1	2	4	7	9	4	3	6	6	6	2	..	..
1	..	..	5	4	7	7	8	7	10	5	7	2	..	..
2	..	..	1	3	5	7	10	7	5	4	3	..	..	..
3	..	2	4	6	5	5	2	7	5	5	..	2	..	..
4	..	2	2	3	6	1	6	5	5	3	2	2	..	..
5	..	1	1	5	7	4	5	3	2	7	..	3	1	..
6	..	..	2	7	4	5	5	3	..	4	3	..	..	..
7	..	2	2	2	2	3	6	2	4	2	2	1	..	..
8	..	1	..	5	6	..	6	1	3	1	..	..	..	..
9	1	..	5	5	2	3	3	4	4	..	1	1	..	..
10	..	1	4	2	1	2	..	2	3	2	..	..	..	..
11	..	1	3	3	1	..	4	1	1	..	1	..	1	..
12	..	..	1	4	2	3	1	3	..	..	..	..	..	..
13	..	2	1	1	1	..	..	3	1	..	..	..	..	..
14	..	1	3	2	2	1	..	2	3	1	1	1	..	..
15	..	..	1	1	1	3	4	..	2	1	..	..	..	..
16	..	..	1	1	2	1	3	1	..	..	..	..	..	..
17	..	..	..	1	4	1	..	2	..	..	..	..	..	..
18	..	..	3	1	1	2	..	4	..	..	..	..	..	..
19	..	..	2	2	1	..	1	1	2	..	..	1	..	..
20	..	1	1	3	3	3	3	4	5	..	..	..	..	..
25	..	1	1	2	6	2	..	6	4	1	1	..	..	..
30	..	1	1	1	4	1	3	2	1	..	..	..	..	..
35	..	..	2	2	3	2	1	3	1	..	..	1	..	..
D. N. C.	..	12	19	15	14	6	2	5	3	..	..	..	..	..
Total......	1	29	67	85	97	71	77	81	70	42	27	16	2	665
75% ile....	..	8	7	5	4	2	2	3	2	2	1	1	..	..
Median....	..	23	14	9	9	6	6	9	7	4	2	5	..	..
25% ile....	..	D. N. C.	D. N. C.	22	27	17	11	20	15	6	6	9	..	..
Quartile....	..	..	..	8.5	11.5	7.5	4.5	8.5	6.5	2.0	2.5	4.0	..	..

nature of the test is obvious, although we do not believe that the puzzle feature is so great as to make the test worthless when used in a group of tests. There are cases of inability to complete the test at all ages up to twelve. Only 4 cases are able to complete the test in less than 10 seconds.

Errors. (Table 13 and Graph 18.) The table and graph for the errors show much the same characteristics as the table and graph for the time. The irregularity of the 25 percentile curve is very marked, and this means a great fluctuation in variability from age to age. The greatest number of errors made by those completing the test is about 35. There are cases at all ages of individuals completing the test without error.

Healy Puzzle "A"

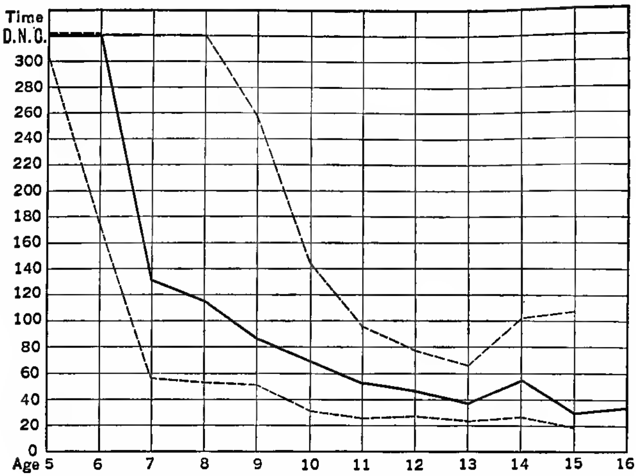
Time. (Table 14 and Graph 19.) The distribution table shows a considerable amount of scattering at all ages. At every age, from four to twelve inclusive, and also at age fourteen, there are children who fail to complete the test within the time limit. The shortest time taken is 5 seconds or less, and there are cases of children who complete the test within this short limit of time at ages eleven, twelve, thirteen and fifteen. The graph of the median shows a fairly constant and steady decrease in time up to age thirteen. There is, however, a rather large variability, as indicated by the percentiles, at all ages up to age eleven, with the usual variation of the medians and percentiles in the upper ages (above age thirteen).

Our results seem somewhat at variance with Healy's norms. He says: "No normal person over eight or nine years should fail to do it in 5 minutes."[2] Referring to our table of distribution, we find 20 out of 122 nine-year-olds, 11 out of 117 ten-year-olds, and a few eleven-, twelve- and fourteen-year-olds who fail to fulfill these conditions. It is hardly conceivable that all these individuals are

TABLE 14. HEALY PUZZLE "A." TIME.

Age	4	5	0	7	8	9	10	11	12	13	14	15	18	
Time														
5	..	..	..	..	..	..	..	2	2	1	..	2	..	..
10	..	..	2	5	7	6	15	15	10	11	7	8	2	..
20	1	..	2	6	15	13	13	19	15	8	11	5	..	..
30	..	..	1	10	10	8	7	7	13	9	1	1	3	..
40	..	3	1	9	4	9	11	6	6	7	2	4	..	..
50	..	1	2	8	3	8	8	10	7	4	2	..	..	..
60	..	..	1	5	5	6	7	5	8	5	..	..	1	..
70	..	..	1	2	5	7	4	8	6	3	3	..	..	..
80	..	1	1	1	10	7	7	4	4	2	5	..	..	..
90	..	..	4	4	4	4	5	5	..	2	1	1	..	..
100	..	..	2	8	5	1	4	4	..	..	4	..	1	..
110	..	..	..	..	8	1	3	2	..	..	1	1	..	..
120	..	..	..	2	1	2	4	..	3	..	1	1	..	..
130	..	..	1	..	1	5	1	1	3	..	..	..	..	..
140	..	1	1	2	3	2	2	2	1	..	..	..	..	..
150	..	..	8	3	6	8	4	3	2	1	2	..	..	..
175	..	3	8	3	6	4	2	1	2	1	2	1	..	..
200	..	..	5	2	3	5	4	1	..	1	..	1	..	..
225	..	..	5	4	..	3	4	4	1	..	..	1	..	..
250	..	..	2	4	3	3	2	2	1	..	..	..	..	..
275	..	1	3	5	3	2	1	..	..	1	..	1	..	..
301	..	1	..	..	..	..	..	..	..	..	..	..	..	..
D. N. C.	4	32	80	57	45	20	11	4	4	..	2	..	..	..
Total.......	5	43	108	138	147	122	117	105	88	58	44	25	7	1005
75% ile.....	..	301+	175	58	53	52	32	25	27	23	28	19	..	..
Median.....	D.N.C.	D.N.C.	D.N.C.	131	117	88	70	54	46	38	55	30	35	..
25% ile.....	..	D.N.C.	D.N.C.	D.N.C.	D.N.C.	258	145	98	78	68	102	108	..	..
Quartile.....	..	..	..	..	..	103.0	58.5	35.5	25.5	21.5	38.0	45.0	..	..

[2] Healy, W.: The Individual Delinquent, Little, Brown and Company (1915), p. 107.



GRAPH 19.—Healy Puzzle "A." Time.

to be classified as abnormal. We do not believe in regard to the ten- and eleven-year-olds who fail to complete the performance within 5 minutes that we are warranted in classifying them as below normal.

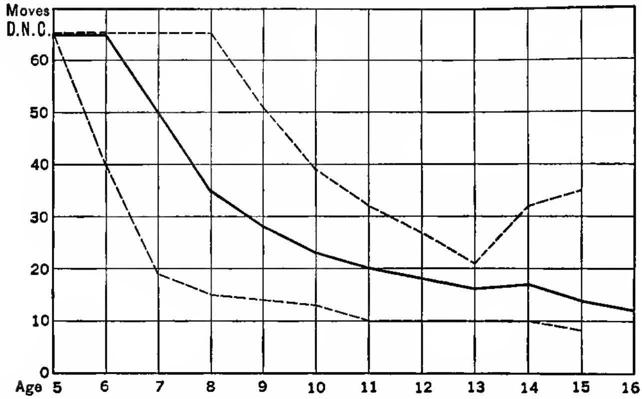
The only other norms for time comparable with ours are those of Bruckner and King[3] for eight- and ten-year-old children. Their median time for eight-year-olds is 140 seconds, while ours is 117 seconds, i.e., considerably shorter. It is interesting to note that the medians for the ten-year-olds differ only by one second, theirs being 69 and ours 70 seconds.

[3] Bruckner, L., and King, I.: "A Study of the Fernald Form Board," *Psychological Clinic,* Vol. ix, No. 9 (1916), pp. 249-258.

Moves. (Table 15 and Graph 20.) The graph and table for the number of moves indicate much the same features as have been noted in dealing with the time. The greatest number of moves taken by any one child completing the test within the time limit is about 100. Five moves is the fewest number by which the test can be completed. This is possible if the child places all the five pieces

TABLE 15. HEALY PUZZLE "A." MOVES.

Age	4	5	6	7	8	9	10	11	12	13	14	15	16	
Moves														
5	1	..	5	10	10	14	24	27	22	13	13	9	2	..
10	..	4	2	16	18	18	12	17	20	12	8	5	3	..
15	..	1	5	11	11	18	20	0	9	15	2	2	1	..
20	..	..	2	12	10	8	7	16	8	7	4	1	..	..
25	..	2	7	9	7	6	14	6	11	3	4	2	..	..
30	..	..	3	3	8	8	7	7	3	1	7	..	..	..
35	..	..	3	1	8	8	5	6	2	..	2	3	1	..
40	..	..	7	2	4	9	3	4	5	1	..	..	..	..
45	..	..	3	5	5	1	2	1	1	..	..	..	..	..
50	..	..	7	2	2	5	1	..	1	..	2	..	..	..
55	..	..	..	2	1	2	4	1	..	1	..	..	..	..
60	..	..	3	2	3	1	2	..	1	2	..	..	..	..
65	..	..	..	..	..	..	..	4	1	..	..	2	..	..
70	..	..	..	3	..	1	2	1	..	..	..	1	..	..
75	..	..	..	3	1	2	1	1	..	1	..	..	..	..
80	..	..	..	..	1	..	1	..	..	..	..	..	..	..
85	..	..	..	..	1	..	..	..	..	..	..	..	..	..
90	..	..	1	..	..	..	..	1	..	..	..	..	..	..
95	..	..	..	..	1	..	..	..	..	..	..	..	..	..
100	..	..	..	..	..	..	1	..	..	..	..	..	..	..
D. N. C.	4	33	60	57	45	20	11	4	4	..	2	..	..	..
Total.......	5	40	108	138	145	121	117	105	88	56	44	25	7	999
75% ile.....	..	D.N.C.	40	19	15	14	13	10	10	10	10	8	..	..
Median.....	D.N.C.	D.N.C.	D.N.C.	50	35	28	23	20	18	16	17	14	12	..
25% ile.....	..	D.N.C.	D.N.C.	D.N.C.	D.N.C.	51	39	32	27	21	32	35	..	..
Quartile.....	..	..	..	..	..	18.5	13.0	11.0	8.5	5.5	11.0	13.5	..	..



GRAPH 20.—Healy Puzzle "A." Moves.

correctly at the first trial. We have a certain number of cases of this nature occurring at every age (with the exception of age five). We feel that this is due to the puzzle nature of the test, which allows an element of chance to enter into the solution.

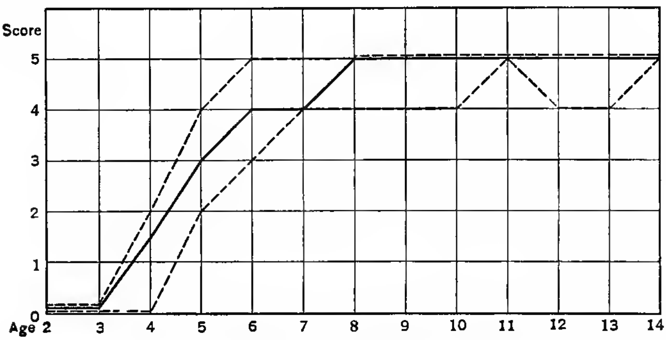
Many of the cases completing the test in 5 moves, particularly among the younger children, are due to pure chance. The child happened to place the first few blocks correctly and the rest of the solution followed without difficulty. In the solution of this test the element of chance seems to enter to a greater degree than in the other tests. We feel that, used by itself, it is very unreliable and that the only justification for its use is in a group of mental tests by means of which the chance element, if at work, will be modified by the performances on the other tests.

The Manikin Test

Score. (Table 16 and Graph 21.) The table shows an excellent distribution with comparatively little scattering. A score of 0 is made by some

Table 16. The Manikin Test. Score.

Age	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Score																
0	4	8	7	8	1	1	..	..	..	..	..	..	..	..	..	..
1	..	..	1	5	2	..	1	..	..	..	..	..	..	..	..	..
2	..	..	5	16	8	5	2	1	..	..	..	..	..	..	..	..
3	..	1	2	20	20	5	3	3	4	2	..	..	..	..	..	..
4	..	..	..	10	49	59	55	21	4	3	5	3	..	..	..	..
5	..	..	1	9	35	50	66	49	17	15	13	5	5	1	1	..
Total...	4	9	16	68	115	120	127	74	25	20	18	8	5	1	1	611
75% ile.	0	0	2	4	5	5	5	5	5	5	5	5	5	..	..	..
Median.	0	0	1.5	3	4	4	5	5	5	5	5	5	5	..	..	..
25% ile.	0	0	0	2	3	4	4	4	4	5	4	4	5	..	..	..
Quartile	..	..	1.0	1.0	1.0	0.5	0.5	0.5	0.5	0	0.5	0.5	0		..	..

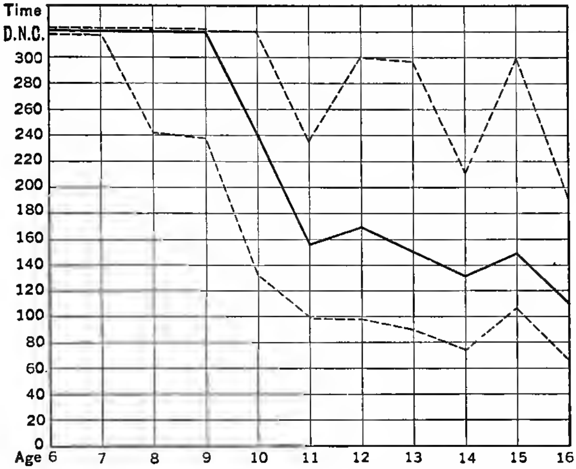


Graph 21.—The Manikin Test. Score.

cases at all ages from two to seven. The highest score is made by only one case at age four, and from then upwards by an increasing percentage of the cases at the other ages. The curve for the medians shows a very decided rise from age three up to age eight, where it reaches the maximum score, at which place it remains for all the other ages. The quartile is never greater than 1.0. The

TABLE 17. THE FEATURE PROFILE TEST. TIME.

Age	4	5	6	7	8	9	10	11	12	13	14	15	16	
Time														
10	..	..	..	..	..	..	..	..	..	..	..	..	..	..
20	..	..	..	..	..	1	..	..	..	..	..	..	..	..
30	..	..	..	..	..	..	..	..	2	2	4	1	1	..
40	..	..	..	..	..	..	5	3	2	4	3	..	..	..
50	..	..	..	..	1	..	2	3	5	2	3	2	1	..
61	..	..	..	..	..	2	2	2	1	2	5	1	3	..
70	..	..	..	..	1	2	2	11	6	6	2	1	1	..
80	..	..	..	..	..	2	3	3	2	1	3	1	3	..
90	..	..	..	1	2	..	1	..	3	1	2	1	..	..
100	..	..	..	1	2	1	5	8	6	6	7	9	1	..
121	..	..	..	..	2	..	4	6	7	6	5	1	1	..
135	..	..	..	..	2	1	4	5	3	4	4	..	..	..
150	..	..	..	1	5	3	3	7	3	4	4	2	1	..
165	..	..	..	1	2	2	3	2	3	3	3	..	2	..
181	..	..	..	1	1	2	5	5	2	1	2	1	2	..
200	..	..	..	2	1	2	5	6	5	2	4	3	1	..
225	..	..	1	..	1	3	4	5	2	1	1	..	..	..
241	..	..	..	1	2	4	8	1	3	4	2	..	1	..
275	..	..	..	..	4	1	4	2	2	3	3	2	..	..
D. N. C.	..	..	45	51	51	55	35	17	25	16	8	9	2	..
Total.....	..	..	46	59	77	81	95	86	82	68	65	34	20	713
75% ile...	..	..	D.N.C.	D.N.C.	241	239	134	99	99	90	75	104	68	..
Median...	..	..	D.N.C.	D.N.C.	D.N.C.	D.N.C.	240	157	170	150	132	150	110	..
25% ile...	..	..	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	237	300+	299	212	300+	190	
Quartile ..	..	..	..	..	..	..	..	69	100+	104	68	98+	61	..



GRAPH 22.—The Feature Profile Test. Time.

test as used with the method of scoring adopted seems to be an excellent one for differentiating abilities below age eight.

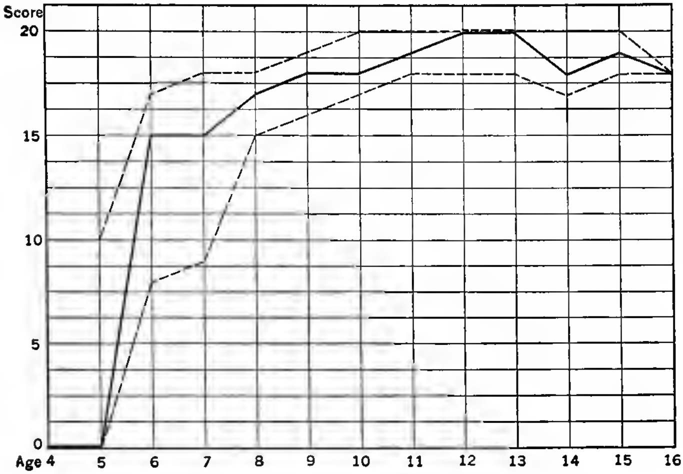
THE FEATURE PROFILE TEST

Time. (Table 17 and Graph 22.) The distribution table indicates the fairly wide range in time taken to solve this test. The shortest time taken to complete the test is made by a nine-year-old child, who took between 20 and 30 seconds. There are cases at all ages of inability to complete the test within the time limit. The curve for the medians shows a rather irregular decrease after age eleven, although the general tendency is downward up to age sixteen. The variation of the middle 50 per cent is rather large and varies considerably in amount for the ages tested. Knox[4] places this test in his thirteen-year-old group with a time limit of 10 minutes. We found, in our cases, that about 80 per cent of the twelve-year-olds, about 90 per

TABLE 18. THE SHIP TEST. SCORE.

Age	4	5	6	7	8	9	10	11	12	13	14	15	16	
Score														
0	1	18	8	6	7	..	..	..	..	..	..	..	..	..
1	..	1	2	1	1	..	..	..	..	..	..	..	..	..
2	..	..	1	2	2	..	..	..	..	..	..	..	..	..
3	..	..	..	1	1	1	..	..	..	..	..	..	..	..
4	..	..	2	3	1	1	..	..	..	..	..	..	..	..
5	..	1	1	1	1	..	..	..	..	..	..	..	..	..
6	..	..	2	..	..	1	...	..	..	..	..	..	..	..
7	..	..	1	2	..	..	..	..	..	..	..	..	..	..
8	..	..	1	5	1	..	..	..	1	..	..	..	..	..
9	..	..	..	1	..	..	1	..	..	..	..	..	..	..
10	..	2	3	4	2	..	2	..	1	..	..	..	..	..
11	..	1	2	1	2	..	1	..	..	..	..	..	..	..
12	..	1	1	2	..	..	1	..	1	..	..	..	..	..
13	..	1	5	8	1	2	3	..	..	..	1	..	..	..
14	..	1	5	4	3	2	1	1	..	..	..	..	..	..
15	..	1	7	7	10	5	..	4	3	..	..	..	..	..
16	..	1	6	8	11	8	9	5	2	3	3	1	..	..
17	..	1	7	6	11	9	9	10	8	6	5	3	..	..
18	..	..	5	10	20	13	13	13	8	4	7	3	2	..
19	..	..	..	3	5	13	9	11	9	4	3	2	..	..
20	..	..	8	10	19	17	28	37	37	25	8	7	..	..
Total	1	29	67	85	98	72	77	81	70	42	27	16	2	667
75% ile	..	10	17	18	18	19	20	20	20	20	20	20	..	..
Median	..	0	15	15	17	18	18	19	20	20	18	19	..	..
25% ile	..	0	8	9	15	16	17	18	18	18	17	18	..	..
Quartile	..	5.0	4.5	4.5	1.5	1.5	1.5	1.0	1.0	1.0	1.5	1.0	..	..

[4] Knox, H. A.: "A Scale, Based on the Work at Ellis Island, for Estimating Mental Defect," *Journal of the American Medical Association,* Vol. lxii (March 7, 1914).



GRAPH 23.—The Ship Test. Score.

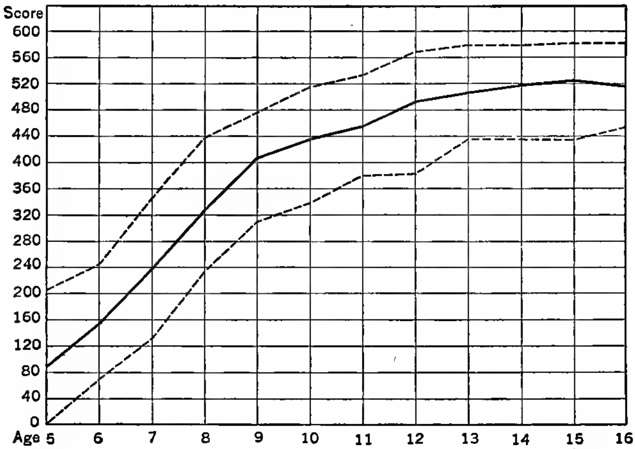
cent of the thirteen-year-olds and about 95 per cent of the fourteen-year-olds finish the test within 5 minutes. This would make it a very easy thirteen-year-old test without taking into consideration the fact that Knox allows a time limit of 10 minutes.

THE SHIP TEST

Score. (Table 18 and Graph 23.) The distribution here is fairly good, inasmuch as the scattering is not very great. A score of 0 is made by some cases at ages four, five, six, seven and eight. The highest score, denoting a perfect solution of the test, is made by some cases at each age from six to fourteen inclusive. The median shows a constant increase in ability to perform the test from age five up to age twelve. The variability at most ages is not very great and the quartile diminishes markedly in the upper ages. The test seems to discriminate well at all ages from five to eleven.

The Picture Completion Test

Score. (Graph 24.) For the table of distribution and an extended discussion of it see Pintner and Anderson.[5]



Graph 24.—The Picture Completion Test. Score.

The curve for the medians shows a steady and gradual increase up to age fifteen. It is interesting to note the slight drop from age fifteen to the adults. The amount of variability in score made by

[5] Pintner, R., and Anderson, M. M.: The Picture Completion Test. *Educational Psychology Monographs,* Warwick and York, Baltimore.

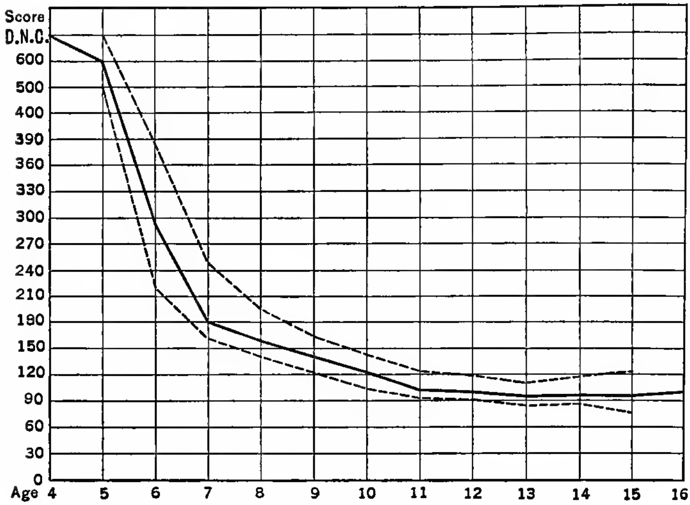
the middle 50 per cent is remarkably constant at all ages.

The Substitution Test

Score. (Table 19 and Graph 25.) This distribution is exceptionally good. There is very little

Table 19. The Substitution Test.' Score.

Age	4	5	6	7	8	9	10	11	12	13	14	15	16	
Score														
60	..	..	..	..	..	..	..	1	1	1	1	1	..	..
70	..	..	..	..	..	..	1	7	7	6	2	4	..	..
80	..	..	..	..	..	1	5	4	7	6	5	..	..	..
90	..	..	..	..	2	2	9	16	12	12	6	4	..	..
100	..	..	..	1	4	5	10	14	12	6	2	1	1	..
110	..	..	..	..	5	5	9	13	14	3	5	1	1	..
120	..	..	..	2	7	13	13	10	5	2	1	4	..	..
130	..	..	..	..	6	9	6	2	4	3	1	1	..	..
140	..	..	..	8	11	6	9	4	3	1	3	..	..	..
150	..	..	2	5	16	11	3	4	1	..	..	..	..	..
160	..	..	3	11	8	5	4	2	2	1	1	..	..	..
170	..	..	2	9	5	3	1	..	1	1	..	..	..	..
180	..	..	1	6	5	4	2	1	..	..	..	..	..	..
190	..	..	1	2	7	3	2	..	1	..	..	..	..	..
200	..	..	4	6	4	2	..	..	..	..	..	..	..	..
210	..	..	2	2	4	3	..	..	..	..	..	..	..	..
220	..	..	4	4	4	..	..	..	..	..	..	..	..	..
230	..	..	3	4	3	..	..	..	..	..	..	..	..	..
240	..	..	1	3	1	..	..	..	..	..	..	..	..	..
250	..	4	12	9	2	..	..	..	..	..	..	..	..	..
300	..	3	18	5	1	..	1	..	..	..	..	..	..	..
400	..	..	5	2	1	..	..	..	..	..	..	..	..	..
500	..	1	2	4	1	..	.	..	..	..	..	..	..	..
600	..	11	6	2	..	..	..	..	..	..	..	..	..	..
D. N. C.	1	10	1	..	1	..	..	..	..	..	..	..	..	..
Total..........	1	29	67	85	98	72	77	80	70	42	27	16	2	666
75% ile..	..	500	224	165	140	123	105	94	91	85	86	77	..	..
Median........	..	600	290	180	156	141	123	107	106	96	99	97	..	..
25% ile........	..	D.N.C.	365	249	195	164	143	123	119	110	119	123	..	..
Quartile........	..	..	80.5	42.0	27.5	20.5	19.0	14.5	14.0	12.5	15.5	23.0	..	..



GRAPH 25.—The Substitution Test. Score.

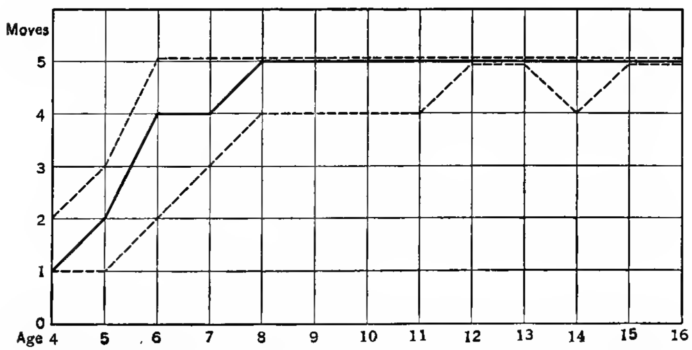
scattering. The lowest score, between 60 and 70, is made by one case at each of the ages from eleven to fifteen inclusive. Inability to do the test is shown only at ages four, five and six, with an isolated case at age eight. The graph for the medians shows a particularly constant and steady decrease from age five to age eleven or twelve, from which point onwards it remains more or less stationary. The small amount of variation at all ages, as indicated by the percentiles, is a noticeable feature of this test, and reflects the compact distribution as shown in the table.

The Adaptation Board

Score. (Table 20 and Graph 26.) The table shows a fairly good distribution, indicating increasing ability to perform the test with increasing age.

Table 20. The Adaptation Board.

Age	4	5	6	7	8	9	10	11	12	13	14	15	16	
Number Correct														
0	3	..	1	2	..	..	..	..	..	..	..	..	..	..
1	8	19	12	9	4	2	1	..	..	..	..	..	..	..
2	2	16	20	13	10	2	3	..	..	..	..	..	..	..
3	2	5	23	20	22	16	8	3	2	1	2	..	..	..
4	..	7	25	40	35	41	15	13	8	4	10	2	1	..
5	..	4	35	48	84	66	49	48	49	42	30	19	12	..
Total	15	51	116	132	155	127	76	64	59	47	42	21	13	918
75% ile	2	3	5	5	5	5	5	5	5	5	5	5	5	..
Median	1	2	4	4	5	5	5	5	5	5	5	5	5	..
25% ile	1	1	2	3	4	4	4	4	5	5	4	5	5	..
Quartile	0.5	1.0	1.5	1.0	0.5	0.5	0.5	0.5	0	0	0.5	0	0	..



Graph 26.—The Adaptation Board. Number of moves correct.

Only a very few of the four-, six- and seven-year-olds are unable to do any part of it. Cases of complete performance begin at age five and the number increases steadily in the upper ages. The curve for the medians shows a constant rise up to age eight, where it reaches the maximum score. It remains at this maximum score for all succeeding ages. The amount of variability is naturally small, since we are dealing with a small amount of possible variation.

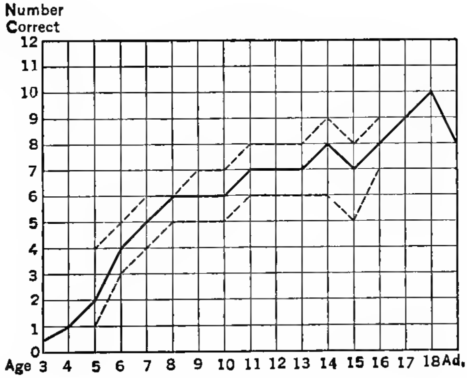
THE CUBE TEST

Score. (Table 21 and Graph 27.) The table shows a good distribution with relatively little scattering. A score of 0 is made by a few cases at ages three to six inclusive. The highest possible score of 12 is made by only one individual, a sixteen-year-old. The curve for the medians shows a constant increase up to age fourteen, from which point onwards we have a drop at age fifteen and again with adults. It is interesting to note that the adults make the same score as fourteen- and sixteen-year-old children. The amount of variation of the middle 50 per cent is fairly small and fairly constant at all ages.

These tables of distribution and graphs will form the basis for the various methods in which our data have been used for the purpose of constructing the different types of scales to be discussed in the

TABLE 21. THE CUBE TEST.

Age	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Ad.	
Number Correct																		
0	4	1	0	1	..	..	..	..	..	..				..	..	..	..	..
1	5	9	13	4	..	2	..	..	1	..	..	..	..	..	..	..	..	..
2	..	2	12	2	3	1	1	1	2	1	2	1	..	..	..	..	..	..
3	..	2	11	9	7	4	6	3	1	3	1	..	4	..	..	..	..	..
4	..	..	9	15	7	3	7	7	8	7	3	..	1	..	..	..	1	..
5	..	..	7	7	18	15	15	14	10	10	8	9	12	2	..	..	2	..
6	..	..	3	6	10	19	16	23	17	19	16	15	10	4	..	..	4	..
7	..	..	..	4	4	7	14	23	23	27	19	17	13	8	1	2	6	..
8	..	..	..	1	6	3	8	13	18	19	14	25	8	7	1	..	10	..
9	..	..	..	..	..	3	2	7	5	13	8	15	6	7	3	..	5	..
10	..	..	..	..	1	..	1	4	8	2	5	7	7	3	2	2	8	..
11	..	..	..	..	..	..	..	..	1	..	1	5	1	1	..	1	3	..
12	..	..	..	..	..	..	..	..	..	..	..	..	..	1	..	..	..	..
Total	9	14	61	49	56	57	70	95	94	101	77	94	62	33	7	5	39	923
75% ile.........	..	..	4	5	6	6	7	7	8	8	8	9	8	9	..	..	9	..
Median.........	1	1	2	4	5	6	6	6	7	7	7	8	7	8	9	10	8	..
25% ile.........	..	..	1	3	4	5	5	5	6	6	6	6	5	7	..	..	7	..
Quartile........	..	..	1.5	1.0	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.5	1.5	1.0	..	..	1.0	..



GRAPH 27.—The Cube Test. Number of lines correct.

chapters that follow. We have thought it wise to present the data in this shape, first of all, so that a rough idea of the reliability of the tests could be obtained and also so that later on other results might be added to those given in our tables of distribution. The addition of more cases would serve to increase the reliability of the norms.

CHAPTER V

THE YEAR SCALE

THE first type of scale that we have tentatively constructed on the basis of the results collected has been of the type made familiar by the Binet Scale. A year scale is a scale in which the tests are grouped according to years, with the presupposition that the average child of a particular age will pass all the tests of the year scale at the age in question and all below that year and none above that year. This is, of course, the ideal; and what we actually find is that a particular child passes tests scattered over several years. An addition of these tests leads to the computation of a mental age.

In the chapter on standardization we have discussed the various methods in common use in order to determine the placing of a test at a particular year. In general, the choice lies between the adoption of the 75 per cent standard or of a standard which fluctuates between 60 and 90 per cent, according to the type of curve exhibited by the results of the test. The advantages of these different methods have been discussed sufficiently. In the year scale here presented the 75 per cent standard is adopted. This method is chosen owing to the nature of the data with which we are dealing. Most of our tests are not amenable to the "all or none" credit method that has been customary up to the present time in year scales. We cannot say that a child passes or does not pass a particular test if we adhere to the method of evaluating the performance which we have adopted. The pass or fail method would have limited each test to use at one particular age only. In this case we should have had to set a definite time limit to each test and credited with a pass all who completed the test within that time limit. The age where the curve showed the most decided rise above the 60 per cent point would have been the age at which to place the test. This method of procedure would have greatly diminished the scope of each test and would have left us with relatively few tests, hardly adequate to form a scale.

We have, therefore, adopted the 75 per cent standard and taken the time or score made by the lowest of the upper 75 per cent at each age as being the time or score which a child must make in order to be credited with a pass at any particular age. This allows us to use most of our tests for a great many ages. The actual procedure has been to take the 25 percentile point at each age as being the lowest time or score or number of moves or errors made by the upper 75 per cent and use this as the limiting point for crediting a pass at a specific age. The other limiting point is set by the 25 percentile of the age below. For example, the 25 percentile at age seven is 31 and the 25 percentile at age eight is 18. This means that at age seven the upper 75 per cent make scores better than 31, therefore 30 is taken as the one limit and this limit extends down to the limit for age eight, which is 17. All those who make scores between 30 and 18 are given a seven-year-old credit, since 75 per cent of the seven-year-olds make scores better than 31; if, however, they make scores better than 18 they are given an eight-year-old credit because they fall within the eight-year-old group. This method seemed the only one possible in dealing with data such as we have in our tests. On the basis of this scheme Table 22 has been constructed. The limiting points for time or score or number of moves or errors are given for each test for each age. The table is to be read as follows: In the Mare and Foal Test any individual making a time record lying between 160 and 92 seconds inclusive is to be credited with a pass at age five, and any individual making a time record between 91 and 77 seconds inclusive is to be credited with a pass at ages five and six; any individual completing the test within from 76 to 59 seconds is to be credited with a pass for all ages from five to seven, and so on for the other ages until we get to time records of 32 or less, for which record an individual is to be credited with a pass for ages five to eleven inclusive and also for age thirteen. The second line giving the number of errors is to be interpreted as follows: for more errors than 11 credit at no age is given, for errors extending from

Table 22. Limiting Points for Age Credit in the Year Scale.

Test	Age	4	5	6	7	8	9	10	11	12	13	14	15
1. Mare and Foal.......	Time Errors	....	160–92 11–5	91–77 4–3	76–59	58–49	48–45	44–40	39–33	 2–	32–	....	....
2. Seguin..............	Time (best of three)	......	43–30	29–26	25–22	21–20	19–16	......	15–13	......	......	12–	......
3. Five Figure Board....	Time Errors	....	....	300–225 31–19	224–146 18–12	145–132 11–8	131–107 7–5	106–91	90–85	84–63	62– 4–	....	....
4. Two Figure Board....	Time Moves	....	....	300–284 45–40	283–118 39–21	115–94 20–18	93–64 17–15	....	63–59 14–12	58–43	42–	 11–	....
5. Casuist Form Board..	Time Errors	....	....	300–274	273–165 30–15	164–128 14–12	....	127–99 11–8	98–84 7–6	....	 5–	83–	....
6. Triangle Test.........	Time Errors	....	....	300–240 36–31	239–123 30–18	122–89 17–14	....	88–86 13–11	....	85–60 10–9	59– 8–	....	....
7. Diagonal Test........	Time Errors	....	....	300–174 40–22	173–141 21–17	....	140–83 16–11	82–49 10–6	....	....	48–39 5–	38–	....
8. Healy Puzzle "A"....	Time Moves	....	....	....	....	300–258 100–51	257–145 50–39	144–96 38–32	95–78 31–27	77–66 26–21	65– 20–	....	....
9. Manikin Test........	Score	1–2	......	3	4	......	5	......	......	......	......	......	......
10. Feature Profile.......	Time	......	......	......	......	......	......	300–237	236–212	......	......	211–190	189–
11. Ship Test............	Score	......	1–6	7–8	9–14	15	16	17	18	......	19–20	......	......
12. Picture Completion...	Score	......	5–42	43–98	99–177	178–270	271–326	327–365	366–395	396–415	416–445	446–460	461–
13. Substitution..........	Score	......	600–385	384–249	248–195	194–164	163–143	142–123	122–119	118–110	109–	......	......
14. Adaptation...........	Moves	1	2	3	......	4	......	......	5	......	......	......	......
15. Cube Test...........	Number Correct	1	2	3	4	5	......	6	......	......	......	7	8–

11 to 5 a pass for age five is allowed, for 4 or 3 errors a pass at ages five and six is given, for 2 or less errors a pass at ages five, six and twelve is allowed.

We have by this method established a system of age credits or passes whereby a specific record gives credit for a certain number of ages, which have been determined by the limits as set by the points at which 75 per cent pass at each age. It is to be noted that by this method failure to pass a test cannot be credited at all, although we may know that failure to pass the test is the median or even 25 percentile performance of the group. It would, however, be impossible to give any specific age credit for a failure, since we do not know whether the failure in question is a two-, three- or four-year-old type of failure.

Table 22 is, therefore, the table to which the worker must constantly refer for evaluation of the results after giving the tests. It will be seen at a glance that the number of tests at each age, or rather that the number of age credits for tests, is different for different ages. In two ways, therefore, our year scale differs radically from the year scale of the Binet type. In the first place, our tests are not given as tests specially adapted to one or at most two or three years, as is the case with the Binet tests. In the second place, the number of tests at each age varies. The Binet and modifications of the Binet have generally adhered to a constant number of tests at each age. We have disregarded this altogether, allowing the tests themselves to determine the number of different years for which they are adapted and thus setting the number that may fall to each age. The following number of tests or performances to be allotted age credit have resulted for each age:

At age 4— 3	At age 10—15
" " 5— 8	" " 11—14
" " 6—18	" " 12— 9
" " 7—17	" " 13—14
" " 8—17	" " 14— 6
" " 9—14	" " 15— 3

This method, adopted for evaluating the results and for arriving at a mental age, is the one proposed by Terman and Childs [1] in their first proposal for a revision of the Binet Scale, namely, that of giving a different test value or index to each of a group of tests according to the number of tests in the group. The Binet Scale allows 1/5 of a year credit for each additional test passed beyond the basal year, because there are 5 tests in each age group. The logical step taken by Terman and Childs was to allow that fraction of a year as credit which corresponded to the number of tests within a group, since some of their groups of tests contained more than the original 5 tests of Binet. If

[1] Terman, L. M., and Childs, H. G.: "A Tentative Revision and Extension of the Binet-Simon Measuring Scale of Intelligence," *Journal of Educational Psychology,* Vol. iii (1912).

there were 6 tests in a group, 1/6 of a year was allowed for each test, and so on. We have adopted this same method in allotting credit to the differing numbers of tests in our year groups. The values obtained by dividing the year's credit into fractions, according to the number of tests in each age group, we have called "test values." These "test values" are constant as long as we use all the tests in the scale. If, however, we are obliged, for any reason, to omit any of the tests, we may nevertheless arrive at a mental age by computing new test values for each year corresponding to the new grouping of tests that has resulted from the omission of some of the tests.

The procedure in computing mental age is the same as with the Binet Scale. After the tests have been evaluated, the basal age is determined, that is, the age at which all the tests are passed. To this basal age are added the extra number of years and fractions thereof that have been obtained by additional tests passed above the basal year. In actual procedure we multiply the number of tests passed at each year by the test value for the year, take the sum of these test values and add this to the basal year.

This procedure will be made clear from an explanation of the test blank used and a description of a sample case. A copy of the test blank is shown on Figure 11. The fifteen tests are numbered and designated by name at the left of the blank. Next follows the record of the test. The

Form 889

Date ________ Examined by ________

Name ________ Sex ________

Date of Birth ________ Nationality ________

PSYCHOLOGICAL CLINIC

THE OHIO STATE UNIVERSITY

School ________ Grade ________

RECORD BLANK FOR PINTNER-PATERSON PERFORMANCE SCALE

Chronological Age (Age last birthday) ________ Diagnosis ________

Median Mental Age ________ Points ________ Point Age ________

Median Percentile ________ Mental Age (Year Scale) ________

Test. Name of Test.	Record	Median Mental Age	Percentile	Points	YEAR SCALE												
					Age	4	5	6	7	8	9	10	11	12	13	14	15
					Test Value	.33	.12	.05	.06	.06	.07	.07	.07	.09	.07	.14	.33
1. Mare and Foal.	Time ____						5	6	7	8	9	10	11		13		
	Errors ____						5	6						12			
2. Seguin Form Board. Time 1 ____ 2 ____ 3 ____	Time of Shortest Trial ____						5	6	7	8	9		11			14	
3. Five Figure Board.	Time ____							6	7	8	9	10	11	12	13		
	Errors ____							6	7	8	9				13		
4. Two Figure Board.	Time ____							6	7	8	9		11	12	13		
	Moves ____							6	7	8	9		11			14	
5. Casuist Form Board.	Time ____							6	7	8		10	11			14	
	Errors ____								7	8		10	11		13		
6. Triangle Test.	Time ____							6	7	8		10		12	13		
	Errors ____							6	7	8		10		12	13		
7. Diagonal Test.	Time ____							6	7		9	10			13	14	
	Errors ____							6	7		9	10			13		
8. Healy Puzzle "A."	Time ____									8	9	10	11	12	13		
	Moves ____									8	9	10	11	12	13		
9. Manikin Test. Time ____	Score ____					4		6	7		9						
10. Feature Profile.	Time ____											10	11			14	15
11. Ship Test. Time ____	Score ____						5	6	7	8	9	10	11		13		
12. Picture Completion. Time ____	Score ____						5	6	7	8	9	10	11	12	13	14	15
13. Substitution Test. Time ____ Errors ____	Score ____						5	6	7	8	9	10	11	12	13		
14. Adaptation Board. Moves 1 ____ 2 ____ 3 ____ 4 ____ 5 ____	Score ____					4	5	6		8			11				
15. Cube Test. A ____ X ____ Y ____ B ____ C ____ D ____ E ____ F ____ G ____ H ____ I ____ J ____	Number Correct ____					4	5	6	7	8		10				14	15
SCORING RECORD FOR TEST 12 PICTURE COMPLETION TEST		Median Mental Age ____	Median Percentile ____	Total Points ____	Number of Tests												
					Test Values												

Name of Piece	log	broken window	flying bird	dog	cat	basket	foot ball	hat	chicken	Total Score
Record										
Score Value										

FIGURE 11.

records actually used for diagnostic purposes are all found in the vertical column directly under the heading "Record." In many tests, however, space is provided for time records, etc., that are not directly used for diagnosis, and in others space is provided for scoring the tests, as in Tests 14 and 15. The next three columns, headed "Median Mental Age," "Percentile," and "Points," are for these three methods of evaluating the tests and they will be discussed in the succeeding chapters. The part of the blank to the right under the heading "Year Scale" is what concerns us here. The first horizontal line shows the ages from four to fifteen. The next line gives the test values computed as we have described above. The other figures below this show at what ages passes are allowed for each test. For example, taking the first horizontal line showing the passes allowed for the Mare and Foal Test (Time), we note that age credit is allowed at ages five, six, seven, eight, nine, ten, eleven and thirteen. And so on for the other tests. Again, reading vertically down the columns, we note that there are three tests at age four, i.e., Tests 9, 14, and 15, and each has a test value of .33. To obtain a mental age of four on the scale, a subject must pass all of these four-year-old tests. At age five there are 8 tests which give five-year-old credit, each having a test value of .12, and so on with the other ages. The horizontal line at the bottom of these age credits, called "Number of Tests," allows a space for noting the number of tests passed at each age. Below this, marked "Test Values," is a line for the total amount of credit arrived at for each age group of tests.

Figure 12 shows a record blank filled out for year scale purposes. The actual calculation of a mental age can be described more accurately by reference to this. The boy completes the Mare and Foal Test in 62 seconds. By reference to Table 22 we find that this gives him credit up to age seven. We, therefore, mark this on the record sheet by putting circles (or any other mark) around all the ages for which credit is allowed on this test up to and including age seven, i.e., ages five, six and seven. Two errors were made and this is equivalent to a twelve-year performance. We, therefore, mark all ages for which credit is given up to and including age twelve. In this case there are only three ages, namely five, six and twelve. From this sample the procedure with the other tests will be clear. It is necessary to mark all ages below the age at which credit is gained, so that when we come to checking up the vertical columns we can see at a glance whether all the tests at a specific age have been passed or not. After all the tests have been evaluated, we proceed to the checking up of the year scale. At ages four, five, six and seven all the numbers on the vertical columns are inclosed by circles, which means that all the tests have been passed. We, therefore, make a check mark below these ages. Age seven is the last age at which all tests have been passed; it is the so-called basal age and we note this below by writing the digit 7. At

THE OHIO STATE UNIVERSITY

Date ______ Examined by ______

Name *Harold H.* Sex *Boy*

Date of Birth *18/4/1909.* Nationality *American*

School *7* Grade *2 B.*

Chronological Age *1* (Age last birthday) Diagnosis ______

Median Mental Age ______ Points ______ Point Age ______

Median Percentile ______ Mental Age (Year Scale) *11.1*

Test. Name of Test.	Record	Median Mental Age	Percentile	Points	Age 4	5	6	7	8	9	10	11	12	13	14	15
					Test Value .33	.12	.05	.06	.06	.07	.07	.07	.09	.07	.14	.33
1. Mare and Foal.	Time *62*					5	6	7	8	9	10	11		13		
	Errors *2*					5	6						12			
2. Seguin Form Board. Time 1 *25* 2 *17* 3 *19*	Time of Shortest Trial *17*					5	6	7	8	9		11			14	
3. Five Figure Board.	Time *52*						6	7	8	9	10	11	12	13		
	Errors *0*						6	7	8	9				13		
4. Two Figure Board.	Time *74*	*9*					6	7	8	9		11	12	13		
	Moves *13*	*11*					6	7	8	9		11			14	
5. Casuist Form Board.	Time *65*						6	7	8		10	11			14	
	Errors *2*							7	8		10	11		13		
6. Triangle Test.	Time *18*						6	7	8		10		12	13		
	Errors *1*						6	7	8		10		12	13		
7. Diagonal Test.	Time *18*						6	7		9	10			13	14	
	Errors *0*						6	7		9	10			13		
8. Healy Puzzle "A."	Time *195*								8	9	10	11	12	13		
	Moves *27*								8	9	10	11	12	13		
9. Manikin Test. Time *30*	Score *5*				4		6	7		9						
10. Feature Profile.	Time *248*										10	11			14	1
11. Ship Test. Time *146*	Score *20*					5	6	7	8	9	10	11		13		
12. Picture Completion. Time *206*	Score *287*					5	6	7	8	9	10	11	12	13	14	15
13. Substitution Test. Time *168* Errors *2*	Score *174*					5	6	7	8	9	10	11	12	13		
14. Adaptation Board. Moves 1 *+* 2 *+* 3 *+* 4 *+* 5 *+*	Score *5*				4	5	6		8			11				
15. Cube Test. A *✓* X *✓* Y *✓* B *✓* C *—* D *✓* E *—* F *—* G *—* H *—* I *—* J *—*	Number Correct *5*				4	5	6	7	8		10				14	15
		Median Mental Age	Median Percentile	Total Points	Number of Tests *✓*	*✓*	*✓*	*✓*	*16*	*12*	*10*	*7*	*4*	*8*	*2*	
					Test Values			*7*	*.96*	*.84*	*.70*	*.49*	*.36*	*.56*	*.28*	*=11.1*

SCORING RECORD FOR TEST 12

PICTURE COMPLETION TEST

Name of Piece	log	broken window	flying bird	dog	cat	basket	foot ball	hat	chicken	Total Score
Record	*✓*	*c. wind*	*cherries*	*baby*	*✓*	*✓*	*pumpkin*	*purse*	*✓*	
	53	*22*	*3*	*2*	*81*	*55*	*1*	*3*	*58*	*287*

age eight 16 tests have been passed and, since this is not a complete age group, we write this number in the column for age eight. Similarly at age nine we count 12 tests, at age ten we have 10 tests, at age eleven we have 7 tests, at age twelve 4, at age thirteen 8, at age fourteen 2. The number of tests at each age is now multiplied by the test value at the top of the sheet and the result is noted on the horizontal line marked "Test Values." Thus, the 16 tests at age eight are multiplied by the test value .06, which gives .96. The 12 tests at age nine are multiplied by .09, which gives .84, and so on for the other ages. The sum of these test values equals 4.19. This is added to the basal age of seven and gives as a final result the mental age of 11.19.

From this sample the computation of mental age on the year scale will be obvious. The procedure is somewhat more complex than is the case with the ordinary year scales. This complexity is due to the fact that we have made use of our tests for many ages according to the quality of the performance, and have thereby abandoned the "plus or minus" method of utilizing a test, which limits the test to use at one or two ages only.

The question of diagnosis based on the mental age arrived at by the year scale is, of course, unanswerable at this time, and must wait for an adequate answer until sufficient cases have been tested with the scale.

CHAPTER VI

THE MEDIAN MENTAL AGE

THE method used in the Binet Scale for the determination of mental age has been so widely used and has become so much a matter of habit in clinical psychology that very little has been done, with the exception of the Point Scale Method, in the way of discussion as to other possible methods of arriving at a mental diagnosis. It has occurred to us that the median mental age of a group of tests might very well serve as a reliable value for the estimation of an individual's mentality. We offer this suggestion as a method that must be worked out and tested. We are, unfortunately, unable to use our data to test adequately the reliability of this method, but we hope to do so in the future.

The method is briefly as follows: Given a group of tests which have been adequately standardized and for which the median performance at each age is available, then the measure of an individual's intelligence is the median of all the mental ages which he approximates in all the tests.

To make our data available for this method, we have constructed Table 23. This has been arranged from the tables of medians given for each

TABLE 23. MEDIAN INTERVALS FOR SCORING TESTS. BY THE MEDIAN MENTAL AGE METHOD.

Test	Age	4	5	6	7	8	9	9.5	10	11	11.5	12	12.5	13	13.5	14	15	16
1. Mare and Foal......	Time Errors	.. ..	150–89 10–5	88–67 4	66–50 3	49–45	44–39	.. 2	38–35	34–33	.. ..	32–31	.. ..	30–29	.. 1	28	27 ..	26– 0
2. Seguin..............	Time	..	50–32	31–25	24–22	21–19	18–17	..	16	15	..	14	..	13–12	..	11	10–	..
3. Five Figure Board...	Time Errors	.. ..	....	300–158 30–11	157–107 10–7	106–88 6	87–74 5	.. ..	73–67 4	66–61	.. ..	60–53	.. 3	52–45	.. ..	44–	.. 2–	.. ..
4. Two Figure Board...	Time Moves	.. ..	300–188 45–27	187–146 26–23	145–89 22–18	88–56 17–14	55–49 13	.. ..	48–43 12	42–37	.. ..	36–32	.. 11	31–26	.. ..	25–	.. ..	.. 10–
5. Casuist Form Board..	Time Errors	.. ..	....	300–227 32–21	226–130 20–11	129–100 10–8	99–86 7	.. ..	85–83 6	82–67 5	.. ..	66–64 4	.. ..	63–61	.. 3	60–54	53– 2–	.. ..
6. Triangle Test........	Time Errors	.. ...	300–192 35–20	191–93 19–12	92–71 11–10	70–62 9	61–57	.. 8	56–52	51–49	.. 7	48–43	.. ..	42–30 6	.. ..	29– 5	.. 4–	.. ..
7. Diagonal Test.......	Time Errors	.. ..	300–163 35–19	162–113 18–12	112–76 11–10	75–62 9–8	61–46 7	.. ..	45–41 6	40–36 5	.. ..	35–29 4	.. ..	28–24 3	.. ..	23– 2–	.. ..	.. ..
8. Healy Puzzle "A"...	Time Moves	.. ..	....	....	300–124 100–43	123–102 42–32	101–78 31–26	.. ..	77–62 25–22	61–50 21–19	.. ..	49–42 18–17	.. ..	41–36 16	.. ..	35–32 15	31– 14–	.. ..
9. Manikin Test.......	Score	1	2	3	4	5	.....	..	.....	.....	..	.....	..	.....	..	.....	..	..
10. Feature Profile Test..	Time	..	.....	.....	.....	.....	.....	..	300–213	212–178	..	177–161	..	160–141	..	140–128	127–	..
11. Ship Test...........	Score	..	1–10	11–15	16	17	.....	18	.....	19	..	20	..	.....	..	.....	..	..
12. Picture Completion...	Score	..	30–121	122–196	197–284	285–368	369–421	..	422–444	445–474	..	475–499	..	500–510	..	511–	..	..
13. Substitution.........	Score	..	600–445	444–235	234–169	168–150	149–132	..	131–115	114–107	..	106–100	..	99–96	..	95–	..	..
14. Adaptation Board....	Moves	1	2	3	4	5	.....	..	.....	.....	..	.....	..	.....	..	.....	..	..
15. Cube Test..........	Number Correct	1	2	3–4	5	.....	6	..	.....	.....	..	7	..	.....	..	8	9–	..

test in Chapter IV. Since we are dealing, in most cases, with large values, the median value for each age is obtained by finding the middle points between each age and by using the interval between these middle points as the median interval. These median intervals are shown in the table. For example, in the Mare and Foal Test, Time, the median time for the six-year-olds is 71 seconds (see Table 1, p. 100). A point midway between this median and the median for the five-year-olds, which is 107 seconds, is about 88 seconds and, therefore, 88 is the upper limit of the median interval for age six. Similarly the median for the seven-year-olds is 62, and a point midway between 62 and 71 is about 67, and, therefore, 67 is the lower limit for the six-year-olds and 66 the upper limit for the seven-year-olds.

In Table 23 the top line gives the mental age. The next line of the table is to be read as follows: In the Mare and Foal Test all time records between 150 and 89 are to be given five-year-old credit, all records between 88 and 67 are to be given six-year-old credit, and so on to age sixteen, where all records below 26 are to be given sixteen-year-old credit.

A difficulty of this method appears in the second line. Here the median number of errors made by eight-, nine-, ten- and eleven-year-olds is the same, namely 2. The question then is, If a child makes a score of 2, which is the median for eight-, nine-, ten- and eleven-year-olds, what mental age are we to credit him with? The only answer to this question that we are able to give at present is to credit the child with the median of the mental ages which have the same median scores. Thus, in the Mare and Foal Test, Errors, we would give a mental age of 9.5 for a performance with only two errors, and similarly a mental age of 13.5 for a performance with one error. This same difficulty is encountered in other tests, as, for example, in the Five Figure Board, Errors, where a score of 3 is the median for ages eleven to fourteen, and where we would give by this method a mental age of 12.5. This same thing also occurs in the Two Figure Board, Moves; in the Ship Test, and rather markedly in the Cube Test. In short, this difficulty will tend to occur in all tests where the method of scoring is not fine enough to allow for slight differences in the performance of a test and, therefore, does not discriminate between the medians of two or three contiguous years.

Whether this difficulty will turn out to be a real one in the actual use of the method is yet to be determined. It may be that by some slight modification of the scoring of a test we shall be able to overcome it in part. It may also be that by the use of many tests, and the use of the median of the mental ages on all the tests, no real difficulty will be present.

The advantage of this method appears to us to lie in its direct comparison of a performance with the median performance of the different ages.

Again, if for any reason a test must be omitted, no change is required in computing mental age, for we only have to take the median of the tests used. New tests may be added as fast as they are standardized and old ones discarded if they are found unsuitable. Furthermore, an inspection of the array of median mental ages will give a kind of mental profile of the individual. We are able to see at a glance what mental age he approximates in the so-called different mental processes being measured by the tests.

Unfortunately, as we have stated before, we were not able to test all the individuals on all the tests, so that we are unable to find the median mental ages on this series of tests for each individual and see what kind of a distribution would result for each chronological age. We have done this, however, for a group of tests for one age. There were 77 ten-year-old children who had all been tested on nine identical tests. Making use of both time and errors on some tests, we have 16 separate values for each child. These values are turned into the equivalent median mental age by the use of Table 23. These median mental ages for each test for each of the 77 children are shown in Table 24. The first sixteen columns of the table show these separate mental ages for each of the tests. The last column shows the median mental age of the child. The table is to be read as follows: Case No. 1 makes a performance on the Mare and Foal Test which is equal to a median twelve-year-old performance, and the number of errors on this test is

Table 24. Mental Ages on Tests. Ten Year Old Children.

Case Number	Mare & Foal		Five Figure		Two Figure		Tri-angle		Diag-onal		Ship	Subst.	Casuist		Healy "A"		Med. M. A.
	Time	Errors	Time	Errors	Time	Moves	Time	Errors	Time	Errors	Score	Score	Time	Errors	Time	Errors	
1	12	13.5	6	6	9	7	14	11.5	5	5	12	14	6	6	7	7	7
2	10	6	11	8	8	7	13	9.5	13	14	12	10	8	10	15	15	10
3	16	16	14	15	7	6	7	6	14	14	8	10	13	9	7	7	9.5
4	6	5	7	6	-5	-5	-5	-5	10	11	5	6	8	7	7	6	6
5	7	16	6	7	13	16	12	14	6	6	9.5	9	9	8	-7	-7	6.5
6	10	16	14	15	7	6	12	13	8	7	8	10	15	13 6	10	10	10
7	7	6	10	10	11	12.5	5	5	6	-5	6	9	8	7	10	11	7.5
8	9	7	11	9	6	7	7	6	14	14	9.6	8	11	7	7	'7	6
9	7	9.5	14	15	13	16	10	9.5	14	14	12	11	15	15	15	15	14
10	8	5	13	12.5	7	6	14	15	13	8	9.5	8	11	7	15	15	10 5
11	16	7	6	7	7	6	7	6	14	13	11	8	15	12	-7	-7	7 5
12	8	7	14	15	8	7	6	14	7	9	12	10	6	7	10	13	8 5
13	9	7	7	8	5	5	6	6	9	11	9.5	7	9	7	12	11	7.5
14	6	6	-6	-6	8	7	13	7	8	6	6	7	7	-6	-7	-7	7
15	11	9.5	14	15	10	16	7	9.5	14	14	12	10	11	15	9	9	11
16	15	13.5	14	12.5	14	12.5	14	15	7	6	12	9	15	13 5	9	8	13
17	16	16	13	12.5	7	5	14	15	10	11	7	10	7	7	7	6	10
18	5	9.5	6	12 5	10	8	-5	-5	-5	-5	7	7	6	6	7	7	6.5
19	11	16	14	12.5	12	8	14	15	9	8	12	11	15	15	12	12	12
20	13	13.5	13	10	-5	-5	13	14	11	12	8	13	7	12	7	9	11.5
21	16	16	11	7	14	12.5	10	9.5	9	7	12	11	11	7	7	7	10 5
22	13	16	13	15	13	16	14	15	10	12	12	11	15	13.5	15	15	13:5
23	10	6	8	7	12	12.6	12	7	12	8	6	9	9	7	15	15	9
24	11	16	14	15	14	12.5	6	6	8	6	8	14	11	9	10	10	10
25	10	7	13	15	7	5	6	5	12	9	9.5	11	14	11	14	11	10
26	14	13 5	13	15	11	16	7	9.5	14	14	12	11	15	15	9	14	14
27	9	13.5	7	7	12	12.5	14	15	14	14	11	14	7	6	8	7	11.5
28	10	16	6	10	6	7	7	6	13	12	6	6	11	12	9	9	9
29	12	13.5	14	15	7	6	13	11.5	12	8	7	13	13	13.5	7	7	12
30	11	9 5	10	7	11	12 5	13	13	6	6	6	9	13	13.5	9	10	10
31	9	13.5	14	15	8	8	13	15	7	6	9 5	10	15	15	11	12	11 5
32	7	5	13	15	5	6	13	15	14	14	9.5	9	15	15	10	12	12.5
33	9	7	7	6	7	7	14	15	11	9	5	6	7	6	9	9	7 5
34	15	16	14	15	13	16	13	15	9	10	9.5	14	12	11	15	15	14
35	12	16	8	12.5	14	16	14	15	11	11	7	11	15	15	15	15	14
36	14	0	14	15	14	16	13	8	14	14	12	13	7	7	11	11	13
37	12	9.5	9	7	6	5	14	15	6	5	8	7	11	9	12	15	9.25
38	15	13 5	13	12.5	7	5	6	5	11	10	7	13	15	11	7	7	10 5
39	16	13.5	9	10	14	16	12	14	14	14	8	10	15	15	-7	-7	13 75
40	9	7	14	15	9	6	-5	-5	12	12	7	9	14	8	15	15	9
41	15	9 5	12	15	7	5	6	7	12	12	11	9	15	13.5	7	7	10
42	9	0.5	7	7	10	12.5	10	8	5	5	12	12	6	12	8	8	6 5
43	12	9.5	14	12.5	13	9	14	15	14	14	12	11	15	15	15	15	14
44	10	13.5	6	9	5	5	14	15	12	12	9.5	11	13	13 5	7	7	10 5
45	12	13.5	7	6	14	16	7	6	5	5	11	14	11	11	10	9	10 5
46	7	16	9	9	7	6	7	9 5	9	8	12	10	11	13.5	15	15	9 5
47	6	13.5	12	12.5	12	9	7	9.5	7	6	9.6	13	12	15	15	15	12
48	16	13.5	9	15	14	16	7	7	14	14	12	11	15	15	15	15	14
49	9	9.5	8	7	7	6	14	15	9	7	12	14	12	15	13	13	10
50	13	13.5	11	10	6	5	11	6	7	6	11	10	15	13.5	13	14	11
51	9	13.5	7	7	7	7	13	14	9	6	8	11	11	9	7	-7	6.5
52	7	13.5	7	7	7	7	7	6	14	14	11	10	15	13.5	10	11	10
53	12	9.5	14	15	13	12.5	12	9.5	11	9	7	10	14	13.5	-7	-7	11 5
54	11	16	12	15	7	6	14	15	6	6	12	10	8	7	10	10	10
55	8	16	6	6	9	9	8	13	8	9	11	14	7	9	-7	-7	8 5
56	8	16	12	15	12	9	7	15	14	14	11	13	14	15	12	15	13 5
57	14	16	12	10	-5	-5	14	15	7	7	6	6	11	9	15	15	10.5
58	16	13 5	12	9	13	16	7	6	5	5	12	12	6	6	13	13	12
59	15	16	11	10	14	16	6	5	14	14	9.5	10	12	12	15	15	12 6
60	11	13.5	6	9	11	12.5	7	7	.5	6	9.5	10	13	13.5	15	15	10 5
61	12	16	12	10	14	16	6	6	7	6	12	12	11	7	12	12	12
62	9	6	6	10	13	10	6	8	7	8	7	13	-6	-6	7	8	8
63	13	16	6	15	8	8	10	13	13	14	12	9	11	16	15	15	13
64	7	7	7	7	11	10	12	6	-5	-5	7	10	11	15	15	15	8 5
65	13	9.6	6	7	10	9	7	6	14	14	12	11	8	7	7	7	8 5
66	10	16	10	12.5	10	8	6	6	13	14	12	9	7	11	15	15	10.5
67	10	9.6	9	10	13	12.6	6	6	6	5	11	9	15	15	15	15	10
68	9	16	7	10	12	12.5	10	15	10	11	12	9	15	15	8	9	10 6
69	10	16	8	9	9	6	12	15	14	14	12	9	11	8	15	15	11 5
70	10	7	9	7	8	6	13	13	12	10	12	9	10	7	15	14	10
71	9	13.5	13	15	13	12.5	7	8	11	10	9 5	14	14	11	11	12	11 5
72	11	9.5	9	6	14	16	14	15	12	14	12	11	11	11	-6	-6	11
73	10	7	8	6	13	12.5	7	5	14	13	6	9	6	16	8	15	8.5
74	12	9.5	8	7	8	7	10	9.5	14	14	12	10	8	6	9	12	0.5
75	12	9.6	10	12.5	11	9	6	6	9	10	6	6	12	11	7	7	9 75
76	10	6	12	12.6	14	16	10	11.5	14	14	12	11	7	7	11	11	11 25
77	6	-5	7	6	5	-5	6	5	9	9	5	7	8	9	11	13	6.5

equal to the median of a thirteen-and-a-half-year-old performance. On the Five Figure Board, Time and Errors, he receives a mental age of six in both cases. And so on with the other tests. The median mental age of all these mental ages is seven, which is shown in the last column. The 16 different mental ages for each child give an indication of the amount of variation in his performance. Child No. 1, for example, makes a rather poor performance on the Diagonal Test, since his mental age here is five, whereas his median mental age is seven; on the other hand, his best performances are on the Substitution and the speed of his performance on the Triangle Tests, his mental age in both cases being fourteen. This child varies in these tests from a mental age of five to one of fourteen. What this amount of variation means will become more obvious if this method of median mental ages is used more generally.

The distribution of these 77 ten-year-olds according to their median mental ages is as follows:

Mental Age	6	7	8	9	10	11	12	13	14	Total
Number	3	6	10	8	21	10	7	6	6	77
Percentage	3.9	7.8	13.0	10.4	27.2	13.0	9.1	7.8	7.8	100.0

Of the total number of ten-year-olds 27.2 per cent make a mental age of ten. Fifty and one-half per cent make a mental age of either nine, ten or eleven. These may well be considered the normal cases. Those above the middle 50 per cent are probably bright and these constitute 24.7 per cent of the cases in this group. Those below the middle 50 per cent, making mental ages of eight and below, are probably backward and these make up 24.7 per cent of the cases.

This sample gives promise of a reliable distribution with a sufficient number of tests and a sufficient number of cases at each age. If the sample distribution of our ten-year-olds should be the ordinary distribution for this age, then we might say that the normal or middle 50 per cent of the children can be expected to test at age or one year above or below. Similar definitions for backward, feeble-minded, bright and very bright children could be arrived at according to the distribution at each age.

To sum up: The median mental age method recommends itself as a quick and simple method of arriving at a mental age. Its reliability will have to be more adequately established by more data. It allows the addition or subtraction of tests without dislocating the whole scale. We believe that this method will prove itself of decided value in the future.

CHAPTER VII

THE POINT SCALE

THE first practical application of allotting credit in points for various kinds of performances on a test was made by Yerkes, Bridges and Hardwick.[1] This method is new in its application to intelligence scales. It is, of course, the world-old device of teachers and pedagogues in marking their pupils, whether on the result of an oral recitation or of a written examination. The teacher or examiner himself determines how many marks or points shall be given for each question or test, and decides, either with great accuracy or with little regard to accuracy, how many points shall be given to each type of answer or performance. This is practically what has been done in the Yerkes-Bridges Point Scale. Each test has been divided as conveniently as possible into parts and one or two points credit have been allowed to each part, very much in the same way as a teacher will allow one mark for each question answered correctly or for each example in arithmetic solved correctly, without any

[1] Yerkes, R. M., Bridges, J. W., and Hardwick, R. S.: A Point Scale for Measuring Mental Ability, Warwick and York (1915).

regard to the differing degrees of difficulty of the various questions or examples.

The allotment of points in the Yerkes-Bridges Scale has been purely arbitrary, as the authors themselves admit, and if any principle can be said to underlie the allotment of points, the only possibility may be the convenience with which a test may be divided into parts. The table of the different mental processes supposed to be measured by the tests [2] and the credits allowed to each mental process might have led some to infer that the authors were seeking an allotment of credits weighted according to their estimation of the value of these mental processes in the total complex called "general intelligence." This, however, is not the case, since the table is merely the statement of the actual number of points allotted to the different mental processes resulting from the group of tests that happened to have been chosen. We do not mean to imply by this that it is unfortunate that the authors did not adopt some such principle as is suggested by the table mentioned. On the contrary, we think it fortunate that they have avoided this pitfall, since there would have resulted much fruitless discussion as to what mental processes are involved in the complex known as "general intelligence" and as to what particular weight or importance should be attached to each one of the processes supposed to enter into intelligence. Even if a point scale were to be drawn

[2] *Idem*, pp. 8, 9.

up from this point-of-view, it would result in as arbitrary an allotment of points as in the present Yerkes-Bridges Scale. The accuracy of the Yerkes-Bridges Scale, in spite of this arbitrary allotment of points, seems, to the writers, to be due to the fact that the tests of the scale have been so thoroughly tried out in the old Binet Scale.

The only other point scale known to the writers which has appeared up to the present time is Haines' Point Scale for the Blind.[3] This scale is modeled on the Yerkes-Bridges Scale and assigns points in the arbitrary manner of the latter. Nowhere does the author raise the question of any guiding principle in the allotment of points. The tests are largely adaptations of the tests used by Yerkes and Bridges, along with others devised by Haines and other workers.

The introduction of the point scale has, nevertheless, challenged the attention of workers and we cannot neglect the inevitable question as to the method of allotting points. Point scales of the future will have to adopt some underlying principle according to which points are to be allotted. We shall attempt a brief discussion of what seem to us, at present, possible principles in the allotment of points, bearing in mind always the type of test that we are discussing in the present volume. Some of what we have to say will hardly bear directly

---

[3] Haines, T. H.: "Mental Measurements of the Blind," *Psychological Monographs,* Vol. xxi, No. 1, Whole No. 89 (April, 1916).

upon the type of test included in the present Binet style of scales. All of what we have to say is theoretical and will doubtless be much modified by practical work in the future. Three possible principles that might be formulated at present may be designated as follows: (1) points allotted according to the discriminative capacity of the test; (2) allotment of an equal number of points to each test; (3) points allotted according to the degree of difficulty of test as determined by the standardization. We shall discuss these briefly in the above order.

Discriminative Capacity of Test

The medians at each age of tests such as ours may be taken to give an idea of the discriminative capacity of the test. This will appear clearly on the curve of the age medians. If the curve is a straight line it means that the medians at each age are the same, and obviously the test does not discriminate between six-, seven-, eight- or ten-year-old intelligence. A test showing medians of this sort would possess in terms of this phraseology no discriminative capacity. If the curve were to show a distinct rise (or fall) from age to age for all the ages tested, then the test would possess discriminative capacity for all those ages. It follows that the suddenness of the rise or fall at each age is a measure of the discriminative capacity possessed by a test. Now, as a matter of fact, most curves for intelligence tests show more or less sudden rises or falls at particular ages, and for the most part tend to become more and more level as we approach the higher ages. The allotment of points according to this principle would be made on the basis of the number of points on the curve that might be said to show the discriminative capacity of the test.

In the actual application of the principle the chief source of difficulty would lie in the determination of the amount of steepness a curve must show in order to indicate a real difference between one age group and another. This steepness is again dependent upon the fineness of the grading or scoring system of the test.

Let us take a few concrete examples from our own curves. Graph 15, p. 117, shows the medians at each age for the time taken to complete the Triangle Test. Beginning with the four-year-olds, we find a steep drop from D.N.C. to 275 seconds at age five. The curve at this point is sufficiently steep or the difference between the two medians is sufficiently great to indicate that the test really discriminates between four- and five-year-old intelligence. Continuing, we note a still steeper drop from age five to age six, from 275 to 108 seconds; a fairly steep drop from age six to age seven, from 108 to 77 seconds; a moderately steep drop from age seven to age eight, from 77 to 64 seconds; a moderately steep drop from age eight to age nine, from 64 to 58 seconds; a less steep drop from age nine to age ten, from 58 to 55 seconds, a drop that might be considered doubtful of having real discriminative capacity; from age ten to age eleven the drop is from 55 to 49 seconds; between age eleven and age twelve there cannot be said to be any drop (from 49 to 48 seconds) and, therefore, no discriminative capacity; from twelve to thirteen the drop is steeper, from 48 to 37 seconds, and may be said to show some discriminative capacity; from there onwards the curve is practically level. Our choice of discriminative points on this test might well be at ages five, six, seven, eight, nine, eleven and thirteen. By omitting ages ten and twelve we obtain a steeper drop, from nine to eleven, and again from eleven to thirteen. We might then say that the test shows 7 discriminative points and allot 7 credits to the test. A credit of 1 for a performance of about 275 seconds, i.e., the median for the five-year-olds; a credit of 2 for a performance of about 108 seconds, i.e., the median for the six-year-olds; and so on. The actual limiting points would be determined midway between these points as was done in the median mental age method.

This illustration is merely a suggestion indicative of how the principle might be applied. It will readily be seen that the decision as to what is really a discriminative point on the curve is, in the last analysis, more or less arbitrary, inasmuch as a difference of opinion as to what shall be considered a steep drop is bound to arise. Using the above illustration again, a different choice of points is readily conceivable, such as five, six, seven, eight and thirteen, by which choice only 4 credits would be allowed. Whether the insistence upon great differences between the medians would give better results than a more moderate standard can scarcely be determined theoretically. The best kind of standard to be employed would be shown by practical work with scales constructed on this principle.

The illustration we have taken has been purposely a fairly difficult one in order to show the difficulties of applying the principle. It is difficult because of the fact that we are dealing with a time test where intervals of one second have been used. However, if we are dealing with a test that does not use such small intervals, either of time or score, the principle is easier to apply. In the Cube Test (Graph 27, p. 137) the method of scoring admits of 12 possible scores. Here the curve, if it rises at all, must rise by one of these large units. The determination of the points is easy, since every age at which the curve rises by one unit may be termed a discriminative point. Inspection of the graph shows such points to be at the ages four, five, six, seven, eight, eleven, fourteen, seventeen and eighteen, which would give us 9 discriminative points and, therefore, 9 or 10 credits. The credits would be allotted as follows:

For	a	score	of	1	— 1 point
"	"	"	"	2 or 3	— 2 points

For	a	score	of	4	— 3	points
"	"	"	"	5	— 4	"
"	"	"	"	6	— 5	"
"	"	"	"	7	— 6	"
"	"	"	"	8	— 7	"
"	"	"	"	9	— 8	"
"	"	"	"	10	— 9	"
"	"	"	"	above 10	—10	"

It can be seen from this example how this principle differs in the allotment of points from the Yerkes-Bridges principle of allowing the test itself to determine the number of points. If the test itself were to determine the number of points there would be 12 points allowed, because there happen to be 12 steps in the test. By the principle under discussion only 10 points can be allowed. On the other hand, it may be objected that a score of 3 is better than a score of 2 and, therefore, ought to be given more points, whereas by this method the same number of points is allowed for a score of 3 as for a score of 2. To which the reply would be that it seems just as easy to score 3 as to score 2 on the Cube Test as shown by the medians, and therefore no more points should be allowed.

We believe this principle to be a decided step in advance over the arbitrary method or lack of method employed in point scales up to the present time. Nevertheless there seems to be a valid argument against it. The objection may be summed up in the statement that easy and difficult tests are allowed the same number of points. Or, stated otherwise, a difficult test does not receive more credit than an easy one. This objection is equally true of the point scale of the Yerkes-Bridges type. Is this a valid objection? Let us try to imagine how it will work out in actual practice. Let us imagine two tests each having 3 discriminative points, and therefore 3 credits, the one an easy test and the other a difficult one. The child who passes both will score 6 points, and the child who passes one, only, will score 3 points. As a rule, of course, the child who makes any score on the difficult test will score three points on the easy test as well, but ought not the child passing the difficult test to be given a greater number of points than the child passing the easy one, because he has passed a much more difficult test? In other words, the difference in the scores of these two children will not show the difference in their ability. The ultimate determination will, of course, be made by reference to the norms established by the scale as a whole, but it would seem only fair that a hard test passed by a child should be given more credit than an easy test. If for some extraneous and uncontrollable reason a child fails on an easy test but passes a hard one, he will be penalized very severely, and will receive the same score as the child who passed the easy test and failed on the hard one, because he did not possess enough intelligence to accomplish it. This objection is of the same nature as the one urged by Stern[4] against the Binet Scale and its system of adding 1/5 of a year for an additional test passed regardless of how difficult the additional test might be. The method he proposed, to overcome this difficulty, was to "weight" the test passed according to the year group to which it belonged.

In view of this objection, then, it seems desirable to look for another principle by which to allot points.

Allotment of an Equal Number of Points to Each Test

We do not offer this principle as one which obviates the difficulty raised in the preceding paragraphs, since it will be seen that the same objection applies with almost equal force. It does not, however, permit the anomaly of giving more points for a correct performance of an easy test than for a correct performance of a more difficult one. It gives to the best type of performance on all tests the same number of points. In this particular, then, it may be regarded as somewhat of an advance over the last principle. It arbitrarily assumes all tests to be of equal value and allots a definite number of points to each test. Each different type of performance into which a test divides itself is given the same number of points. The

[4] Stern, W.: The Psychological Methods of Testing Intelligence, Trs. by Whipple, *Educational Psychology Monographs,* No. 13, Warwick and York.

number of different types of performance into which a test divides itself is determined, as in the previous case, by the median performances at each age. Having determined upon the number of median intervals which seem to be discriminative, this number is divided into the definite number of points which have been decided upon. If, for example, we have decided to allot 20 points to each test and if we have 5 discriminative points or types of performance as determined by the actual results of the individuals tested, then we shall allot 4 points to each type of performance. The poorest type of performance will score 4 points, the next 8, and so on up to the best or complete performance, which will have a score of 20. We shall not discuss the application of this principle any further at this point, because we have drawn up a point scale for our tests on this basis, and the application of the principle will be best seen in actually dealing with the tests themselves later on.

The objection urged above to the first principle still holds good here, although perhaps not with equal force. This compels us to look around for some method whereby this objection may be overcome.

Points Allotted According to Degree of Difficulty of Tests

The logical conclusion that we have been driven to, and the only one which seems possible, is to allot points for various types of performance according to the varying degrees of difficulty of different types of performance. Our different types of performance can be determined only by the results from children themselves, and the only question remains as to the determination of the degrees of difficulty of these different types. What is to be the criterion of the degree of difficulty in different steps of the same test and in different steps of different tests? For example: How difficult is a certain performance on the Adaptation Board in comparison with a certain performance on the Cube Test? *A priori* we may say, after seeing the two tests, that a perfect performance on the Adaptation Board is easier than a perfect performance on the Cube Test, but much beyond this we cannot go. A little experience with the tests would lead to a little better comparison, and more experience to still better comparisons of different degrees of difficulty of the one test with the other. These comparisons, it is to be carefully noted, are all based upon the ease or difficulty with which different children perform the tests. So, logically, we are driven back to the best available comparison in the norms for the tests at various ages. To do all the five moves on the Adaptation Board correctly is about as easy or difficult as to do six of the Cube Lines correctly, because average eight-year-olds can do both of these performances. Two moves on the Adaptation Board are about equal to two lines on the Cube Test, because these are the norms for five-year-olds, and so on with all the other tests. Fundamentally, the last judge of the ease or difficulty of a test is the child himself. Average five-year-olds set standards of ease or difficulty for all five-year-olds. Unless we abandon the customary and sound hypothesis as to the development of intelligence with age, the chronological age of the child is the measure of our test. This is the fundamental truth at the bottom of Binet's system of the measurement of intelligence, which no critic of his system has been able to controvert. However clumsy we may deem the way in which Binet himself made use of this fundamental truth, we must admit that the critics of Binet who have put forth the point scale method as a superior device have merely adopted a still more clumsy device in their attempt to apply this fundamental truth. The Yerkes-Bridges Point Scale has worked admirably in practice, thanks to our years of experience with the tests in the Binet Scale. Binet at least attempted to "weight" his tests according to the ability of normal children. The Yerkes-Bridges tests are not "weighted" according to any principle, unless the caprice of the constructor of the scale may be termed a guiding principle.

We have in this discussion been driven to the only logical conclusion that seems possible, namely, that in the allotment of points the underlying principle is the chronological age of the child. The natural application of this is to give to each type of performance a number of points corresponding to the chronological age for which this performance is distinctive. For example, since average five-year-olds can do 2 moves on the Adaptation Board, we must give a score of 5 to 2 moves on this board; for 4 moves on the board we must give a score of 6, since 4 moves is the average performance of six-year-olds. Similarly 5 points must be given to 2 Cube Lines, and 6 to 4 Cube Lines, and 7 to 5 Cube Lines, since these lines are passed correctly by average five-, six- and seven-year-olds respectively.

We might conceivably decide not to give the same number of points as number of chronological years, but begin with an arbitrary number of one point for four-year-old performances, 2 points for five-year-old performances, and so on. Or we might try a still further refinement and argue that, since there is a greater difference in intelligence between the earlier years than the later years of a child's life, we should make a greater difference in the number of points at the earlier years than at the later. For instance, we might give one point to a four-year-old performance, 6 points to a five-year-old performance, 10 points to a six-year-old, gradually decreasing our additional increment of points as we come to the higher ages. All of these schemes seem to us to be more or less justifiable. All of them are a recognition of the fundamental principle that the chronological age determines the ease or difficulty of the test, and reference to it is the only way of properly "weighting" our tests.

Having arrived at this conclusion, the question arises as to whether we gain anything by allotting points, since we are forced to allot points in accordance with the performances in terms of chronological years. If we allot points according to years and then establish norms, will we arrive at anything differing from the median mental age method discussed in Chapter VI? If we establish our age norms for the scale on the basis of averages, they will differ only slightly from the medians and will, we believe, be slightly less desirable. If we use medians, we are doing exactly the same thing as using the median mental age, and this method we have discussed at length in the previous chapter.

We have, therefore, in this discussion of the underlying principles of a point scale method been forced back to the median mental age and are compelled to question the validity of a point scale that differs in principle from the median mental age. At present we see no loophole in this argument. A point scale, as such, seems to have no right to exist. It can only be a modified form of the median mental age method.

Tentative Point Scale

There are, nevertheless, practical reasons in favor of the use of points. There is a certain ease in calculation and in the establishment of a coefficient of mental ability. We have drawn up a tentative point scale on the basis of our second method, i.e., the allotment of an equal number of points to each test. To draw up a point scale according to the third method would have resulted merely in a repetition of the median mental age method. It may be that some workers will desire to use the point scale system, and results expressed in scores lend themselves to all kinds of mathematical treatment.

Taking the principle of the allotment of an equal number of points to each test, the number of discriminative places that each test seemed to show was first ascertained from the tables of medians in Chapter IV and the discriminative steps more or less arbitrarily determined. The number of discriminative steps for each test can be seen in the list of tests on page 175, with their discriminative points, the number of points credit allowed to each discriminative point, and the total points for each test.

A number was chosen which would allow as nearly as possible an equal division into all tests, so that the use of fractions might be avoided. Thirty points for each test was the number chosen, this being large enough to allow differentiation between tests having a different number of discriminative steps, and on the other hand not too large to make the addition of credits for the tests as a group too cumbersome. Thirty or 28 or 32 can be divided by all the different groups of discriminative steps of the tests. The column headed "Points Credit" gives the number of points for each step of the test, and the next column headed "Total"

TEST	SCORE	DISCRIMINATIVE POINTS	POINTS CREDIT	TOTAL POINTS
Mare and Foal	Time	5	6	30
" " "	Errors	5	6	30
Seguin Form Board	Time	5	6	30
Five Figure Board	Time	6	5	30
" " "	Errors	6	5	30
Two Figure Board	Time	6	5	30
" " "	Moves	4	7	28
Casuist Form Board	Time	7	4	28
" " "	Errors	7	4	28
Triangle	Time	7	4	28
"	Errors	7	4	28
Diagonal	Time	6	5	30
"	Errors	8	4	32
Healy Puzzle "A"	Time	8	4	32
" " "	Moves	7	4	28
Manikin	Score	4	7	28
Feature Profile	Time	4	7	28
Ship	Score	6	5	30
Picture Completion	Score	10	3	30
Substitution	Score	8	4	32
Adaptation Board	Moves	5	6	30
Cube	No. Correct	8	4	32
				652

gives the total number of points for each test. Each test is given from 28 to 32 points. The total for all the tests is 652 points.

The number of points being determined, the assignment of them to the discriminative intervals is

Table 25. Table of Credits for the Point Scale.

Test											
1. The Mare and Foal:											
Time	Points	6	12	18	24	30	....	....	....	....	....
	Score	150–89	88–55	54–35	34–25	24–	....	....	....	....	....
Errors	Points	6	12	18	24	30	....	....	....	....	....
	Score	10–5	4–3	2	1	0	....	....	....	....	....
2. Seguin Form Board:											
Time	Poiots	6	12	18	24	30	....	....	....	....	....
	Score	42–32	31–25	24–17	16–13	12–	....	....	....	....	....
3. Five Figure Board:											
Time	Points	5	10	15	20	25	30	....	....	....	....
	Score	300–158	157–107	106–88	87–74	73–53	52–	....	....	....	....
Errors	Points	5	10	15	20	25	30	....	....	....	....
	Score	30–11	10–7	6	5–4	3	2–	....	....	....	....
4. Two Figure Board:											
Time	Points	5	10	15	20	25	30	....	....	....	....
	Score	300–188	187–146	145–89	88–55	54–37	36–	....	....	....	....
Moves	Points	7	14	21	28	....	....	....	....	....	....
	Score	45–20	10–13	12–11	11–	....	....	....	....	....	....
5. Casuist Form Board:											
Time	Poiots	4	8	12	16	20	24	28	....	....	....
	Score	250–130	129–100	99–86	85–73	72–67	66–59	58–	....	....	....
Errors	Points	4	8	12	16	20	24	28	....	....	....
	Score	31–21	20–10	9–8	7–6	5	4	3–	....	....	....
6. Triangle Test:											
Time	Points	4	8	12	16	20	24	28	....	....	....
	Score	300–192	191–93	92–71	70–61	60–47	46–39	38–	....	....	....
Errors	Points	4	8	12	16	20	24	28	....	....	....
	Score	35–20	19–12	11–9	8	7	6	5–	....	....	....
7. Diagonal Test:											
Time	Poiots	5	10	15	20	25	30	....	....	....	....
	Score	300–163	162–113	112–70	69–46	45–34	33–	....	....	....	....
Errors	Poiots	4	6	12	16	20	24	28	32	....	....
	Score	35–19	16–12	11–9	8–7	6	5	4	2	....	....

Table 25. Table of Credits for the Point Scale.—(*Continued*)

Test											
8. Healy Puzzle "A":											
Time....................	Points	4	8	12	16	20	24	28	32	....	....
	Score	300–124	123–102	101–78	77–62	61–50	49–42	41–34	33–	....	....
Moves....................	Points	4	8	12	16	20	24	28	....	....	....
	Score	100–43	42–32	31–26	25–22	21–19	18–15	14–	....	....	....
9. The Manikin Test:											
Score....................	Points	7	14	21	28	....	....	....	....	....	....
	Score	1–2	3	4	5	....	....	....	....	....	....
10. The Feature Profile Test:											
Time....................	Points	7	14	21	28	....	....	....	....	....	....
	Score	300–199	198–154	153–143	142–	....	....	....	....	....	....
11. The Ship Test:											
Score....................	Points	5	10	15	20	25	30	....	....	....	....
	Score	1–10	11–16	17	18	19	20	....	....	....	....
12. Picture Completion Test:											
Score....................	Points	3	6	9	12	15	18	21	24	27	30
	Score	30–121	122–190	197–284	285–368	369–421	422–444	445–474	475–499	500–510	511–
13. Substitution Test:											
Score....................	Points	4	8	12	16	20	24	28	32	....	....
	Score	600⁺–445	444–235	234–169	168–150	149–132	131–115	114–100	99–	....	....
14. Adaptation Board:											
Moves Correct.............	Points	6	12	18	24	30	....	....	....	....	....
	Score	1	2	3	4	5	....	....	....	....	....
15. The Cube Test:											
Lines Correct.............	Points	4	8	12	16	20	24	28	32	....	....
	Score	1	2	3–4	5	6	7	8	9⁺	....	....

the next step. These discriminative intervals with the number of points allotted to each are shown in Table 25. For each test or part of a test for which credit is allowed, the score and the number of points are shown. Thus, for the Mare and Foal Test (Time), 6 points are given for all scores (in this case time in seconds) that lie between 150 and 89. No credit is allowed for scores greater than 150. Twelve points are allowed for scores from 88 to 55; 18 points for scores from 54 to 35; 24 points for scores from 34 to 25; and 30 points for all scores less than 25. The rest of the table is to be read in the same way.

The next logical step in the construction of a point scale would have been to work over the original data allotting points for all the tests and thus arriving at norms for each age. This the writers were unfortunately unable to do, since the children from whom our data were obtained were not tested on all the tests. To arrive at age norms for the scale as a whole, each individual should be tested on all of the tests of the scale. If this is not done, there remains the further possibility of taking the average or median for each test at each age and using the totals of these medians or averages for the age norms for each age. This is the method we have employed to arrive at tentative age norms.

Practically, there will be little difference in the age norms thus established from the age norms arrived at by taking averages for all children tested on all the tests of the scale. We find a practical

TABLE 26. MEDIAN SCORES FOR EACH TEST AND EACH YEAR. POINT SCALE METHOD.

Test	Age												
	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Mare and Foal:													
Time...........	6	0	12	12	18	18	18	24	18	24	24	24	30
Errors..........	0	6	12	12	18	18	24	18	18	24	18	18	30
2. Seguin:													
Time...........	0	6	12	18	18	18	24	24	24	30	30	30	30
3. Five Figure Board:													
Time...........	0	0	5	10	15	20	25	25	25	30	25	25	30
Errors..........	0	0	5	10	15	20	20	20	25	25	25	20	30
4. Two Figure Board:													
Time...........	0	5	10	15	20	25	25	25	25	30	30	25	30
Moves..........	0	7	7	7	14	21	14	21	14	21	21	21	28
5. Casuist Form Board:													
Time...........	0	0	0	4	8	12	16	20	24	16	28	20	28
Errors..........	0	0	4	8	12	16	20	20	24	28	28	24	28
6. Triangle Test:													
Time...........	0	4	8	12	16	20	20	20	20	28	24	28	28
Errors..........	0	4	8	12	16	16	16	20	20	28	24	24	28
7. Diagonal Test:													
Time...........	0	5	10	15	15	20	25	20	20	30	30	25	30
Errors..........	0	4	8	12	12	20	20	12	16	28	32	24	32
8. Healy Puzzle "A":													
Time...........	0	0	0	4	8	12	16	20	24	28	20	32	32
Moves..........	0	0	0	4	8	12	16	20	24	24	24	28	28
9. Manikin Test:													
Score...........	7	14	21	21	28	28	28	28	28	28	28	28	28
10. Feature Profile Test:													
Score...........	0	0	0	0	0	0	7	14	14	21	28	21	28
11. Ship Test:													
Score...........	0	0	10	10	15	20	20	25	30	30	20	25	30
12. Picture Completion:													
Score...........	0	3	6	9	12	15	18	21	24	27	30	30	30
13. Substitution Test:													
Score...........	0	4	8	12	16	20	24	28	32	32	32	28	32
14. Adaptation Board:													
Score...........	6	12	24	24	30	30	30	30	30	30	30	30	30
15. Cube Test:													
Score...........	4	8	12	16	20	20	20	24	24	24	28	24	32
Total.......	23	88	182	247	334	401	446	479	503	586	579	554	652

example of this in Yerkes-Bridges' [5] norms. From a table showing the average score for each test and each year, we were able to arrive at a total number of points for each year, and to compare the total points with the age norms established by taking the average of the total scores made by the children on all the tests of the scale. At 4 ages the two norms are exactly the same, at 7 ages there is a difference of only 1 point, and at 1 age only is there a difference of 2 points. The discrepancy is, therefore, not very great. We may, therefore, obtain from our data tentative norms in this manner until a more accurate standardization is judged feasible.

From the tables of medians for each age the number of points was determined, and these median number of points for each test for each year are given in Table 26. The table is to be read as follows: In the Mare and Foal Test (Time) the median number of points scored by the four-year-olds is 6, by the five-year-olds 6, by the six-year-olds 12, and so on up to 30 points by the sixteen-year-olds. The other lines are to be read in the same way. At the bottom of the table is given the total score for each age, and these values are the age norms for use in practical testing. Graph 28 represents these age norms graphically. It will be noted that there is a steady increase in the norm from age four up to age thirteen. Ages fourteen and fifteen fall below age thirteen and the norms

[5] Yerkes, Bridges and Hardwick: *Op. cit.*, p. 123.

Graph 28.—Age Norms for Point Scale Method. Ordinates indicate points scored; abscissae, age by years.

here are not to be relied upon. Age sixteen is given a perfect score, since we have no data beyond age sixteen. Individuals at any age making a score of 600 would undoubtedly show excellent ability on these tests. The results of individuals could be compared with these norms in the usual way and the coefficient of mental ability obtained by dividing the score made by the norm for the age of the child tested. As to where the limiting points between feeble-mindedness and backwardness and normality would lie, there would be the usual difference of opinion, but these points could be set as easily on this scale as on any other.

One drawback of the point scale, as the writers see it, has been touched upon in this discussion: namely, that for the establishment of valid norms the cases must be tested on all the tests of the scale. In other words, when we set out to standardize a point scale, we must fix upon our tests beforehand and depart from them afterwards at the risk of spoiling the standardization or of having to resort to some such device as we have been compelled to use.

Another drawback, due to this interlocking character of any point scale, is due to the fact that with any individual case that we are testing we must use all the tests before we can employ our results to any advantage. The age norm is established on the basis of all the tests and we run a serious risk of doing injustice to a case if we omit any test. A child often scores points where an examiner may have felt that any score was impossible. This objection is, of course, equally valid for the year scale, but not for the median mental age method nor for the percentile method. If we do omit a test on the point scale, we should have to go back to our table and compute new norms with this test omitted. This would be a very inconvenient and laborious procedure in practical work.

The discussion, in this chapter, of the principles of the point scale is an attempt on our part to raise the whole question of the validity and justification of the point scale method. We do not claim to have given a final answer to this question. The results of our tests have been presented for use as a point scale, although we are well aware of the tentative nature of the norms as established and the limitations of the point scale method.

CHAPTER VIII

THE PERCENTILE METHOD

THE presentation of the results of tests in the form of percentile tables is a comparatively recent innovation in the history of mental tests. It has arisen naturally with the testing of large groups of individuals. The method would be impossible with few cases. It has arisen, also, from a desire to know what the distribution of a group really is in respect to the various portions that go to make up the total group. Our belief that individuals, in regard to all kinds of abilities, distribute themselves on a normal curve with the very good ones at one end and the very poor at the other, rather than into distinct types, is leading us to insist more and more upon a presentation of results that can be interpreted in this manner. The 25 and 75 percentiles so commonly used at present are the result of our desire to know what the middle 50 per cent or "normal" group of the individuals tested can do. The addition of other percentile points gives us a finer means of discrimination. It has long been customary to consider the middle 50 per cent normal, the upper 20 or 15 per cent bright, the uppermost 10 or 5 per cent very bright, the lower 20 or 15 per cent poor, and the lowest 10 or 5 per cent very poor. The division into 10 percentiles will allow us to increase our groups greatly, and in time to attach a definite meaning to each of the 10 percentile abilities.

A further very decided advantage of the percentile method for purposes of mental testing is that it allows us to compare each individual with individuals of the same age. The individual is ranked according to the performances of individuals of like age, while at the same time allowing cross comparisons with any percentiles of any other age.

Woolley[1] seems to have been the first to suggest the percentile method for practical use in mental testing. She presents the results of all her tests in tables of 10 percentiles. Using these as indices of the child's ability on each test, an average of the percentiles for a number of tests gives an index for the child. Her distribution tables, giving the averages of the percentile ranks, show a decided tendency toward the normal type of distribution. In addition to this she has taken the next logical step and presented a percentile table of average percentile ranks. That is, the average percentile rank of an individual resulting from all the tests of a scale can now be interpreted in the light of all the average per-

---

[1] Woolley, H. T.: "A New Scale of Mental and Physical Measurements for Adolescents, and Some of Its Uses," *Journal of Educational Psychology*, Vol. vi, No. 9 (1915), pp. 521-550.

centile ranks of all the children tested on the scale, and can be placed in its proper percentile. In this way norms can be established in terms of the average percentile rank of a group of tests. As a matter of fact, these norms could be used as age norms in much the same way as the points in the point scale; but if we once begin to think in terms of percentiles, we will take the next logical step, as Woolley has done, and convert them into a percentile table of distribution from which we may read off the percentile of the average percentile rank.

As we have said in a previous chapter, it appears to us that this type of standardization is the most thorough and may ultimately prevail over all other types. It allows of the finest differentiations and the most just comparisons of an individual with individuals of the same age. The only drawback that we can see to it at present is that, for a reliable determination of all of our ten percentile points, a very large number at each age will be necessary.

We have presented our results in tables of percentiles so that they may be used by workers adopting this method. These percentiles have been computed from the tables of distribution in Chapter IV. The accumulation of more data and the addition of this to the tables of distribution would lead, of course, to a gradual perfection of our percentile points. It is conceivable that some day 5 percentile points may be practicable, if mental testing ever has need of such fine discriminations. At the present time, however, 10 percentile points give us finer differentiations than we have so far been accustomed to use.

The Tables

Tables 27 to 48 give the ten percentile distribution at each age for all the twenty-two tests used. Table 27 is to be read as follows: On the Mare and Foal Test, Time, the best or 100 percentile score for five-year-olds is 50, and (continuing down the vertical column) the 90 percentile score, or the score reached by the 90 per cent child, from the poorest upwards, is 63; for the 80 percentile child the score is 70, and so on down the columns. All the other tables are to be read in the same way.

Cross comparisons can readily be made from these tables. For example: In Table 27 we note that the best five-year-olds (time 50) are equal to the 80 percentile six-year-olds, about the 60 percentile seven-year-olds, the 40 percentile eight-year-olds, the 25 percentile nine-year-olds, the 15 percentile ten-year-olds, the 10 percentile eleven-year-olds, the 20 percentile twelve-year-olds, and make scores as bad as the very poorest thirteen- and fourteen-year-olds. The very poorest thirteen- and fourteen-year-olds can perform this test as quickly as the very brightest five-year-olds. These kinds of comparisons may lead to some norm in the future and may materially help our understanding of the general development of intelligence.

The practical working out of the percentile method will be shown in the later discussion of

TABLE 27. THE MARE AND FOAL TEST. PERCENTILES. TIME.

Age	5	6	7	8	9	10	11	12	13	14
Per Cents.										
100	50	30	20	20	20	20	20	10	10	10
90	63	44	34	30	25	24	23	22	21	22
80	70	50	38	35	30	28	25	25	23	25
70	78	61	44	39	34	32	28	29	25	28
60	95	66	49	44	38	34	32	32	27	29
50	107	71	62	48	41	36	34	35	29	31
40	116	78	66	52	45	39	37	39	30	35
30	150	86	73	57	48	43	39	42	32	36
20	190	98	87	66	53	47	44	49	36	39
10	280	136	109	83	64	54	49	57	42	46
0	D.N.C.	D.N.C.	D.N.C.	D.N.C.	120	120	70	79	49	50

TABLE 28. THE MARE AND FOAL TEST. PERCENTILES. ERRORS.

Age	5	6	7	8	9	10	11	12	13	14
Per Cents.										
100	2	0	0	0	0	0	0	0	0	0
90	4	1	1	0	0	0	0	0	0	0
80	5	2	1	1	1	0	1	0	0	1
70	6	2	2	1	1	1	1	1	1	1
60	7	3	2	1	1	1	1	1	1	1
50	7	3	3	2	2	1	2	2	1	2
40	8	4	3	2	2	2	2	2	2	2
30	10	5	4	3	3	2	2	3	2	2
20	12	6	5	4	4	3	3	3	3	3
10	14+	8	7	5	5	4	4	4	4	5
0	D.N.C.	D.N.C.	D.N.C.	D.N.C.	8	12	7	8	5	6

TABLE 29. THE SEGUIN FORM BOARD. PERCENTILES. TIME.

Age	5	6	7	8	9	10	11	12	13	14
Per Cents.										
100	22	18	15	14	13	12	9	10	9	9
90	26	20	18	16	15	13	12	11	11	9
80	29	21	20	17	15	14	13	12	11	10
70	32	24	21	19	16	15	14	12	12	10
60	34	24	22	20	17	15	14	13	12	11
50	37	26	23	20	18	16	15	14	12	11
40	39	27	24	21	18	17	15	14	13	12
30	42	29	25	21	20	18	16	15	13	12
20	44	31	27	23	21	19	16	15	14	13
10	50	34	28	25	24	21	17	17	15	14
0	75	41	38	32	34	27	24	23	17	17

TABLE 30. THE FIVE FIGURE FORM BOARD. PERCENTILES. TIME.

Age	5	6	7	8	9	10	11	12	13	14
Per Cents.										
100	40	30	30	30	30	20	20	20	20	20
90	107	78	60	56	44	40	38	34	29	29
80	124	95	74	64	53	47	45	39	33	41
70	185	132	85	71	58	52	49	46	36	49
60	224	159	97	85	68	61	57	51	39	54
50	D.N.C.	200	117	97	79	69	64	58	47	59
40	D.N.C.	249	149	120	92	82	73	66	53	66
30	D.N.C.	D.N.C.	191	137	116	96	81	77	59	72
20	D.N.C.	D.N.C.	D.N.C.	184	160	117	99	94	67	100
10	D.N.C.	D.N.C.	D.N.C.	285	D.N.C.	145	162	121	86	190
0	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	262	150	237

TABLE 31. THE FIVE FIGURE FORM BOARD. PERCENTILES. ERRORS.

Age	5	6	7	8	9	10	11	12	13	14
Per Cents.										
100	1	0	0	0	0	0	0	0	0	1
90	6	3	3	2	1	1	1	1	1	1
80	7	5	3	3	2	1	1	1	1	2
70	16	7	4	3	3	2	2	2	2	3
60	19	10	6	5	4	3	3	2	2	3
50	D.N.C.	14	7	6	4	4	4	3	3	3
40	D.N.C.	21	9	7	5	4	5	4	4	4
30	D.N.C.	D.N.C.	14	10	7	6	6	5	5	5
20	D.N.C.	D.N.C.	D.N.C.	15	10	8	7	6	5	7
10	D.N.C.	D.N.C.	D.N.C.	25	D.N.C.	11	11	11	7	13
0	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	30	22	20

TABLE 32. THE TWO FIGURE FORM BOARD. PERCENTILES. TIME.

Age	5	6	7	8	9	10	11	12	13	14
Per Cents.										
100	30	20	20	10	10	10	10	10	10	10
90	65	44	42	29	26	26	23	23	16	19
80	85	54	51	37	31	29	25	28	21	34
70	113	76	59	43	36	35	28	31	24	28
60	137	126	77	49	40	40	31	36	27	29
50	200	175	116	62	47	47	38	39	29	35
40	300+	300	157	79	60	62	44	46	32	45
30	D.N.C.	D.N.C.	224	102	79	101	56	55	38	55
20	D.N.C.	D.N.C.	D.N.C.	149	106	122	75	64	45	78
10	D.N.C.	D.N.C.	D.N.C.	275	224	198	123	89	79	120
0	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	300+	160	D.N.C.

TABLE 33. THE TWO FIGURE FORM BOARD. PERCENTILES. MOVES.

Age	5	6	7	8	9	10	11	12	13	14
Per Cents.										
100	10	9	9	9	9	9	9	9	9	9
90	12	10	11	9	9	9	9	9	9	9
80	15	11	12	11	10	10	10	10	10	9
70	16	14	13	11	11	11	10	11	10	10
60	18	18	16	13	11	11	11	11	10	10
50	21	24	20	14	12	13	11	12	11	11
40	35	44	24	16	13	15	12	13	11	12
30	D.N.C.	D.N.C.	32	20	17	19	14	14	13	12
20	D.N.C.	D.N.C.	D.N.C.	24	21	24	17	16	13	13
10	D.N.C.	D.N.C.	D.N.C.	40	40	37	25	19	19	21
0	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	45	45	D.N.C.

TABLE 34. THE CASUIST FORM BOARD. PERCENTILES. TIME.

Age	5	6	7	8	9	10	11	12	13	14
Per Cents.										
100	80	50	40	40	30	20	20	30	20	20
90	112	98	65	56	48	49	43	44	49	41
80	174	132	84	65	59	56	50	51	55	44
70	249	163	105	79	65	63	55	55	62	49
60	281	228	118	93	79	71	62	62	67	53
50	D.N.C.	300	154	106	93	78	68	66	75	58
40	D.N.C.	D.N.C.	186	130	114	89	79	75	86	63
30	D.N.C.	D.N.C.	236	151	155	116	94	94	99	79
20	D.N.C.	D.N.C.	D.N.C.	189	224	149	114	116	120	90
10	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	208	179	224	149	173
0	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	300	300	300

TABLE 35. THE CASUIST FORM BOARD. PERCENTILES. ERRORS.

Age	5	6	7	8	9	10	11	12	13	14
Per Cents.										
100	5	0	1	0	0	0	0	0	0	0
90	10	6	4	2	1	1	1	1	1	1
80	18	10	6	3	2	2	2	2	2	1
70	29	13	8	4	3	3	3	3	2	1
60	31+	21	9	6	5	4	4	3	3	2
50	D.N.C.	30	12	8	7	5	5	4	3	3
40	D.N.C.	D.N.C.	14	12	10	7	6	5	5	3
30	D.N.C.	D.N.C.	23	14	11	9	7	6	6	4
20	D.N.C.	D.N.C.	D.N.C.	18	21	14	10	8	7	8
10	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	21	13	16	10	12
0	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	31+	29	17

TABLE 36. THE TRIANGLE TEST. PERCENTILES. TIME.

Age	5	6	7	8	9	10	11	12	13	14
Per Cents.										
100	30	10	10	10	10	10	10	8	10	10
90	72	30	28	25	25	17	20	19	15	14
80	103	55	38	35	33	29	27	27	19	18
70	178	70	45	43	38	37	34	34	26	28
60	220	85	56	52	46	46	44	39	32	33
50	275	108	77	64	58	55	49	48	37	39
40	D.N.C.	190	103	79	81	74	65	58	43	48
30	D.N.C.	D.N.C.	154	106	104	84	77	76	49	70
20	D.N.C.	D.N.C.	266	173	139	98	108	99	78	100
10	D.N.C.	D.N.C.	D.N.C.	250	176	137	128	158	100	125
0	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	265	175	250

TABLE 37. THE TRIANGLE TEST. PERCENTILES. ERRORS.

Age	5	6	7	8	9	10	11	12	13	14
Per Cents.										
100	5	0	1	0	1	0	0	0	1	1
90	10	4	3	2	3	3	2	2	2	2
80	11	5	5	4	4	4	3	3	3	2
70	19	7	6	5	5	5	4	5	4	4
60	22	11	8	6	7	6	7	6	4	4
50	27	13	11	8	8	8	7	7	5	6
40	D.N.C.	18	14	10	11	9	10	8	6	7
30	D.N.C.	D.N.C.	23	16	15	10	12	10	8	8
20	D.N.C.	D.N.C.	39	24	17	14	15	14	9	10
10	D.N.C.	D.N.C.	D.N.C.	33	23	22	19	19	12	12
0	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	35+	25	16

TABLE 38. THE DIAGONAL TEST. PERCENTILES. TIME.

Age	5	6	7	8	9	10	11	12	13	14
Per Cents.										
100	10	10	10	10	10	9	10	9	9	10
90	49	25	23	19	17	16	16	13	13	13
80	94	41	39	25	22	22	22	16	15	16
70	120	66	50	34	29	29	26	26	19	19
60	155	85	62	49	37	35	41	35	20	20
50	275	150	75	76	49	42	54	54	25	25
40	D.N.C.	200	94	106	79	52	89	70	32	29
30	D.N.C.	290	137	175	108	70	120	90	39	35
20	D.N.C.	D.N.C.	250	225	159	107	149	137	56	49
10	D.N.C.	D.N.C.	D.N.C.	D.N.C.	300+	198	224	200	88	98
0	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	170	200

TABLE 39. THE DIAGONAL TEST. PERCENTILES. ERRORS.

Age	5	6	7	8	9	10	11	12	13	14
Per Cents.										
100	0	0	0	0	0	0	0	0	0	0
90	4	2	2	1	0	1	1	1	0	0
80	7	5	4	3	1	2	2	1	1	1
70	10	9	6	5	2	4	4	3	2	1
60	13	10	8	6	5	5	6	4	3	1
50	23	14	9	9	6	6	9	7	4	2
40	D.N.C.	19	12	16	9	8	13	9	5	4
30	D.N.C.	38	19	21	15	9	18	14	6	6
20	D.N.C.	D.N.C.	35	32	20	15	25	20	7	7
10	D.N.C.	D.N.C.	D.N.C.	D.N.C.	37	24	34	27	10	14
0	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	25	25

TABLE 40. HEALY PUZZLE "A." PERCENTILES. TIME.

Age	5	6	7	8	9	10	11	12	13	14	15
Per Cents.											
100	40	10	10	10	10	10	8	8	5	10	5
90	138	90	33	26	26	18	16	17	15	15	10
80	274	159	46	38	36	28	22	23	20	21	15
70	D.N.C.	199	75	71	45	41	28	30	26	29	20
60	D.N.C.	262	118	89	68	52	38	38	33	30	24
50	D.N.C.	D.N.C.	131	117	86	70	54	46	38	55	30
40	D.N.C.	D.N.C.	D.N.C.	183	132	92	68	60	46	82	48
30	D.N.C.	D.N.C.	D.N.C.	D.N.C.	181	120	85	71	60	90	75
20	D.N.C.	D.N.C.	D.N.C.	D.N.C.	215	186	107	118	70	109	120
10	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	300+	225	174	95	174	213
0	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	275	D.N.C.	275

TABLE 41. HEALY PUZZLE "A." PERCENTILES. MOVES.

Age	5	6	7	8	9	10	11	12	13	14	15
Per Cents.											
100	10	5	5	5	5	5	5	5	5	5	5
90	15	19	11	7	9	8	7	7	7	7	6
80	D.N.C.	30	16	13	13	10	9	9	9	9	7
70	D.N.C.	44	22	18	17	15	12	12	12	11	9
60	D.N.C.	60	28	25	19	18	14	14	14	16	11
50	D.N.C.	D.N.C.	50	35	28	23	20	18	16	17	14
40	D.N.C.	D.N.C.	D.N.C.	47	39	27	24	24	18	25	17
30	D.N.C.	D.N.C.	D.N.C.	D.N.C.	42	32	29	26	20	30	27
20	D.N.C.	D.N.C.	D.N.C.	D.N.C.	60	50	38	31	23	34	37
10	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	130	69	45	40	50	67
0	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	75	D.N.C.	70

TABLE 42. THE MANIKIN TEST. PERCENTILES. SCORE.

Age	5	6	7	8	9	10	11
Per Cents.							
100	5	5	5	5	5	5	5
90	5	5	5	5	5	5	5
80	4	5	5	5	5	5	5
70	3	5	5	5	5	5	5
60	3	4	5	5	5	5	5
50	3	4	4	5	5	5	5
40	2	4	4	4	5	5	5
30	2	4	4	4	4	4	5
20	2	3	4	4	4	4	4
10	0	3	4	4	4	3	4
0	0	0	0	1	2	3	3

TABLE 43. THE FEATURE PROFILE TEST. PERCENTILES.

Age	6	7	8	9	10	11	12	13	14	15	16
Per Cents.											
100	225	90	50	20	40	40	30	30	30	30	...
90	D.N.C.	212	135	148	75	71	61	55	49	70	...
80	D.N.C.	D.N.C.	172	212	116	79	85	78	66	99	...
70	D.N.C.	D.N.C.	236	260	155	111	106	106	89	102	...
60	D.N.C.	D.N.C.	D.N.C.	D.N.C.	196	133	133	131	111	108	...
50	D.N.C.	D.N.C.	D.N.C.	D.N.C.	240	157	170	150	132	150	110
40	D.N.C.	D.N.C.	D.N.C.	D.N.C.	281	192	224	180	139	208	...
30	D.N.C.	D.N.C.	D.N.C.	D.N.C.	300+	224	300+	266	191	285	...
20	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	299	D.N.C.	300+	241	300+	...
10	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	300+	D.N.C.	...
0	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	D.N.C.	...

TABLE 44. THE SHIP TEST. PERCENTILES. SCORE.

Age	5	6	7	8	9	10	11	12	13	14	15
Per Cents.											
100	17	20	20	20	20	20	20	20	20	20	20
90	14	20	20	20	20	20	20	20	20	20	20
80	11	18	18	19	20	20	20	20	20	20	20
70	5	17	17	18	19	20	20	20	20	20	20
60	0	15	16	18	19	19	20	20	20	19	20
50	0	15	15	17	18	18	19	20	20	18	19
40	0	13	13	16	18	18	18	19	20	18	18
30	0	10	11	15	17	17	18	18	19	17	18
20	0	5	8	14	16	16	17	17	17	17	17
10	0	0	3	3	15	14	16	16	17	16	17
0	0	0	0	0	3	9	14	8	16	13	16

TABLE 45. THE PICTURE COMPLETION TEST. PERCENTILES. SCORE.

Age	5	6	7	8	9	10	11	12	13	14	15	Ad.
Per Cents.												
100	394	525	646	646	646	646	646	646	646	646	646	646
90	259	354	463	507	577	578	583	583	646	646	646	646
80	207	260	364	446	499	522	578	578	581	578	583.5	583
70	169	219	318	422.5	450	509	519	547	566	570	577.5	578
60	126	185	263	381	440	478	501	509	515	521.5	567	569
50	89	153	239	328	407	435	455.5	493	505	515	525	515
40	49	99	189	296	372	411	443	445	494	501	501	509
30	28	76	153	261	325	367	411	467	446	444	444	469
20	5	62	107	201.5	297	313	369	369	417	415	380	441
10	2	14	60	129	240	251	296	313	310	366	329	360
0	0	0	0	3	62	35	122	36	47	106	144	162

TABLE 46. THE SUBSTITUTION TEST. PERCENTILES. SCORE.

Age	5	6	7	8	9	10	11	12	13	14
Per Cents.										
100	250	150	100	90	80	70	60	60	60	60
90	280	175	147	116	109	92	79	78	75	78
80	360	212	160	133	121	100	92	88	82	85
70	600	233	168	144	126	108	96	95	89	90
60	600	265	177	152	133	116	101	100	93	95
50	600	290	180	158	141	123	107	106	96	99
40	600	333	210	168	152	129	113	111	100	110
30	D.N.C.	365	239	188	158	137	119	117	107	116
20	D.N.C.	400	278	206	173	146	125	125	119	124
10	D.N.C.	500	300	221	190	165	146	143	135	143
0	D.N.C.	D.N.C.	600	D.N.C.	219	300	180	190	170	160

TABLE 47. THE ADAPTATION BOARD. PERCENTILES. MOVES.

Age	5	6	7	8	9	10	11	12	13	14
Per Cents.										
100	5	5	5	5	5	5	5	5	5	5
90	4	5	5	5	5	5	5	5	5	5
80	3	5	5	5	5	5	5	5	5	5
70	2	4	5	5	5	5	5	5	5	5
60	2	4	4	5	5	5	5	5	5	5
50	2	4	4	5	5	5	5	5	5	5
40	2	3	4	4	4	5	5	5	5	5
30	1	3	3	4	4	4	5	5	5	5
20	1	2	3	3	4	4	4	5	5	4
10	1	1	2	3	3	3	4	4	4	4
0	1	0	0	1	1	1	3	3	3	3

TABLE 48. THE CUBE TEST. PERCENTILES.

Age	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Ad.
Per Cents.																	
100	..	..	6	8	10	9	10	10	11	10	11	11	11	12	..	..	11
90	..	..	5	6	8	8	8	9	9	9	9	10	10	10	..	..	10
80	..	..	4	6	6	7	7	8	8	8	8	9	9	9	..	..	10
70	..	..	4	5	6	6	7	7	8	8	8	8	8	9	..	..	9
60	..	..	3	4	5	6	6	7	7	7	7	8	7	8	..	..	8
50	1	1	2	4	5	6	6	6	7	7	7	8	7	8	9	10	8
40	..	..	2	4	5	5	5	6	6	6	7	7	6	7	..	..	8
30	..	..	1	3	4	5	5	6	6	6	6	7	6	7	..	..	7
20	..	..	1	3	4	5	4	5	5	5	6	6	5	6	..	..	7
10	..	..	1	1	3	3	3	4	4	4	5	5	5	6	..	..	6
0	..	..	0	0	2	1	2	2	1	2	2	2	3	5	..	..	4

actual cases tested by the scale. At present we may refer to one obvious drawback of the method as applied to some of our tests. When some contiguous percentile points show the same scores, the question arises as to what percentile rank should be assigned to the score in question. For example, in Table 44 we note that at age eleven the 60, 70, 80, 90 and 100 percentile points are all the same, i.e., a score of 20. If an eleven-year-old child scores 20 on this test, what percentile rank are we to assign to him? There seem to be two possibilities. We may, in the first place, give him the rank which is the median of these four percentiles, i.e., 75. Or, secondly, we may give him the lowest percentile rank, i.e., 60, arguing that a 60 percentile eleven-year-old child can do such a performance, and that, therefore, such a performance is a 60 percentile type of performance for eleven-year-olds. This latter method would of course penalize the bright child on a test where a perfect performance is a relatively easy performance for him. The first method of allowing the median percentile rank is open to the objection of allowing too much credit to some children, although it is less severe on the really bright child.

Similar percentile points tend to occur in all tests which do not allow of fine gradations. They are very rare in time tests, rather uncommon in tests showing the number of moves or errors, but very common in tests having a limited range of scores. Table 44 may be taken as a sample of this phenomenon where the similar percentile points occur in the upper percentiles and in the higher ages. The children, however bright, cannot make higher scores, because at a score of 20 the test stops. Table 43 shows the converse of this. Here D.N.C. score appears in the lower percentiles and at the lower ages. The children fail on the test and we do not discriminate among different kinds of failure. Tables 48 and 28 show a scattering of similar percentiles all over the tables. Tables 47 and 42 show the most pronounced cases of similar percentile points. This is due to the fact that the range of scores is merely from 0 to 5. To make tests adaptable to the percentile method, it may prove desirable to devise means for a finer differentiation of the performances on such tests, i.e., adopt a finer system of scoring. Tests of this type, with a narrow range, are best adapted to a year scale, since they do not discriminate between many age groups. In the final perfection of a percentile scale it may prove desirable to limit the scope of these tests and refuse to give any credit to a perfect score above certain ages. In such a case the test would be omitted from the scale and the median or average percentile rank of the other tests would be taken. For example, in the Manikin Test (Table 42) a perfect score at any age above five is hardly discriminative. It might prove best to omit such a test with all children age six and above making a perfect score. Conversely a D.N.C. score on the Feature Profile Test is not discriminative from age nine downwards and a similar procedure might be adopted here.

This last point suggests the great advantage of the percentile method in the omission and admission of tests. As in the median mental age method, we are able to add and subtract tests much more readily than with the year or point scales. To be sure, the establishment of norms of percentile ranks for all ages is the desirable goal of this method and would require a standardization based on the same tests for all children, but we need not wait for such norms for rough diagnostic purposes. This possibility of the addition and subtraction of tests gives a great flexibility to the scale.

CHAPTER IX

ILLUSTRATIVE CASES

THIS chapter will deal with two sample cases tested on the complete scale, with the results worked out by all four methods of computing mental age. Figures 13 and 14 show the two records of the cases, with the mentality computed by all the four methods. Our record blank is so arranged that a worker may use any or all of the four methods for arriving at the mentality of the case.

Figure 13 shows the record of Arthur S., a fifteen-year-old boy, who is in the fifth grade of school. The record for time, errors, moves and score for the fifteen tests is shown under the column headed "Record." All the data actually required for further elaboration is found in the vertical column directly under the heading "Record." Additional space is provided in several tests for the scoring of the test or for time records that may be required. Thus, in Test 2, we take the record of the first, second and third trials, although only the shortest trial is required for the scale. In Tests 9, 11 and 12 space is provided for the time, although we have not deemed it wise to take this into account in scoring the tests. For Test 12 further space is pro-

Date 21 Sept. 1916. Examined by M. Pintner.
Name Arthur S. Sex Boy
Date of Birth 8 Oct. 1900. Nationality American. School Grade 5 A.

PSYCHOLOGICAL CLINIC

THE OHIO STATE UNIVERSITY

RECORD BLANK FOR PINTNER-PATERSON PERFORMANCE SCALE

Chronological Age (Age last birthday) 15 Diagnosis
Median Mental Age 10·25 Points 484 Point Age 11·2
Median Percentile 41 Mental Age (Year Scale) 13·2

Test. Name of Test.	Record	Median Mental Age	Percentile	Points	Age	4	5	6	7	8	9	10	11	12	13	14	15
					Test Value	.33	.12	.05	.06	.06	.07	.07	.07	.09	.07	.14	.33
1. Mare and Foal.	Time 24	16	83	30			5	6	7	8	9	10	11		13		
	Errors 1	13·5	70	24			5	6						12			
2. Seguin Form Board. Time 1 34 2 21 3 18	Time of Shortest Trial 18	9	0	18			5	6	7	8	9		11			14	
3. Five Figure Board.	Time 76	9	32	20				6	7	8	9	10	11	12	13		
	Errors 5	9	30	20				6	7	8	9				13		
4. Two Figure Board.	Time 32	12	55	30				6	7	8	9		11	12	13		
	Moves 9	16	90	28				6	7	8	9		11			14	
4. Casuist Form Board.	Time 46	15	76	28				6	7	8		10	11			14	
	Errors 1	15	80	28					7	8		10	11		13		
6. Triangle Test.	Time 60	9	35	20				6	7	8		10		12	13		
	Errors 5	14	55	28				6	7	8		10		12	13		
7. Diagonal Test.	Time 89	7	15	15				6	7		9	10			13	14	
	Errors 10	7	15	15				6	7		9	10			13		
8. Healy Puzzle "A."	Time 143	7	15	4						8	9	10	11	12	13		
	Moves 26	9	33	12						8	9	10	11	12	13		
9. Manikin Test. Time 87	Score 4	7	15	21		4		6	7		9						
10. Feature Profile.	Time 206	11	40	7								10	11			14	15
11. Ship Test. Time 79	Score 18	9·5	35	20			5	6	7	8	9	10	11		13		
12. Picture Completion. Time 270	Score 530	14	55	30			5	6	7	8	9	10	11	12	13	14	15
13. Substitution Test. Time 101 Errors 1	Score 103	12	42	28			5	6	7	8	9	10	11	12	13		
14. Adaptation Board. Moves 1 + 2 + 3 + 4 + 5 +	Score 5	8	65	30		4	5	6		8			11				
15. Cube Test. A + X + Y + B + C + D + E + F + G — H — I — J —	Number Correct 8	14	70	28		4	5	6	7	8		10				14	15
		Median Mental Age 10·25	Median Percentile 41	Total Points 484	Number of Tests	✓	✓	✓	✓	✓	13	14	12	8	6	5	2
					Test Values					8	·91	·98	·84	·72	·42	·70	·66 = 5·23

13·23

SCORING RECORD FOR TEST 12

PICTURE COMPLETION TEST

Name of Piece	log	broken window	flying bird	dog	cat	basket	foot ball	hat	chicken	Total Score
Record	✓	✓ c. wind.	✓	✓	✓	cherries.	cherries ✓	✓	✓	
Score Value	52	32	87	64	81	7	84	65	58	530

vided, at the bottom of the blank, for recording the moves made and for filling in the score value. In Test 13 space is provided for time and errors, from which two values the score is computed. In Tests 14 and 15 the moves are recorded, either plus or minus, during the progress of the test, and the final score is in each case the number of moves correct.

By reference to Table 23, p. 152, we are able to fill in the column headed "Median Mental Age." The figures in this column give the approximate mental ages to which the performances opposite to them correspond. Thus, on the Mare and Foal Test, Time, the boy makes a performance equal to that of a median sixteen-year-old, whereas the quality of his performance (one error) is about equal to that of a 13.5-year-old child; and so on down the column. The mental ages on the tests vary from 7 to 16. The poorest performances are on the Diagonal Test, the Healy "A" and the Manikin. His best performances are on the Mare and Foal (Time) and the Two Figure Board (Moves). The median of all the 22 mental ages is a mental age of 10.25, which is an interpolated median between 9.5 and 11.

The computation of the mentality according to the percentile method is shown under the column headed "Percentile." The values are found by reference to the tables of percentiles (Tables 27 to 48). The time on the Mare and Foal Test is about equal to an 83 percentile for fourteen-year-old boys.[1] One error on this test is about a 70 percentile performance. The performance on the Seguin is zero, since it is below the lowest record. The rest of the percentiles have been obtained in the same way. It will be noted that rough interpolations between ten percentile points given in the tables have been made. Wherever the values were the same for several contiguous percentile points we have taken the median percentile value. Thus in Test 11 a score of 18 is found at both the 30 and 40 percentiles and so we have recorded a percentile performance of 35. The final value is a median percentile of all the percentile points recorded. It is again an interpolated median between 40 and 42. This median percentile of 41 means that on the scale as a whole the child's performance is somewhat below the median.

The point scale method is illustrated by the figures under the column headed "Points." The values for each test are obtained from Table 25, and the total number of points, 484, corresponds to a mental age of 11.2 by reference to Table 26. This value, 11.2, we have designated "Point Age," which means the mental age obtained by the point scale method.

The year scale method is illustrated on our blanks and has been filled out as previously explained in Chapter V. It will be noted that the basal age of

[1] We have used age fourteen or lower ages in working out the percentiles, where data for age fifteen or higher were not available.

the case is 8, but the boy makes additional credits in other years to the extent of 5.23 years and this brings his mental age up to 13.2.

We thus have three mental ages and a percentile estimate. The three mental ages are:

Median Mental Age	10.25
Point Age	11.2
Mental Age (Year Scale)	13.2

Which of these is the most significant and which is the truest estimate of the child's mentality cannot be determined at present.

This case was also tested on the Yerkes-Bridges Point Scale. He scored 63 points, which gives him a mental age of 10.8 (using the combined norms of the authors) and gives him a C.M.A. of .79. The mental age of 10.8 on this scale lies between the median mental age of 10.25 and the point age of 11.2 on our scale.

The next record shown in Figure 14 is that of another fifteen-year-old boy. Looking down the records of the tests we note at once a decidedly inferior performance as contrasted with the previous case. There are five tests which he fails to complete.

The median mental age is about 5. The median actually falls between 5 and —6. The ages for the different tests fluctuate between —5 and 10. The great number of minus quantities shows the need for norms of younger children and for simpler

Date 21 Sept. 1916 Examined by M. Lininer

PSYCHOLOGICAL CLINIC
THE OHIO STATE UNIVERSITY

Name Philip A. Sex Boy

Date of Birth 2 Jan. 1901 Nationality American

School Jackson Grade 3

Chronological Age (Age last birthday) 15 Diagnosis

Median Mental Age 5 Points 118 Point Age 5.5

Median Percentile 0 Mental Age (Year Scale) 6.05

Test. Name of Test.	Record	Median Mental Age	Percentile	Points	Age	4	5	6	7	8	9	10	11	12	13	14	15
					Test Value	.33	.12	.05	.06	.06	.07	.07	.07	.09	.07	.14	.33
1. Maze and Foal.	Time 159	5	0	0			5	6	7	8	9	10	11		13		
	Errors 39	-5	0	0			5	6						12			
2. Seguin Form Board. Time 1 77 2 46 3 40	Time of Shortest Trial 40	5	0	6			5	6	7	8	9		11			14	
3. Five Figure Board.	Time D.N.C	-6	0	0				6	7	8	9	10	11	12	13		
	Errors 64	-6	0	0				6	7	8	9				13		
4. Two Figure Board.	Time D.N.C	-5	0	0				6	7	8	9		11	12	13		
	Moves 59	-5	0	0				6	7	8	9		11			14	
5. Casuist Form Board.	Time D.N.C	-6	0	0				6	7	8		10	11			14	
	Errors 65	-6	0	0					7	8		10	11		13		
6. Triangle Test.	Time 225	5	5	4				6	7	8		10		12	13		
	Errors 49	-5	0	0				6	7	8		10		12	13		
7. Diagonal Test.	Time 112	7	5	15				6	7		9	10			13	14	
	Errors 10	7	15	15				6	7		9	10			13		
8. Healy Puzzle "A."	Time 82	9	25	12						8	9	10	11	12	13		
	Moves 25	10	33	16						8	9	10	11	12	13		
9. Manikin Test. Time 81	Score 3	6	0	14		4		6	7		9						
10. Feature Profile.	Time D.N.C	-10	5	0								10	11			14	15
11. Ship Test. Time	Score D.N.C	-5	0	0			5	6	7	8	9	10	11		13		
12. Picture Completion. Time 315	Score 11	-5	0	0			5	6	7	8	9	10	11	12	13	14	15
13. Substitution Test. Time 450 Errors 11	Score 549	5	0	4			5	6	7	8	9	10	11	12	13		
14. Adaptation Board. Moves 1 + 2 + 3 + 4 − 5 +	Score 4	7	15	24		4	5	6		8			11				
15. Cube Test. A + X + Y − B − C − D − E − P − G − H − I − J −	Number Correct 2	5	0	8		4	5	6	7	8		10				14	15
		Median Mental Age 5	Median Percentile 0	Total Points 118	Number of Tests	✓	6	5	3	3	4	3	2	1			
					Test Values	4	.72	.25	.18	.18	.28	.21	.14	.09			= 2.05

6.05

SCORING RECORD FOR TEST 12

PICTURE COMPLETION TEST

Name of Piece	log	broken window	flying bird	dog	cat	basket	foot ball	hat	chicken	Total Score
Record	cage. pumpkin. candle. hat.	clock	cage	b. wind. baby.	f. bird.	b. wind. chicken.	b. wind. baby. f. bird. s. ean.	pumpkin hatchet	bucket shoe	
Score Value	0	0	7	2	2	0	0	0	0	11

FIGURE 14

tests. It is interesting to note that the highest mental ages are obtained by the performance on the Healy Puzzle "A." We have discussed previously the degree of chance entering into this test, and this record seems to be a good example of this factor.

The median percentile is 0. There are 15 out of the 22 performances in which a 0 is scored. This score of 0 shows the case to belong to the very poorest of fifteen-year-olds and we may take this record to mean feeble-mindedness.

The total points scored by the point scale method is 118, which is equivalent to a point age of about 5.5.

The year scale record shows that the case obtained a basal age of 4 and additional credits to the extent of 2.05 years, which brings his mental age up to 6.05.

The three mental ages according to the different methods are:

Median Mental Age	5
Point Age	5.5
Mental Age (Year Scale)	6.05

The Yerkes-Bridges test of this boy shows a score of 25, which is equivalent to a mental age of 5.4 with a C.M.A. of .31. In this case the Yerkes-Bridges record agrees closely with the point age.

The discussion of these two cases will have made clear the method of keeping the record sheet and of computing the different mental ages. Just which of the different methods is the best it is impossible to tell at this time. The practical worker cannot, of course, be expected to work out the results for all four methods, but, nevertheless, we have made it possible to use any method on our record sheet. When we have accumulated a sufficient number of complete records a study of these different methods will be made.

CHAPTER X

CONCLUSION

We shall attempt in this concluding chapter to summarize briefly the main points covered in the preceding chapters.

1. A scale of performance tests as a means of estimating mentality is needed for those children who are deficient or wanting in language.

2. Such a scale is the only means that can be used to measure the intelligence of the deaf, the speech defective and the non-English speaking individual.

3. Language ability is not uniformly correlated with general intelligence and, therefore, a scale of performance tests will be a useful supplement to other scales which depend entirely or in part upon language responses.

4. The need for a more adequate standardization of most of the performance tests in common use has led to an effort on our part to supply this deficiency.

5. The value of such performance tests is greatly enhanced when they are grouped together in some kind of a scale.

6. The results of the tests are presented in tables

of distribution so that additional results may be added from time to time and the reliability of the norms thereby increased.

7. Four different methods of arriving at an index of mental ability have been discussed.

8. The year scale method has the advantage of leading to a result that is easy to interpret, but it has the disadvantage of requiring a great many different tests. This would make the scale unwieldy and would lengthen, beyond practical limits, the time taken to examine a case.

9. We have attempted to construct with our tests a modified type of year scale. This type of year scale differs somewhat from the type of year scale in common use. This difference is necessary if we are to overcome the disadvantages in the year scale method mentioned in the preceding section.

10. The median mental age method is simple in computation and permits the addition or subtraction of tests without dislocating the whole scale. Difficulties arise when the medians are the same for several consecutive ages. The diagnostic significance of the median mental age has yet to be determined.

11. The point scale method has been subjected to a discussion in order to find out the most satisfactory underlying principle upon which to base a point scale. The result seems to lead back to a method closely akin to the median mental age method and one showing no superiority over that method.

12. A point scale has been constructed on the principle of the allotment of the same number of points to each test, although the value of this method of procedure is doubtful.

13. The percentile method seems to offer the best possibilities for future work. The percentile division used can be made as small as the delicacy of the tests will warrant. This method is especially desirable because it permits us to compare an individual's performance with the performances of other individuals of the same age. It would seem at present, however, to require, for purposes of standardization, a very great number of unselected individuals at each age.

14. These different methods lead to different estimates of mentality for the same individual. Which leads to the truest estimate of intelligence is a problem still to be solved.

15. The correlation of this scale with scales of the Yerkes or Binet type has not yet been attempted. Whether a scale of performance tests or a mixed scale of performance and language tests will yield the best estimate of intelligence has yet to be determined.

From the nature of these concluding remarks it should be obvious that we have attempted to avoid being dogmatic upon the subject of scales and methods of testing intelligence. We feel keenly that the present stage of development of the work of mental testing is one in which all methods and devices must be tried in order to lead to more and more accurate work. Any attempt to set up one scale or method as perfect would merely serve to retard the splendid progress that this branch of psychology has made within the last decade. In accordance with this belief we have presented our work in many different forms, with the result that no one cut-and-dried method lies ready for the mere "mental tester." For the practical psychologist, however, we would recommend the median mental age method, because of its simplicity and because it enables the examiner to see at a glance whether the subject's performances on the various tests have been uniform or erratic. The examiner, however, must always exercise great caution in the interpretation of his final result.

We are well aware of the tentative nature of our work and only too conscious of what still remains to be done. We shall feel amply repaid if we have made one step forward in the construction of a different type of scale for the measurement of mental ability.

INDEX OF AUTHORS

INDEX OF AUTHORS

INDEX OF SUBJECTS

INDEX OF SUBJECTS