1924

Study of performance tests

Wood, Miriam Wealthy
Boston University

http://hdl.handle.net/2144/5423
Boston University
BOSTON UNIVERSITY

SCHOOL OF EDUCATION

Thesis

A STUDY OF PERFORMANCE TESTS

Submitted by

Miriam Wealthy Wood

1924

In partial fulfilment of requirements for the degree of Master of Education.
A STUDY OF PERFORMANCE TESTS.

OUTLINE

I. Introduction
   A. Psychologists have been seeking measures of intelligence for thirty years.
   B. There were distinct tendencies in the devising of tests.
      1. Tests emphasizing physical attributes. Description of the studies of workers in this field.
      2. Age-level tests. Description of the work of Binet and Simon and their followers.

II. Historical data in regard to performance tests.
   A. Knox introduced a scale for testing immigrants. Description
   B. Pintner and Patterson inaugurated a scale for testing the deaf.
      1. Purpose of this scale.
      2. Composition of the scale.
      3. The conclusions of the authors.
   C. Wooley Scale.
      1. Description of tests used.
   D. Healy schedule of tests for clinical use.
      1. Tests included.
   E. Outstanding contributors of single performance tests.
      Dearborn, Stenquist, Knox, Witmer, Goddard, Healy.

III. Conditions of the present statistical study of performance tests.
   A. Data collected at Judge Baker Foundation.
      1. Age, sex, and I.Q. of 3600 children.
      2. Records of their performance on Healy Construction.
A and B, Mare and Foal Picture Board, Goddard Form Board, Witmer Cylinders, Dearborn Formboard #2 and #3, Healy Puzzle Box, Knox Feature Profile, Stenquist Mechanical Assembly.

B. Method of gathering the data.
   1. Minimum schedule of tests given.
   2. Supplementary tests.

C. Actual basis for this study.
   1. Records of 289 children.
   2. Nature of the group.
   3. Distribution table.

IV. The tests studied.

A. Healy Construction A.
   1. Description
   2. Discussion
      a. Attempts at standardization and their value.
      b. Recent Judge Baker Foundation study and its conclusion.
   3. Method of ranking performances in this study.

B. Healy Construction B.
   1. Description
   2. Discussion
      a. Four standardizations and their value.
      b. Recent study at Judge Baker Foundation and its conclusions.
   3. Method of ranking performances in this study.

C. Mare and Foal Picture Board.
   1. Description
   2. Discussion of norms and standardizations.
   3. Method of ranking performances in this study.
D. Goddard Form Board
1. Description
2. Discussion of norms and standardizations.
3. Method of ranking performances in this study.

E. Witmer Cylinders
1. Description
2. Discussion of norms and standardizations.
3. Method of ranking in this study.

F. Dearborn Formboard #2
1. Description
2. Discussion of norms and standardizations.
3. Method of ranking in this study.

G. Dearborn Formboard #3
1. Description
2. Discussion of norms and standardizations.
3. Method of ranking in this study.

H. Healy Puzzle Box
1. Description
2. Discussion of norms and standardizations.
3. Method of ranking in this study.

I. Knox Feature Profiles
1. Description
2. Discussion of norms and standardizations.
3. Method of ranking in this study.
J. Stenquist Mechanical Assembly

1. Description

2. Discussion of norms and standardizations.

3. Method of ranking in this study.

V. Test results.

A. Tabulation of ranks on each test for each child.

B. Use of results to test out Construction A as an indicator.

C. Coefficients of correlation between Construction A and each of the other tests.

VI. Conclusion of statistical study.
A STUDY OF PERFORMANCE TESTS

INTRODUCTION

For at least thirty years psychologists have been concerned with finding measures of intelligence. In the beginning their studies concentrated upon finding physical tests which would show some correlation with a child's school grade and mental brightness as estimated by his school teacher. As early as 1893 Townsend (1) correlated the height and weight of 33,500 school children with their age and grade. In 1894 Gilbert (2) correlated weight, height, lung capacity, simple reaction time, reaction time with discrimination and choice, and time memory with mental ability. The teacher's judgment was the determinant of general mental ability and each teacher classified her pupils as dull, average, and bright. Approximately 100 children of each age group from six to seventeen were studied.

Following these studies the American Psychological Association appointed a committee to investigate the subject of mental and physical tests. This committee agreed upon and reported a series of tests which it recommended be tried on college students in the various psychological laboratories of the country. (3) The following list of the tests suggested shows clearly that the emphasis was still upon physical attributes:

Preliminary Data. - Date of birth; birthplace; birth-
place of father, birthplace of mother; occupation, including
class in college; occupation of father; any measure previously
made. Color of eyes; color of hair; right or left handed.
Mother's maiden name; number of brothers; sisters; order of
birth; age of parents at birth; birthplace and occupation of
grandparents. Asymmetry of body; color of eyes, hair, com-
plexion; degeneracy or other stigma of head, eyes, ears, mouth,
teeth, hands, feet, posture; gait; manners; coordination and
speech; indications of intellectual, emotional and moral
characteristics.

Physical Measurements. - Height, weight and size of head.
Breathing capacity. Height sitting.

Keenness of Vision
Color Vision
Keenness of Hearing
Perception of Pitch
Fineness of touch
Sensitiveness to Pain
Perception of Weight or force of movement
Dynamometer pressure of right and left hand
Rate of movement
Fatigue
Will power
Voluntary attention
Right and left hand movement
Rapidity of movement. - Taps on telegraph, short marks, trill-
ing with two fingers or five.

Accuracy of aim
Reaction time for sound
Reaction time with choice. - Card sorting
Rate and discrimination of movement. - Marking out 100 a's in
500 letters, one of a number of geometrical figures, or
colors, or pictures, or objects.

Quickness of distinction and movement. - Rate at which cards
are sorted, combine with reaction, with choice, with ef-
facts of practice.

Perception of size. - Draw a line equal to a model 5 cm. in
length, bisect it, erect a perpendicular of the same
length, and bisect the right hand angle.

Perception of time. - The accuracy with which a standard time
can be reproduced.

Memory. - The accuracy with which eight numerals heard once
can be reproduced, and the accuracy with which a line drawn
by the observer at the beginning of the hour can be reproduced at the end of the hour; line to be identified, not drawn; ten numerals; nine numerals. A combined test of memory, association and finding time as described in the catalogue of the Columbian Exposition, accuracy of observation and recollection as proposed by Cattell and Bolton.

Memory type. - Variation in the use of ten numerals, compare results for indication of memory type and kind of imagery preferred.

Apperception Test of Ebbinghaus Imagery.

This committee was greatly assisted in its choice of tests by the work of J. McKean Cattell (4) who had inaugurated a series of experiments with undergraduates at Harvard University in 1887 and had continued his experiments at University of Pennsylvania and Bryn Mawr College in 1888 and 1889, and in the following years at Columbia University.

The next few years saw the accomplishment of a great amount of testing along these same lines. Smedley (5) in 1900 correlated height, standing and sitting; weight; ergograph and dynamometer records, and lung capacity with age and school standing of children between the ages of eight and eighteen inclusive. Reaction time tests of many kinds were tried out by various investigators and the results of their work were reviewed by Whipple in 1904 (6). He concluded that any reaction time is conditioned upon a large number of independent factors and when these are eliminated or controlled in the laboratory "we have left no residuum of individual variation that can be turned to account in estimating the observer's general intelligence or mental ability". In 1901 Wissler (7) published the results of Cattell's work for a period of seven years with a long series of tests and anthropometrical measurements. The general conclusions were that the laboratory mental tests show
little correlation in the case of college students; that the physical tests show a general tendency to correlate with themselves but only to a very slight degree with mental tests; that the markings of students in college classes correlate with themselves to a considerable degree but not with the tests made in the laboratory.

Burt departed somewhat from the thus far established trend in testing (8) in his study of two sets of English schoolboys. He correlated general intelligence with tests of discrimination of two points upon the skin, of lifted weights, of pitch, and of length of lines. To these he added two motor tests, tapping and card dealing; two sensory-motor tests, card sorting and alphabet finding; tests of immediate memory of concrete words, abstract words and nonsense syllables; the tracing of a geometrical pattern seen in a mirror, a test of the power to acquire new coordinations; the reproduction from memory of a pattern of spots presented by the tachistoscope upon squared paper; and a test of voluntary attention, which consisted of prickng an irregular line of dots passing rapidly before the subject. Of the results of these tests Burt says "Of the twelve tests, six furnish coefficients below .50 and six above .50. The former six - the simple sensory and motor tests - are thus of little use in the empirical diagnosis of intelligence. Among the latter six, no single test, at any rate in its present form, can be claimed as a self-sufficient instrument for measuring and detecting ability in individuals. But they indicate the direction in which such a test may hopefully be sought . . . . McDougall's dotting
machine seems to be the most scientific. Where the external conditions could be kept most uniform, both the amalgamated and the average raw coefficients reached .84. The mirror test can be procured with but little trouble and expense, and needs no trained superintendent. It, too, requires further improvements, especially in procedure and calculations to eliminate the influence of possible previous practice, and to elicit more completely the significance of the figures observed. If called upon to recommend a simple test for immediate use upon untrained subjects, I should be inclined to advocate the alphabet test as perhaps the simplest and most satisfactory test of all.

However, the problem of the best methods of clinical testing still remained unsolved. Burt's conclusions were valuable from the standpoint of an interpretation of intelligence but his tests could not be taken over for clinical work for several reasons. In the first place time was too large a factor in them; in the second the apparatus used in the tests which correlated most highly with general intelligence was too unusual and formidable not to interfere with the ease and confidence of the child requisite for reliable results. Furthermore in some tests previous practice might vitiate results. Therefore Binet and Simon (9) compiled an entirely different set of tests when in 1904 they undertook the task of segregating the mentally defective children in the public schools of Paris. Binet undertook to arrange a series of tests capable of practical application to young children. He strove to eliminate the quantitative measurement of results and substitu-
ted a qualitative measure. The requirement of laboratory conditions was discarded in favor of situations more in accord with the normal everyday life of children. There is no necessity for a detailed description of each one of these tests. They were arranged into a scale, approximately five tests at each age level through fourteen, and consisted of such tasks as comparing weights, copying designs, repeating digits and sentences heard, defining words, describing pictures, naming colors, comparing objects, making change, answering common-sense questions, solving problems. An attempt was made to use only those tests which are unaffected by educational advantage. It was assumed that there is regular progress in mental growth and that, therefore, a mental age as well as a chronological age, could be determined.

Binet revised his first scale in 1908. Then followed wide use of this scale both in Europe and America. In 1911 Binet made a second revision. Many experimenters in America added their testimony as to the usefulness of this series of tests. Goddard (10) applied it to four hundred inmates of the Vineland School for feebleminded children and to two thousand public school children. Kuhlman (11) gave the tests to institution children at Faribault, Minnesota; Terman and Childs (12) gave them to a large group of normal children in California. These experimenters concluded that there were certain outstanding fallacies in the scale, namely:

1. The assumption of serial mental development from early childhood to adult age.

2. The omission of tests of socially significant abilities.
3. Failure to distinguish certain innate abilities from a certain expression of them due to age or experience.

4. Failure to use tests which are an accurate measure of mental development of normal children.

5. The assumption that a defective is quantitatively rather than qualitatively different from a normal individual.

6. Failure to include tests which are a clue to industrial possibilities.

The third distinct departure in the field of testing came with the realization of the inadequacies of both the early physical measurements and the Binet-Simon series. A great deal of criticism had been directed particularly against the tests of the Binet Scale which required language responses. Ayres (13) made this criticism shortly after the scale had come into general use. Just how much the ability to handle language was indicative of intelligence was the question at issue. Was a test valid when ability to pass it depended not merely upon comprehension of language but also upon the ability to frame an adequate language response? This language difficulty became very pronounced as the use of the scale spread to such practical workers as clinical psychologists in large cities face to face with the problems of foreign, deaf, speech-defect and other difficult children. In 1911 Healy and Fernald (14) worked out a set of tests of an entirely different character from any which had preceded them. They had two purposes in mind; first, to supplement the Binet-Simon series, second to overcome the language difficulty confronting clinical workers by devising tests which would reduce to a minimum the use of language on the part of the examiner and eliminate the necessity for its use by the subject. Their first three tests were of the type
now commonly known as performance tests or tests which measure ability to work with concrete material.

It is with the performance test which measures ability to work with concrete material that this study has to deal, the interest being especially centered in ten specific tests in this field.
HISTORICAL

Once research had been started in the field of performance tests the devising of new ones and the standardization of them went on apace. Knox (15) at Ellis Island had the problem of testing non-English speaking immigrants and found it impossible to use the existing scales in which language responses were required, even though he employed interpreters. He therefore devised a series of performance tests which he constructed into a kind of scale. Many of his tests were excellent but his scale was admittedly rough and lacking in standardization.

Pintner and Patterson (16) were faced with the problem of testing the mentality of deaf children and found the ordinary scales of intelligence absolutely inadequate. They, therefore, gathered together existing performance tests, standardized them, devised others, and gradually developed an entire scale of performance tests. These they have presented in great detail in their book on the subject. (17)

They felt that their scale could be used as a supplement to the ordinary scales of intelligence involving language as well as independently in dealing with deaf, foreign and speech-defect children. They included in their scale:

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 Patterson.
4. The Two Figure Board, devised by Pintner.
5. The Casuist Form Board, a copy of the original board devised by Knox.
6. The Triangle Test, devised by Gwyn.
7. The Diagonal Test, devised by Kempf.
10. The Feature Profile Test, devised by Knox and Kempf.
11. The Ship Test, devised by Glueck.
12. The Picture Completion Test, devised by Healy.
13. The Substitution Test, devised by Woodworth and Wells.
15. The Cube Test, devised by Knox and modified by Pintner.

As a result of their work they stated frankly that of the four methods of arriving at an index of mental ability, which they discussed, the one which would lead to the truest estimate had not yet been determined. Furthermore it remained an open question whether a scale of performance tests or a mixed scale of performance and language tests would yield the best estimate of intelligence.

Subsequent research and invention in the field of performance tests has consisted largely in revision of the Pintner and Patterson scale or devising new single tests. There have been few complete scales. Wooley (18) contributed an unusual and original scale for adolescents. It is the outcome of measurements by from 600-800 adolescents of ages fourteen and fifteen. None of the tests are taken from the Binet Scale. They cover a wide range including physical tests, tests of motor ability, as well as purely mental tests and many involve the use of language. The scope of the scale is, of course, very limited and it has been standardized for ages fourteen and fifteen only.

The tendency has been to work away from the use of a scale and from the emphasis upon a single mental age score or rank gained from such a scale, and more and more to emphasize
the findings of each individual test especially as they indicate special abilities or disabilities. Many clinics have followed the lead of Healy (19) in Chicago who evolved a schedule of tests which he found helpful in clinical work. These were not incorporated into a scale but the results of each were evaluated separately. Performance tests were included but did not play a major part in the schedule. The group of tests also included a great variety of tests of special abilities such as memory powers, powers of attention, ability to give testimony, motor coordination, associative processes, perceptions of form and color relationships, learning ability, language powers, mental representations, apperceptions.

Some of the outstanding contributors of single performance tests are Dearborn, Stenquist, Knox, Witmer, Goddard. Some of the most widely used tests involving work with concrete material are Dearborn Formboard #3, Stenquist Mechanic Assembly, Knox Feature Profile, Witmer Cylinders, Goddard Formboard, and Healy Construction Puzzles A and B. In general tests have sprung up like mushrooms, too often quite unanalyzed as to their application and unaccompanied by norms or any helpful guides regarding their use.
EXPERIMENTAL CONDITIONS.

In connection with the foregoing descriptive study of the rise and growth of performance tests, a small study of actual test results with a group of children who were given a considerable number of performance tests, has been attempted. Stated briefly, the aim of the study was to examine certain data which had been collected relative to approximately 3600 children studied at Judge Baker Foundation during the years 1917 to 1924. This consisted of the age, I.Q., and sex of each child and their scores on the following ten performance tests: Healy Construction Puzzle A, Healy Construction Puzzle B, Mare and Foal, Picture Board, Goddard Form board, Witmer Cylinders, Dearborn Formboard #2, Dearborn Formboard #3, Healy Puzzle Box, Knox Feature Profile, Stenquist Mechanical Assembly. The treatment of results will be concerned with the examination of correlations existing between the various ratings under consideration.

The investigation differs from many which have preceded it, in that the tests were given, not as a part of a definitely planned experiment, but as a part of the routine testing of every child referred to Judge Baker Foundation for study. The fundamental purpose in regard to the group was to determine general ability as well as special abilities and disabilities. Each child was given a minimum schedule of tests which included:

1. Stanford Binet
2. Thorndike Reading
   Ayres' Spelling
   Courtis Arithmetic
Whenever a child had a vocational or educational problem, a considerable number of tests were given beyond the minimum schedule, including several performance tests. When this study was started it was with the hope that the group of children who had been given a battery of performance tests would be sufficiently large to afford a valid basis for conclusions as to intercorrelations of performance tests. However of the 3600 there proved to be only 289 children who had been given some performance tests besides the Healy A and Healy B of the minimum schedule. The accompanying table shows the distribution of this group in relation to the ten tests which have been selected for consideration. (I) This is not a highly selected or homogeneous group. Children of all ages, nationalities and social strata are represented. Undoubtedly it is a great disadvantage that only a few of the group were given certain tests while the numbers with other tests are much larger. However, it was of course impossible to fill in the gaps, and, it was thought worth while to see what could be gleaned from the material at hand, rather than stage a new experiment where scientifically controlled conditions would not be lacking. It was at first thought possible to work out some comparisons between performances of boys and
### Table I. The Distribution of the Tests.

<table>
<thead>
<tr>
<th>Test</th>
<th>Number Boys</th>
<th>Number Girls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construction A</td>
<td>213</td>
<td>49</td>
<td>262</td>
</tr>
<tr>
<td>Construction B</td>
<td>212</td>
<td>44</td>
<td>256</td>
</tr>
<tr>
<td>Mare and Fleal</td>
<td>18</td>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>Goddard Form Board</td>
<td>23</td>
<td>12</td>
<td>35</td>
</tr>
<tr>
<td>Witmer Cylinders</td>
<td>38</td>
<td>9</td>
<td>47</td>
</tr>
<tr>
<td>Dearborn #2</td>
<td>36</td>
<td>6</td>
<td>42</td>
</tr>
<tr>
<td>Dearborn #3</td>
<td>116</td>
<td>39</td>
<td>155</td>
</tr>
<tr>
<td>Puzzle Box</td>
<td>83</td>
<td>3</td>
<td>86</td>
</tr>
<tr>
<td>Knox Feature Profile</td>
<td>31</td>
<td>7</td>
<td>38</td>
</tr>
<tr>
<td>Shenquist Assembly</td>
<td>139</td>
<td>4</td>
<td>143</td>
</tr>
</tbody>
</table>
girls, but because of the smallness of the group all consideration of sex differences had to be abandoned.

THE TESTS STUDIED.

The tests being considered in this study include Stanford, Revision of Binet for general intelligence rating and the following ten tests of ability to work with concrete material:

1. Healy Construction Puzzle A.
2. Healy Construction Puzzle B.
3. Mare and Foal Picture Board
4. Goddard Formboard
5. Witmer Cylinders
6. Dearborn Formboard #2
7. " #3
8. Healy Puzzle Box
9. Knox Feature Profile
10. Stenquist Mechanical Assembly

The following descriptions of the tests together with the methods of ranking each will help to clarify the treatment of results. The lack of standardization in several instances necessitated an arbitrary scheme of ranking performances.

Healy Construction Puzzle A.

This test was always given as a part of the minimum schedule and therefore in some cases was not given on the same day as the remainder of the tests with the exception of Healy Construction B. However the interval of a few days was not considered an appreciable deterrent to the comparison of test results.

Description.

The test was originally described by Healy and Fernald, the authors (14). It is made up of an outer frame and five pieces, two of which are identical in size and
shape, which fill up the frame opening when properly placed. The test may be accomplished with a minimum of five moves, one for each piece. There are eleven possible errors without repetition. In giving the test, records of the number of moves and the time for its accomplishment are kept. The moves are grouped under the three headings: 1. possible moves, 2. impossible moves, 3. repetitions of impossible moves. The frame is placed before the child with the pieces scattered on the table beside it and he is told that the pieces will exactly fill the frame if he finds the right way to put them in. The result is counted a failure if the task is not accomplished in five minutes. The record of time and moves is not so important in evaluating the performance as the subjects' method of procedure. Deliberation and planfulness in attacking the problem show much better ability in handling concrete material than swiftness of execution by trial and error method. The authors say, "this test brings out perception of relationship of form, and also the individual 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."

Discussion. An article about to be published in the Journal of Applied Psychology by the writer and two associates, gives a full account of the standardizations of this test and also a new set of norms based upon 1596 records.
taken from the files of Judge Baker Foundation. (20) There have been at least ten attempts at standardization prior to the Judge Baker Foundation study. The first real attempt was that of Schmitt (21) who tested approximately 150 children in the first six grades of a private school. "We feel that this study was invalidated by the utilization of a small, highly selected group, and the presentation of the results in a form so complicated and obscure as to offer no basis for comparison," (20). In the same year Hall (22) standardized the tests from the performance of 180 children and concluded that Construction A is a test for nine year olds since no greater proficiency is attained at higher levels.

Bruckner and King (23) regarded the test as a learning test, a use for which it is not intended as is pointed out by Bronner (24) in her article on the test published a few months later. She presented norms obtained from testing 437 children. Her group was too small and too highly selected to warrant reliable conclusions. Pintner and Paterson (17) also standardized the test and concluded that it was an age level test.

The Judge Baker Foundation study by the writer and associates (20) gives norms based upon the performances of 1596 children. The conclusions of this study were:

1. Construction Test A cannot be considered an age-level test.

2. The test results cannot be differentiated on the basis of sex.
3. Construction Test A is best considered a test of special ability.

4. "There is a wide range within which it is safe to assume the performance of the average child should fall, the older children tending to approximate the higher level."

5. "By comparison with this range, very superior and inferior performances can be readily detected."

In ranking the performances on this test to determine coefficients of correlation, first consideration was given to moves. The finer differentiations were then made on the basis of time. For instance, of two performances in 5 moves that of 5 moves in 5 seconds was ranked as higher than that of 5 moves in 7 seconds. Method of procedure was not taken into consideration in the ranking.

**Healy Construction Puzzle B.**

This test also given as a part of the minimum schedule and therefore the same time difference which was noted in regard to Construction A in relation to the other tests, is also found here.

**Description.**

This is another of the Healy - Fernald tests (14) It consists of eleven pieces to be arranged to fit six openings of a frame. Three of the pieces are of identical shape and size and four others are paired in the same way. Two of the openings are the same in size and shape. Three of the openings sustain a one to one relationship with the pieces which will fill them, thus leaving three openings to be filled with eight pieces. In the accomplishment of the puzzle only one arrangement of pieces is possible, with the exception of the two identical openings which permit of an alternate arrangement of their respective pieces.
In doing the test one may perceive the relationship between all the openings and the pieces so perfectly as to accomplish the task with no error. In the actual performance of the test it is usually accomplished by first placing the pieces which have a one to one relationship with their openings and thus reducing the task to its simplest form. As some of the pieces when put together will fill some of the openings, but leave the task unaccomplished because of pieces and openings which do not fit, there is the possibility of trial and error which may bring success. In this trial and error the subject does not take into account all of the openings and all of the pieces, but only the relationship of part of the openings and pieces. In this type of reaction to the test there are sixteen possible errors. In addition to the two types of performance above, there is another lower type of reaction in which pieces are placed without reference to their spatial relationships; as for example when a circular piece is put into a rectangular opening. If this type of reaction succeeds, it does so by chance. In such a performance the only ability measured is that of recognizing success, and of being able to keep the goal in mind until it is obtained. There is, too, a still lower type of performance where the subject is quite unable to conceive the object of the task. A record of time and moves is kept. The separate placings of each individual piece are recorded. Each placing is considered a move. The result is counted a failure if the task is not accomplished in five minutes. The method of procedure is the important aspect of the performance of this test as well as of Healy Construction A.
Discussion

This test has had much less subsequent standardization than Healy Construction A. Four attempts at standardization are known. The first was by Schmitt (21). The same objections can be made to her treatment of this test and her choice of a highly selected group as were made in regard to Healy Construction A. The second study was by Hall (22) who concluded that Construction B is properly a test for 11 year olds, since no greater proficiency is attained at higher levels. She tested only thirty children at each level. Another special study is that of Weidensall (25) who standardized the test for 88 adult delinquent women. She clearly intimates that she does not consider it an age-level test. A recent contribution is that of Dewey, Childs, and Ruml (26). They used for their subjects Jewish children from the New York City Public Schools and explicitly state that their norms can be used only in reference to racial groups and do not apply to the average American child. The study by the writer (20) and her associates gives ample evidence for the conclusion that this test, like Construction A cannot be considered an age-level test but is best rated as a test of special ability. Here, too, there is a wide range within which it is safe to assume the performance of the average child should fall. By comparison with this range, very superior and inferior performances can be readily detected.

In this study performances on Construction B are ranked just as on A. The first consideration is given to
moves and further differentiation is on the basis of time.

**Mare and Foal**

This is perhaps the simplest of the group of tests under consideration. Its bright colors and animal pictures excite interest immediately. Its simplicity disarms all suspicion of difficulty. It is given largely to very young subjects.

**Description**

The test was devised by Healy and Fernald and is described by them (14). It is a board measuring 29 by 24.5 centimeters 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. Eleven pieces have been cut out of the picture. Seven pieces are cut on the natural lines of some of the objects in the picture, and four are cut on geometrical lines. Two of these last somewhat resemble each other and the third is an isosceles triangle divided into two right angle triangles. Many experimenters have omitted these four pieces from the test by simply gluing them into place. That has been our procedure also.

The puzzle is placed before the child with the pieces scattered at random on the table with the instructions that the subject put the pieces in the right places as quickly as possible. Time and the number of errors are taken. There is a time limit of five minutes. The test
presents a one to one relationship between the openings and the pieces to be placed. Many children make no errors at all in placing because the situation permits of easy solution by means of planfulness and deliberation. However, it is possible to use a trial and error method.

Discussion: This test has been standardized by Schmitt (21) who used all of the pieces and gave results on 132 children. She found that "trial and error is small after the kindergarten, when eighty per cent, or more, or each grade perform the puzzle with less than three errors; 89 per cent of the kindergarten children make less than six errors."

Pintner and Patterson (17) also standardized the test but they modified the original board by omitting the four geometrical pieces. They had two reasons for this. In the first place they felt the geometrical pieces differed radically in nature from the other pieces and were decidedly more difficult. In the second place they had another test which involved inserting two pieces together to form a triangle and preferred not to demand the same performance twice in their scale. They established norms on the basis of test results from 667 children in 2 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. It was felt that this combination would include a fair sampling of all grades of intelligence. Their table of distribution shows relatively little scattering. They conclude that it is obviously a test where ability to deal with the situation increases
fairly rapidly from age five to age ten. The only five children who failed to complete the test were aged eight or below. "The curve for the medians shows a steady and uniform decrease to age eleven, from which age onwards no 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 variation at any age (with the exception of age five) is not great."

In grading the performances on this test the first rank was given to the performance accomplished in the least amount of time, and so on.

**Goddard Formboard**

The so-called Goddard Form Board is practically identical with the Seguin Form Board and is called by either name by the various writers. It is especially adapted to the testing of very young children.

**Description**

Sylvester (27) who standardized the test on the basis of the performances of 1537 children, describes this board 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 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, 1½ x ⅛ 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."  This description differs only slightly from that of the Goddard Form Board as manufactured by Stoelting. There is great similarity in the age averages found by both Goddard and Sylvester.

Sylvester's method of procedure is: "The form board lies horizontally on a table, its lower edge even with the edge of the table 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 most workers have taken a record of three trials and called the shortest the child's form board index. The time limit is 5 minutes.

Many other writers have worked with this test. Goddard (29) used it in his testing at Ellis Island. Wallin (30) worked out norms not only for the best of three trials but also for the averages of three trials. Young presented another standardization.(31) However, he used quite a different board and his data cannot be compared with that of the other workers for this reason. Pintner and Patterson (17) embodied this test in their scale of performance tests but simply took over Sylvester's norms without additions or alterations. They note that in the table of the record of 1537 children from 5 to 14 the longest time record made by a 5 year old child, is 75 seconds, while the shortest, made by 14 individuals, is 9 seconds. The table shows a general upward trend and the medians for each age a constant and steady decrease.
The test was also standardized by the Bureau of Analysis and Investigation of the New York State Board of Charities (22). They conclude that "If it is desired to use the form-board as a special age test, it seems properly to be a test of 9 year mentality when done in 18 seconds." They give norms based upon the testing of 1049 children.

In this study the records used for this test are those of the shortest of three trials. Performances are ranked, entirely according to speed. No consideration of moves is made.

**Witmer Cylinder**

This test as used in this study was devised in 1915 and was adopted for use in the Psychological Clinic of the University of Pennsylvania. It was an outgrowth of the need felt by this Clinic of a test similar in nature to the Witmer Formboard but more difficult. The Montessori cylinders were brought into use and the test grew out of these.

**Description**

The Witmer Cylinder Test is a circular board having 9 series of recesses about its outer edge into which are fitted eighteen cylinders corresponding in size and order to those of the Montessori apparatus. However, there are no duplications of sizes, and the largest and smallest cylinders have been omitted. There is a central aperture into which the blocks may be thrown and jumbled. The largest cylinder is two and an eighth inches in each dimension. Starting from this point there are seven blocks of constant height but steadily decreasing diameters. The next six have the diameter of one inch but decrease in height to one inch.
Beyond these the blocks increase in both height and diameter up to the original block.

This test was standardized at the Psychological Clinic where it was first used and Paschal (32) reports the results. The test was given to 1722 children and 477 adults. Three trials were given and the subject was dismissed if he had not completed the first trial in 5 minutes. The shortest trial was taken as the best quantitative measure of performance ability for this test.

Discussion

Paschal's tables show that there is a continuous decrease in the time of the means, quintiles and medians with increasing age for each sex. He felt the investigation showed that the test possessed the requisite qualities of a performance test. It is applicable to a wide age range and was shown to give an increased performance with increasing age. Furthermore, the language factor was very largely eliminated so success was not dependent upon language ability.

The test affords a splendid chance for observation of the subjects' manner of handling concrete material. There are eighteen moves to be made, at least, and the cylinders are so alike that in all but the most careful and slow performances the number of moves is in excess of this. The examiner can note the subjects' plan as shown by the blocks picked up; his powers of perception as shown by the placings attempted; his speed in learning, as shown by his profit from errors.

Miller used this test in his study of fifty college
students (33). He says, "If intelligence be defined as the ability to solve what for the individual is a new problem, the test is primarily one of intelligence. This, however, is by no means the only ability involved. On the motor side may be observed the rate of discharge of energy, coordination, complexity of response, and in some cases endurance. The performance likewise displays some degree of analytic and distributed attention, observation, understanding and trainability when more than 1 trial is given."

In our use of the test we have followed Paschal's procedure. Performances have been ranked entirely according to speed.

**Dearborn Formboard #2**

**Description**

This test was first described by Dearborn, Anderson, and Christiansen (34). It consists of a board containing eight irregular depressions and three blocks which will fill each of the holes. The blocks are unequally sided trapeziums with their sides in the relation 2:3:4 with one extra side of 2, 3, and 4 respectively.

The board is placed before the subject who is asked to pick up the 3 blocks and fill the first depression. Three minutes is allowed for this. If the task is not accomplished in that time the examiner puts the blocks in before the subject and remarks how well they fit in. Then the child is asked to fill up the second hole with these same blocks. If no successful placement is made in two minutes the attempt is called a failure.
and the child proceeds to the third hole, and so on, through the eight depressions. Time is taken for each depression.

Discussion

This test was found to be much too difficult for the majority of younger children. Only rarely is a child under ten years of age able to fill all the depressions. The variability of performance was shown to be very great. Several adults, graduate students at Harvard, failed to complete one or more of the forms within the limits of the time set. It was thought that the test would give a measure of the subjects' ability to profit by practice and experience and Dearborn's tables show a gradually decreasing average time as the depressions are gone through successively, thus bearing out the authors' expectations. This decrease is subject to marked exceptions, however, which indicate that some of the depressions are proportionately more difficult, than others.

Further studies and standardization of this test have not been published either by the authors or others so far as the writer knows. The original article was accompanied by data in regard to the performances of 30 persons between the ages of 5 and 16, but these do not of course, provide even tentative norms. Workers who have used the test have been obliged to rely upon their own approximate judgment of each subject's performance rather than any more definite standards.

The board used in this study is a modification of the Dearborn Formboard #2 as described above. Our board had only four depressions. The time taken to fill each depression was recorded separately and then the total time for all four
holes computed. Performances are ranked according to speed only. This is an admittedly arbitrary method for attaining a usable criterion in the judgment of performances. It has no precedent in the work of others with this test so far as the writer knows. However it appears to be entirely in harmony with the purpose and methods of the test as outlined by the authors.

**Dearborn Formboard #3**

**Description**

Dearborn (34), Anderson, and Christiansen describe this test in the same article in which they describe Dearborn Formboard #2. The test was devised by Dearborn for use with defective children, with the purpose in mind of arranging a more difficult performance test than the Seguin form board, but one of essentially the same sort. Originally it was a block test. There were twenty blocks of geometrical forms which would fit into and exactly fill the nine depressions of a rather large formboard. The board was modified by Healy and Fernald and used by them at the Psychopathic Institute in Chicago. However Dearborn claims that the modification used by them is subject to certain criticisms which do not apply to the original board, or at least with equal force. He, therefore, prefers the original board.

The method of procedure is to present to the subject four distinct problems in relation to the filling of the board. In Problem A the board is filled with the exception of two spaces and there is one square piece left over. The problem is to prepare a place for the square by making as few as possi-
ble changes of the blocks as placed. The number of moves taken is recorded. Time is not a factor in the performance. In Problem B, places must be prepared for two squares, in Problem C for 4 blocks, in Problem D for 5 blocks. Each problem can be accomplished in a minimum number of moves. The performance is judged according to the number of moves beyond the minimum which were required for the subject to accomplish the task.

Discussion

So far as the writer knows, there has been no standardization of this test. In using it at Judge Baker Foundation a performance accomplished in the minimum number of moves was considered very good at any age level and the use of double the required moves or more was considered poor at any age. Performances lying within this range were roughly judged as good and fair.

In ranking the performances on this test the total number of moves used by each subject in accomplishing all 4 problems was computed and the grading was made entirely on the basis of efficiency.

Puzzle Box

Description

This is another of the Healy Farnald Tests (14). It is quite different from Healy Construction A and B in which the child has to analyze more or less complicated sets of spatial relationships. Here he analyzes a set of functional relationships of a contrivance, all the parts of which are open to view and involve no complex mechanical principles such as
the lever or pulley. The test consists of a box which the child must open as quickly as he can. There are seven steps necessary to accomplish this; three inner rings must be loosened from their confining posts, the staple at the back must be removed, also the ring from the hook at the front of the lock, and the hook itself from the lock and the lid must be raised. The puzzle box is made in such a way that the steps have to be accomplished in a certain order and a tool must be used for the removal of the three inner rings. A long hook resembling a shoe button hook is provided for this purpose.

In giving the test the box with the hook on top is placed lock side before the child and he is told that he may look all over the box inside and outside and any way he wishes to examine it to see if he can find a way to open it, and that he may do anything he thinks will help in opening it, or use anything he thinks will help. No further hint concerning the tool is given except in case the child tries to accomplish step one without the tool and leaves it to try something else since he can not succeed. He is then told that he may use the hook. He is then recorded as having to be told about the tool as opposed to the child who sees for himself the need of the tool and uses it without suggestion or asking permission to do so.

The record of data on each performance should include the amount of time spent in preliminary survey, each step or attempt, and the total time for accomplishment. Wrong attempts are recorded as errors. The original time limit was
15 minutes. After he has successfully opened the box the child is asked to close it. Time is recorded on this.

Discussion

Schmitt (21), who standardized this test, found that among the children under the fifth grade, from 70 to 100 percent were unable to do the test by a method superior to that of trial and error. The lack of the idea of using the tool is generally associated with the trial and error method of doing the test. She found that the total time decreased slightly as the school grade progressed and that for each grade the time for doing the test by the planned method averaged less than the time for the trial and error method. She thought the correlation of time with grade was due to increase of motor ability.

Halt and Brandenburg made an interesting study of the Puzzle Box (35). Their purpose was threefold: first, to standardize the time element; second to discover the nature of the ability it discloses; and third, to investigate the correlation of this ability with general intelligence. They gave the test to three groups; first, a miscellaneous group of 28 students, both sexes, who came to the psychology department at Purdue University for intelligence tests; second, to a group of 18 freshmen boys in a local high school who were simultaneously taking manual training; and third, to 29 senior engineers who were taking the Mental Measurements course at Purdue.

They arrived at the following conclusions:

1. Many individuals who have a reputation for excellence or failure in Mechanical lines show this by their degree of speed in opening the Puzzle Box.

2. A minute and a half to open and about twice that
to close, may be taken as approximate standards for the Puzzle Box.

3. According to one's performance in opening the Puzzle Box, so will he tend to do in shop work.

4. Accepting tentatively the definition of mechanical ability as an innate readiness of that and hand in analyzing and constructing processes and devices whose parts are related in a logical and more or less complicated way, and assuming that it is a reasonable expression of the qualities determining success in shop work, the time in opening the Puzzle Box is a fair test for mechanical ability.

5. Improvement in grade of shop work with reference to the time of opening the box, is fairly uniform according to percentile grouping.

6. General intelligence has little in common with speed in opening or closing the Puzzle Box.

7. Providing this test is a fairly accurate measure of mechanical ability, the man who excels in this ability should not be expected to possess necessarily a high or low level of general intelligence."

Knox Feature Profile

Description

This test was devised by Knox and Kempf (15). It consists of a large board in the form of a head. There are 4 small pieces which will fit into a square aperture to form the ear and three larger pieces which will form the face or profile. The test is placed before the subject with the large board in correct position nearest the child and the three pieces forming the profile are separated from each other by the four pieces forming the ear and are placed at the top of the head. The child is told to put this together as quickly as he can. The time limit is 5 minutes. Knox told his subjects that the figure was a head but Pintner and Paterson (17) who standardized the test, very explicitly warn against telling the child what the test represents (36). A
record of the time alone is kept.

Discussion

Knox called this his highest and most difficult performance test but considered it eminently fair because everyone has seen a human head, and, furthermore, he told his subjects that it was a head. The test demands the synthetic ability of seeing the parts of a whole and putting these together. Knox places it among a group of tests headed "At From Thirteen Years Onward" and his time limit is 10 minutes. Pintner and Patterson used a time limit of 5 minutes and found that 76 per cent of the thirteen-year-olds passed the test so they placed the test at thirteen years.

Stenquist Mechanical Assembly

Description

This test was devised by Stenquist (37) who does not claim that it measures all that is important in the field of mechanics, but does claim that it gives one definite indication of the relative ability of pupils in a standardized task of a mechanical nature. The test does not attempt to measure trade skill, but individual differences in native talent. It consists of a series of ten common mechanical objects arranged in order of difficulty in a special long box with ten compartments. Each mechanical model comes disassembled, packed into its proper compartment. The test consists in assembling as many of these models as possible in 30 minutes. It is not a mere speed test as this is ample time for the average person. The subject is given a screw-driver with which to work. The devices used in the test are 1. a linked chain, 2. Rubber hose Shut-off, 3. Paper clip, 4. Clothes Pin, 5. Bicycle
bell, 6. Push button, 7. Monkey wrench 8. Valve, 9. Lock, 10. Mouse trap. The child is not told what the parts will make or helped in any way. Each correct device is scored 10. Some credit is given partial performances. There are well established norms.

Discussion.

Stenquist has found that the most satisfactory criterion by which to judge what this test measures is shop and science teachers' ranks. Where the true abilities of pupils are known, correlations as high as .87 have been obtained between the assembly test and such ranks. This means that the test measures the same general qualities considered by shop and general science teachers in ranking pupils according to general mechanical ability. Stenquist does not claim to test general intelligence in this test. He points out the fact that mechanical ability and general intelligence, as measured by such tests as the Army Alpha, or Binet, are largely independent traits, the correlations rarely being higher than .4. He considers general abstract intelligence a poor basis for judging mechanical ability but notes that the two abilities often run along together.
TABULATION OF RESULTS.

In order that the individual record may be seen, the accompanying table of distribution gives the age, sex and I. Q. of each child as well as the result on each of the ten performance tests given.

(II)

DISCUSSION OF RESULTS.

As the distribution table shows, the records available on some of the tests are very few in number; in fact only on Construction A and B, Dearborn #3 and Stenquist, are they large enough to give a fairly satisfactory basis for comparisons. However, it was thought worth while to study the records on all of the tests with at least one purpose in mind. This purpose was to find out the relationship of Construction A to the remaining 9 tests. At Judge Baker Foundation this test has been used as an indicator of a child's ability to work with concrete material. If he fails on it, usually no other test of the performance type is given. If he succeeds, he is given Construction B. If he does both of them very well, he is often given the remaining tests, especially the more difficult ones. No scientific basis for this procedure is known. None of the research done in connection with the test has covered this particular point so far as the writer knows.

Therefore, to ascertain the relation of Construction A. to the other tests a series of correlations between the ratings on it and on each of the other 9 tests was calculated. In each case the coefficient of correlation was obtained by the rank difference method. The correlations are as follows:
Table II - continued. (2)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>B</td>
<td>83</td>
<td>94</td>
<td>177</td>
<td>44</td>
<td>31</td>
<td>44</td>
<td>254</td>
<td>56</td>
<td>56</td>
<td>254</td>
</tr>
<tr>
<td>14</td>
<td>B</td>
<td>110</td>
<td>105</td>
<td>215</td>
<td>39</td>
<td>31</td>
<td>39</td>
<td>178</td>
<td>39</td>
<td>30</td>
<td>178</td>
</tr>
<tr>
<td>15</td>
<td>B</td>
<td>88</td>
<td>83</td>
<td>171</td>
<td>54</td>
<td>44</td>
<td>44</td>
<td>223</td>
<td>54</td>
<td>54</td>
<td>223</td>
</tr>
<tr>
<td>18</td>
<td>B</td>
<td>81</td>
<td>78</td>
<td>159</td>
<td>47</td>
<td>36</td>
<td>36</td>
<td>183</td>
<td>47</td>
<td>47</td>
<td>183</td>
</tr>
<tr>
<td>19</td>
<td>B</td>
<td>89</td>
<td>83</td>
<td>172</td>
<td>44</td>
<td>33</td>
<td>33</td>
<td>178</td>
<td>44</td>
<td>44</td>
<td>178</td>
</tr>
<tr>
<td>20</td>
<td>B</td>
<td>86</td>
<td>84</td>
<td>170</td>
<td>47</td>
<td>34</td>
<td>34</td>
<td>180</td>
<td>47</td>
<td>47</td>
<td>180</td>
</tr>
<tr>
<td>21</td>
<td>B</td>
<td>100</td>
<td>94</td>
<td>194</td>
<td>47</td>
<td>36</td>
<td>36</td>
<td>207</td>
<td>47</td>
<td>47</td>
<td>207</td>
</tr>
<tr>
<td>22</td>
<td>B</td>
<td>110</td>
<td>105</td>
<td>215</td>
<td>47</td>
<td>39</td>
<td>39</td>
<td>188</td>
<td>47</td>
<td>47</td>
<td>188</td>
</tr>
<tr>
<td>23</td>
<td>B</td>
<td>88</td>
<td>83</td>
<td>171</td>
<td>44</td>
<td>31</td>
<td>31</td>
<td>178</td>
<td>44</td>
<td>44</td>
<td>178</td>
</tr>
<tr>
<td>24</td>
<td>B</td>
<td>89</td>
<td>83</td>
<td>172</td>
<td>44</td>
<td>33</td>
<td>33</td>
<td>178</td>
<td>44</td>
<td>44</td>
<td>178</td>
</tr>
<tr>
<td>25</td>
<td>B</td>
<td>86</td>
<td>84</td>
<td>170</td>
<td>47</td>
<td>34</td>
<td>34</td>
<td>180</td>
<td>47</td>
<td>47</td>
<td>180</td>
</tr>
<tr>
<td>26</td>
<td>B</td>
<td>100</td>
<td>94</td>
<td>194</td>
<td>47</td>
<td>36</td>
<td>36</td>
<td>207</td>
<td>47</td>
<td>47</td>
<td>207</td>
</tr>
<tr>
<td>27</td>
<td>B</td>
<td>110</td>
<td>105</td>
<td>215</td>
<td>47</td>
<td>39</td>
<td>39</td>
<td>188</td>
<td>47</td>
<td>47</td>
<td>188</td>
</tr>
<tr>
<td>28</td>
<td>B</td>
<td>88</td>
<td>83</td>
<td>171</td>
<td>44</td>
<td>31</td>
<td>31</td>
<td>178</td>
<td>44</td>
<td>44</td>
<td>178</td>
</tr>
<tr>
<td>29</td>
<td>B</td>
<td>89</td>
<td>83</td>
<td>172</td>
<td>44</td>
<td>33</td>
<td>33</td>
<td>178</td>
<td>44</td>
<td>44</td>
<td>178</td>
</tr>
<tr>
<td>30</td>
<td>B</td>
<td>86</td>
<td>84</td>
<td>170</td>
<td>47</td>
<td>34</td>
<td>34</td>
<td>180</td>
<td>47</td>
<td>47</td>
<td>180</td>
</tr>
</tbody>
</table>
Table II - continued
Table II - Continued (4)

<p>| Age Group | Sex | Full Height | Total | Standing | Sitting | Feature | Weight | Race | Location | - 36 2 - |
|-----------|-----|-------------|-------|----------|---------|---------|--------|------|----------|---|---|
| 16-18 B   | 110 | 148        | 50    | F       |         |         |        |      |           |   |   |
| 16-18 B   | 117 | F          | 89    |         |         |         |        |      |           |   |   |
| 17-19 B   | 117 | 115        | 150   | F       | 143     |         |        |      |           |   |   |
| 18-20 B   | 116 | F          | 83    |         |         |         |        |      |           |   |   |
| 19-21 B   | 110 | 12-32      | 28    | 130     | 86      | 1-15    |         |      |           |   |   |
| 20-22 B   | 117 | 24-125     | 68    | 155     | 54      | 42      |         |      |           |   |   |
| 21-23 B   | 117 | 117-117    | 94    | 170     | 94      |         |         |      |           |   |   |
| 22-24 B   | 116 | 5-15       | 24    | 78      |         |         |         |      |           |   |   |
| 23-25 B   | 115 | 115-111    | 95    | F       | 95      |         |         |      |           |   |   |
| 24-26 B   | 115 | 111-117    | 92    | 170     | 11     | 23      |         |      |           |   |   |
| 25-27 B   | 113 | 5-10       | 74    | 66      | 34      | 120     | 70     | 52    |           |   |   |
| 26-28 B   | 93  | 5-12       | 38    |         |         |         |         |      |           |   |   |
| 27-29 B   | 160 | 10-19      | 31    |         |         |         |         |      |           |   |   |
| 28-30 B   | 103 | 12-32      | 84    | 34      | 49      | 20      |         |      |           |   |   |</p>
<table>
<thead>
<tr>
<th>Combination</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construction A with Construction B</td>
<td>.274</td>
</tr>
<tr>
<td>Construction A with Mare and Foal Picture Board</td>
<td>.58</td>
</tr>
<tr>
<td>Construction A with Goddard Form Board</td>
<td>.42</td>
</tr>
<tr>
<td>Construction A with Witmer Cylinders</td>
<td>.22</td>
</tr>
<tr>
<td>Construction A with Dearborn Form Board #2</td>
<td>.28</td>
</tr>
<tr>
<td>Construction A with Dearborn Form Board #3</td>
<td>.226</td>
</tr>
<tr>
<td>Construction A with Healy Puzzle Box</td>
<td>.22</td>
</tr>
<tr>
<td>Construction A with Stenquist Assembly</td>
<td>.355</td>
</tr>
<tr>
<td>Construction A with Knox Feature Profile</td>
<td>.30</td>
</tr>
</tbody>
</table>

A mere inspection of the coefficients listed above will show that while all of the correlations are positive, not one can be considered significant. In general, it may be stated that coefficients between $\pm .30$ and $\pm .75$ show that the same factors are operative in the two series to some degree, but the correlation can hardly be regarded as significant unless a coefficient greater than $\pm .75$ is found. An immediate conclusion can be drawn, therefore, either to the effect that the material employed as a basis for the correlations was faulty and so brought nothing significant, or that the performances on the ten tests do not involve the same abilities.

In reconsidering the test results which were used as the basis for the correlations, of course we must grant that the total number of performances of Mare and Foal Picture Board, Goddard Form Board, Witmer Cylinders, Dearborn Formboard #2, and Knox Feature Profile, each of which totals under fifty, is truly too small to give valid results. However, the same objection cannot be made in the case of Construction A, Construction B, Dearborn
Formboard #3, Healy Puzzle Box or Stenquist Mechanical Assembly, therefore the correlations of these tests are worthy of consideration.

It is interesting first of all to note that the highest correlation of the group is between Construction A and the Mare and Foal Picture Board. If such small numbers bring such a high correlation it is quite possible that larger numbers would show an even closer relationship. At any rate, it seems apparent that these two tests do require the same capacities to a considerable degree. The same conclusion can be drawn in regard to Construction A and Goddard Form Board.

Of the remaining tests, Stenquist Mechanical Assembly correlates most highly with Construction A. However, the coefficient of correlation +.355, is lower than the coefficient of approximately +.4 which Stenquist reports between general intelligence and the Mechanical Assembly test and which he considers significant as showing general intelligence and mechanical ability are largely independent traits. We must, therefore, claim no more than this for Construction A and Mechanical Assembly. In short, on the basis of our coefficients, we can claim nothing further than this for any of the tests as compared with Construction A. The tests do not involve diametrically opposed capacities for the coefficients are all positive. But the capacities required for a successful performance on Construction A are certainly not identically the same as those required for any one of the other tests.

In conclusion we may say, then, that so far as the correlations show, there are no valid grounds for using Construc-
tion A as an indicator. It may not be presumed that because a child does well on Construction A, he will do well on any of the other tests, or vice-versa. Whether Construction B, or any other test, would serve as a better indicator, is not, of course, shown by our figures. Further experimentation might well be done in search of more information concerning the interrelations of performance tests.

SUMMARY

Performance tests are a late outgrowth in the history of mental tests which extends back over the past thirty years. They were devised to meet the needs of clinical workers among the foreign and handicapped, and they avoid the limitations of the early physical tests as well as the later age-level tests involving language, by simply presenting to the child a concrete situation in which he has to put something together with his hands. This study deals with the performances of 289 Boston children, tested at Judge Baker Foundation, on ten performance tests; namely Healy Construction A, Healy Construction B, Mare and Foal Picture Board, Goddard Form Board, Witmer Cylinders, Dearborn Formboard #2, Dearborn Formboard #3, Healy Puzzle Box, Knox Feature Profile, Stenquist Mechanical Assembly. The aim was to test out the value of a performance on Construction A as an indicator of probable performances on each of the other nine tests. The correlations between Construction A and the other tests were:

Construction B  .274
Mare and Foal Picture Board  .58
Goddard Form Board  .42
Witmer Cylinders  .22
Dearborn Formboard #2  .28
Dearborn Formboard #3  .226
Healy Puzzle Box  .22
Stenquist Assembly  .355
Knox Feature Profile  .30

The obvious conclusion was that, to a certain extent, all of these tests do require the same capacities as does Construction A, but that ability to do well on one is considerably independent of the ability to do well on another. Furthermore, the correlations give no grounds for the use of the performance on Construction A as an indicator of the performance on any one of the other nine tests.
BIBLIOGRAPHY.


Nouvelles recherches sur la mesure du niveau intellectuel chez les enfants d'école. ibid vol. XVII P. 145.

L'intelligence des imbeciles ibid vol. XV. p.1.


32. Paschal, F. C. A Report on the Standardization of the Witmer Cylinder Test. The Psychological Clinic vol. XII No. 2, April 15, 1918, pp. 54-60.


Some tests of General Mechanical Ability - ibid.


