A Thesis

submitted to The Committee on Graduate Study and Degrees in partial fulfillment of the requirements for the degree of Master of Arts, University of Buffalo

by

Olive P. Lester

January 1926

A STUDY OF
THE RELATIVE VALUE OF PERFORMANCE TESTS IN SECTIONING FIRST-GRADE CHILDREN.

SECTION I

INTRODUCTORY

This study is an outgrowth of an interest in the question whether the standard psychological tests most frequently employed in the school system, i.e., the Stanford Revision of the Binet Test, the Pressey Primary Test, and the Otis Primary Group Intelligence Scale, are entirely adequate in sectioning first-grade children, especially in the sectioning of the children of recent immigrants to this country.

An historical survey is given which attempts to trace, or rather summarize, the work up to the present time that has been done in the field of mental testing.

The new data presented here have been collected from two sections of first-grade children, one a bright group of children of American parentage, the other a group of children of foreign parentage, in an attempt to ascertain two things:

1. To what extent various performance tests measure capacities and abilities of young children as compared with the more verbal type of test, such as the Binet Test.

2. If certain of these performance tests can be advantageously combined with the Stanford Revision of the Binet Test so as to classify children more adequately than they are now being classified in the public school system.

I might say that in carrying on this investigation I have been primarily interested in the group of children of foreign parentage, though I believe it was of value to be able to try out a similar set of performance tests on the group of children of American parentage in order to find out if results obtained in the foreign group were substantiated by or applicable in the American group, or if the results were at variance with this group.

I realize on account of the comparatively small number of cases, sixty-three (63) in all, no final conclusions or generalizations can be drawn, yet I hope that the study will be of some value by presenting the results obtained with performance tests in these particular groups and by proving suggestive for further and more extensive investigations in this particular field.

SECTION II

HISTORICAL

The first extensive and practical test to measure mentality dates back to the work of Alfred Binet and Thomas Simon, who in the first decade of the twentieth century published their scale for the measurement of intelligence. However, in a very real sense attempts had been made earlier to measure innate abilities of individuals, but these attempts were not directed so much toward the measurement of intelligence as a whole as toward the development of tests for measuring various capacities of human beings. Sir Francis Galton, one of the most versatile of British scientists, was the pioneer worker in this field of individual testing.[1] In 1883 he published an elaborate account of individual and racial differences.[2] The object of his tests and measurements he described as follows:[3] "It is to obtain a general knowledge of the capacities of a man by sinking shafts, as it were, at a few critical points. In

---

1. Much of Galton's data was gathered by rather unscientific methods such as mere casual observation, but still we find evidences of the careful investigator in the elaborate questionnaire he used for studying differences in mental imagery, in his experiments to test abilities to discriminate weights and in the devising of a particular kind of whistle to test sensitivity to pitch.
2. Inquiries into Human Faculty.
3. Herbert Woodrow, Brightness and Dullness in Children, p.19.

order to ascertain the best points for the purpose the set of measures should be compared with an independent estimate of the man's powers. We may thus learn which of the measurements are the most instructive."

The work of Galton was followed by that of Kraeplin in Germany, who revised some of the tests with the idea of using them in studies of the differences between the mentally normal and the insane.

In this country the work of Galton in the direction of mental testing was largely made known and developed by James McKeen Cattell. Cattell's service in the field of mental testing is well stated by his most distinguished pupil, Professor E. L. Thorndike, who says, "Cattell refined Galton's methods and won recognition for such measurement of individuals as a standard division of psychology, and of psychological training in universities, beginning at Pennsylvania the systematic inventory of mental traits which became such an important feature of the Columbia laboratory, and which was for so many of us an introduction to the whole topic of individual psychology."[1]

---

1. Columbia University Contributions to Philosophy and Psychology, Vol. XXII, No. 4 (1914) p. 92.

Cattell devised a set of tests which for a number of years were given to freshmen entering Columbia University. These were designed to measure such capacities as the following: strength of grip; sensory discrimination by the skin, indicated by the distance that must separate two compass points in order that they may be felt as two; the sense of pain, measured by the amount of pressure on the ball of the hand required to produce a painful sensation; the ability to discriminate weight; reaction time; visual space perception, determined by the ability to bisect a 50 centimeter line; time estimation, shown in the ability to reproduce an interval of ten seconds by taps made on the table; and memory, manifested by the number of letters that can be repeated correctly after one hearing.[1]

It can be seen that these earlier attempts at mental testing concerned themselves chiefly with what may be designated as the sensory and motor phases of ability, and gave scant notice to the more elaborate phases of intelligence. It is interesting in view of their later work to note that Binet and Henri,[2] the French psychologists, criticised the work of the American investigators on the ground that the tests employed by them measured

---

1. "Mental Tests and Measurements" in Mind, Vol. 15, pp. 373-380.
2. "La Psychologie Individuelle" in L'année Psychologique, Vol. 2, 1895, p. 426.

mental processes which were too simple -- which did not sufficiently involve the "superior mental faculties".

More in agreement with the ideas of Binet and Henri just expressed was the work done at this time in Germany by Ebbinghaus. According to Spearman[1] "the first great attempt to grapple adequately with the problem of the nature of intelligence was made by Ebbinghaus in 1897 in a still far from sufficiently appreciated paper".[2] Ebbinghaus in studying and analyzing the performances and behavior, by which he believed people actually distinguished themselves for intelligence in every day life, arrived at the conclusion that such activity consisted in "bringing together a multitude of independent concomitant impressions into a unitary meaningful or in a way purposive whole." Accordingly for him the essential nature of intelligence could be summed up in the term "combination". With this idea as to the nature of intelligence Ebbinghaus set himself to the task of devising a test that would measure this ability or capacity, and as a result we have his famous completion test, in which certain words or syllables of a paragraph are omitted and have to be filled in by the subject taking the test. This is the first test that had

---

1. The Nature of Intelligence, p. 3.
2. Zeit. f.Psychologie, XIII, p. 401.

as yet been devised to test a large group of individuals at the same time and that attempted to measure the higher mental abilities and capacities.[1]

But the successful means of measuring intelligence, as we have already intimated, was accomplished by the French psychologist, Binet, who collaborated with the French physician, Simon, in establishing a scale by which intelligence might be submitted to measurement.

We must not for a moment suppose that this accomplishment, regarded by some psychologists as being the most brilliant achievement of modern psychology, had not been preceded by years of painstaking and careful observation and experiment.[2]

---

1. Spearman makes another rather interesting comment in referring to this test: "This simple procedure -- referring to the completion test -- which to the uninitiated might seem to be only a childish puzzle, did actually attain to an extraordinary degree of success. Not all the restless invention of psychologists in every quarter of the globe has to this day been able to construct any other single test which, on the whole, yields higher correlations with intelligence as evaluated in other ways." - From The Nature of Intelligence, p. 4.

2. In 1895 Binet founded L'année Psychologique which proved to be the main avenue for the publication of his future works, and it is largely by a perusal of this that we may get a glimpse of his experimentation and interest in intelligence testing which culminated in the first set of tests. As early as 1896 we find him interested along with Henri in an article discussing the field of individual psychology. In 1898 he contributed an article with the significant title "Measurement in Individual Psychology". In this article he mentioned for the first (cont'd)

In L'année Psychologique for 1905 there appeared for the first time the idea of a scale of intelligence -- "une échelle métrique d'intelligence." The thirty tests (sometimes referred to as the 1905 Scale) are arranged in order of difficulty, and are not grouped according to age. The tests are scored with whole, half, or no credit, and no attempt is made to ascertain mental age. Apparently the authors had not as yet arrived at the concept of mental age.

---

time, I think, different tests which later on proved so useful in the scale, namely, drawing a square from memory; suggestibility to length of lines; memory for numbers; rearrangement of dissected sentences; answers to questions involving moral judgments; comprehension of an abstract passage; folding paper test. - Rudolf Pintner, Intelligence Testing, p.27.

Looking back it would seem as if nothing could have been more simple and natural than for Binet to have directly proceeded toward the scale of tests, but the way he took was long and devious, and before arriving at the methods foreshadowed in the above article, we find him trying out and experimenting with many other possible means of measurement.

In the following year he published a study on "Attention and Adaptation", his most important work prior to the scale proper. In this, Binet attempted to study voluntary attention in relation to intelligence by means of giving tests to two groups of children -- one bright, one dull.

In the next year Binet, as Pintner phrases it, is "again off on one of his numerous tangents" in which he compared head measurements of intelligent and unintelligent pupils. This work led to still further writing on Binet's part throughout the next year concerning cephalometric research.

In 1903 his book entitled L'étude Expérimental de l'intelligence appeared. It is in this book that he takes the attitude that intelligence is equivalent to the higher mental processes. In 1904 there appeared another of Binet's studies on the question of the relation of handwriting to intelligence.

This summary, though far from complete, serves to show that the measurement of intelligence had been one of (cont'd)

Perhaps Binet would have experimented for a still longer period on the problem of measuring general intelligence before arriving at a scale similar to the above had not his research been influenced by practical necessity. In 1904 the French Minister had appointed Binet as a member of a commission to organize classes for subnormal children in the public schools of Paris.[1]

In 1908, as a result of further investigation, Binet and Simon published their first revision of the Scale, which is referred to as the 1908 Scale. The important thing to note is that the tests are now arranged according to their appropriate ages, and that in connection with the scale the authors have worked out the idea of mental age.

The tests are arranged in order of their increasing difficulty. To find out which tests are suitable for each age Binet and Simon tested supposedly normal

---

Binet's main thoughts and interests for the previous ten years; almost every study had been concerned with the problem of measuring intelligence, and these attempts had ranged, as we have seen, from tests proper to the study of head measurements and graphology.

1. The question arose as to how these subnormal children could be distinguished and classified. Up to this time rather hazy notions had surrounded the problem and it was important that the selection of children should be accomplished by more objective methods than had previously been employed. It was under this incentive and practical difficulty that Binet and Simon worked out and published their measuring scale of intelligence.

children at each age (some two hundred in all) in the schools of Paris. If a test were passed by 60% to 90% of children at one of these particular age levels it was considered standard for that age. For example, if it were found that a certain test was passed by only a very small portion of the six-year-olds, and seven-year-olds, and that it was passed by 60% to 90% of the eight-year-olds, the test would then be considered by Binet a test of eight-year intelligence. In like manner tests passed by 60% to 90% of nine-year-olds were considered tests of nine-year intelligence, and so on. By trying out many different tests in this way Binet was successful in obtaining five tests for each age from three to ten years (excepting age four which has only four tests) five for age twelve, five for fifteen, and five for adults, making fifty-four (54) tests in all.[1]

After standardizing the tests for particular ages, though the method and number of children tested, as Pintner remarks, may seem unsatisfactory to us now, it was possible to obtain an estimate of intelligence in terms of mental age and to classify children on this basis. The mental age of any child is expressed by the age that he reaches in the graded series of tests.[2]

---

1. Lewis M. Terman, The Measurement of Intelligence, p. 37.
2. Let us suppose that the child being tested is seven years of age. If he goes as far in the tests as normal (cont'd)

We find Woodrow relating how Esquirol in 1828, Duncan and Millard in 1866, and Down in 1887, plainly thought in terms of mental ages when they made comparisons between feeble-minded and normal children on the age basis, without, however, using any tests, but it remained for Binet to crystallize this concept of mental age and on the basis of his experimentation and observation to change it from a useless concept to one of vital importance for all psychologists and educators.

In 1911 as a result of further study the 1911 Scale was published. Undoubtedly there would have been many further revisions of this measuring scale, for no one realized more than Binet himself the complexity of the problem he had undertaken and the many shortcomings of the scale he had devised for submitting intelligence to measurement. But this work was halted by his untimely death in the same year in which the 1911 Scale was published.

Some authorities have contended that the success of the Binet-Simon Tests lies in the fact that they attempted to test the higher and more complex mental functions

---

seven year old children go, then we say that the child has a mental age of seven years, which in this case is normal (the child being seven years of age); if he goes on as far as children eight years old go, then he would have a mental age of eight years and would be considered above average in intelligence. On the other hand, if in the test he could (cont'd)

instead of the simpler and more elementary ones. There is undoubtedly some truth in this, but the basis for Binet's success, I think, lies in the fact that after years of experimenting with separate tests he came upon the idea of combining a large number of these tests in order to measure intelligence. In other words, he realized that intelligence, whatever it is, was too complex a thing, that it had too many aspects for any one type of test to display it adequately, and that hence a combination of tests would serve better.[1]

Although the success of the Binet Tests is due primarily, as we have said, to the arrangement of a number of tests in combination, their rapid and widespread popularity can be attributed to the utilization of age standards or norms, i.e., to the arrangement of the tests by ages.

Perhaps an undue amount of space has been given over to the work of Binet and the descriptions of his scale, but it seemed wise to do this and then treat in

---

go only as far as five year old children, he would have a mental age of five years and be considered definitely low in intelligence.

1. Binet's belief that intelligence has many different aspects is clearly brought out in the type of test that he employed in order to display it. "Some tests bring out differences of memory; others, differences in power to reason, ability to compare, power of comprehension, time orientation, facility in the use of number (cont'd)

a less intensive way the revisions and new scales that have appeared since 1908, for these are practically outgrowths of the experimentation of Binet and are based on his fundamental work.

Henry H. Goddard was the first in this country to introduce the tests and to use them extensively. He adapted the scale to American conditions, making as few changes as possible. His work with the scale made him doubt somewhat if all of the tests were properly placed. This led him to standardize the tests on a group of two thousand children, and his subsequent revision of the Binet scale was published in 1910[1] and served as a standard until the publication of Terman's more complete revision.

Kuhlman[2] also revised Binet's tests in his work with sub-normal children at Faribault, Minnesota, and contributed a standardization of tests for children below

---

concepts, power to combine ideas into a meaningful whole, the maturity of apperception, wealth of ideas, knowledge of common objects, etc." - Terman, Measurement of Intelligence, p. 36.

1. "Binet-Simon Measuring Scale for Intelligence." The Training School, January 1910.

2. "A revision of the Binet-Simon system for measuring the intelligence of children." Journal of Psycho-Asthenics, monograph Supp., Volume I, No. 1, 1912.

The Measurement of Mental Development. (1917)

the age of three. In 1915 Yerkes, Bridges and Hardwick[1] published a point-scale revision of Binet's tests.

In 1912 Terman and Childs[2] published a tentative revision and extension of the Binet 1908 Scale. The results of this study did not prove very satisfactory and accordingly a new investigation was undertaken and carried out independently by Terman[3] at Leland Stanford University.

The outcome of this later investigation is the scale of tests, so often referred to as the Stanford Revision

---

1. A Point-Scale for Measuring Mental Ability. (1915)
2. "A Tentative Revision and Extension of the Binet-Simon Measuring Scale of Intelligence." Journal of Educational Psychology. Vol. 3, February, March, April and May, 1912.
3. Terman himself describes the work he attempted in his revision of the Binet Tests thus: "As far as possible the original Binet Tests have been retained in the form in which they were used by their author, although in a number of cases it has seemed advisable to introduce alterations either in procedure or scoring. While it is not claimed that the revision here offered is satisfactory in every respect, the authors believe that it possesses a number of distinct advantages over other versions of the Binet Scale. Among these advantages are the following:
    1. Correction of the too-great ease of the original scale at its lower end and its too-great difficulty at the upper end.
    2. The revision not only contains a much larger number of tests than any other series but also brings into operation a much greater variety of mental functions. This is especially true for the upper part of the scale.
    3. It is believed that the detailed directions set forth in the companion volume for giving and scoring the tests should tend materially to promote uniformity of procedure."

of Binet Tests.[1] This scale, based on a study of approximately twenty-three hundred subjects, including seventeen hundred children, contains ninety tests[2] as contrasted with fifty-four in the original Binet series, and is the form of the Binet Test, so to speak, used so extensively in our school system. No other scale has had such a thorough and extensive foundation.

The other important contributions of Terman, along with the revision of the scale, is the introduction of the Intelligence Quotient (I.Q.) as a means of indicating an individual's mentality.

Binet expressed the amount of a child's intelligence in terms of mental age. Terman considers the mental age of a child meaningless if considered apart from chronological age. He believes that it is only the ratio of mental age to chronological age (i.e., the I.Q.) which has significance.

This method of expressing a child's mentality has certain advantages, but it also involves dangers that

---

1. Lewis M. Terman, The Measurement of Intelligence. (1916)
   Lewis M. Terman, The Stanford Revision of the Binet-Simon Scale. (1917)
2. Of these ninety tests there are six for each age level from three to ten, eight for the age of twelve and six for the age of fourteen. There are also six tests for average adults and six for superior adults. A number of alternate tests for the various ages is also provided.

must be guarded against when the I.Q.'s are used for administrative purposes. The chief value of the I.Q. lies in the fact that it apparently expresses the child's innate intelligence in a more or less arbitrary way. It is intended to indicate his actual mentality irrespective of his age. According to Terman the I.Q. is the basis for prediction in regard to the child's later mental development. The possibility of such prediction comes from the fact that he has found in the large majority of cases that the I.Q. remains fairly constant. This would mean that if a child of five chronologically was mentally four years old he would have an I.Q. of 80; at ten years chronologically he should have a mental age of eight and still an I.Q. of 80. This contention of Terman's seems on the whole to be substantiated by the facts.

The latest revision of the Binet Scale has been constructed by Herring.[1] This is a Point Scale. It correlates very high with the Stanford Revision of the Binet Scale, and as it is made up of similar though different tests it can very well be used for re-testing purposes, thus eliminating the practice effect that always results when the same scale is given the second time.

---

1. The Herring Revision of the Binet-Simon Tests. (1922)

The finding of a means for testing intelligence was such an outstanding accomplishment that for a time at least the proponents of the new method were too oblivious to possible limitations, and blind to some of the defects of the scales. One of the most outstanding limitations of the Binet scale proper, and its revisions, is that the tests are too verbal in their nature, they rely too much on words, and too little on activities, i.e., they appeal too much to abstract intelligence.

Another type of intelligence test was developed, therefore, which attempted to meet this difficulty by showing the functioning of intelligence without the necessity of accuracy in the use of language. This is the performance test. The performance test does not represent, as does the Binet Test, the work of any single individual, nor does it designate a specific group of tests. It is rather the name of a type of test or a method of procedure in testing. As the name indicates, tests of this sort emphasize the doing in a more objective sense, usually with the hands, certain tasks in response to given directions. This type of test is not only of immense value in supplementing the more verbal type of test, but it is practically indispensable in the testing of foreign children where language difficulties render the verbal test entirely inadequate. More will be said later concerning the use of the performance

tests in classifying foreign-speaking children. In fact, our whole study is concerned with these performance tests and their use in classifying children, especially foreign children.

The names of Healy, Fernald, Knox and Porteus stand out conspicuously in the development of the performance test.

Healy in his work with delinquents found that the Binet Test did not succeed in evaluating the capabilities of the individuals he was concerned with, and as a result he and Fernald[1] were stimulated into working out and developing various form boards[2] and puzzle boards to supplement, and to be used independently of the Binet Test.

Knox[3] in his work with immigrants at Ellis Island developed the cube test, which has proved very useful in classifying foreign-speaking individuals. The particular value of the test, as might be imagined from the

---

1. The Psychological Review Monograph Supplement No. 83, July 1915.
2. Most of the form boards now included among the performance tests are modifications of the form board originated by Seguin. We may also consider the picture puzzle test, now so common, as a variation of the form board.
3. Journal of the American Medical Association, 1914, Vol. LXII, pp. 741-747.

circumstances and surroundings under which it was developed, lies in the fact that the understanding of the problem presented and subsequent success are so little dependent on language.

Porteus[1] made a very valuable contribution to the field of performance testing with the maze. This test was used extensively twenty years ago, in the earlier days of animal psychology, and it is to Porteus that is due the credit for adapting it to human beings. The maze test when used with human beings is a paper and pencil test of the performance type, the maze being printed on a sheet of paper, and the person tested being required to find his way through the maze.

At present all we can do is to site the outstanding early contributions to the field of performance testing. The more recent contributions of this type are many[2] and, combined with the more specialized types of performance tests designed for testing the blind and deaf, are too numerous to be outlined here.

The group test is the most recent development

---

1. "Mental Tests for Feebleminded; A New Series." Journal of Psycho-Asthenics, Vol. XIX, No. 4, pp. 200-213.
2. Pintner and Paterson in 1917 constructed a scale of performance tests referred to as the Pintner-Paterson Performance Scale which represents a collection of fifteen of the more common tests of this type.

in the field of mental testing. Heretofore the necessity of testing large numbers of children for the purpose of classification and teaching was hardly recognized. This need was first felt not in the school but in the army, and the success of the group test there resulted in the construction of a number of group tests, mostly of the verbal type for use in the various school systems. The advantage of the group test, of course, lies in the fact that it can be administered to a large group of individuals at the same time, which means an enormous saving when we consider the amount of time that would necessarily be consumed in determining by individual tests the intelligence of large groups of school children. But, nevertheless, there are times when the individual tests are desirable and really indispensable in the further classification of individuals and even entire groups.

There are a great number of group tests available at the present time. Pintner cites twenty-seven and there are undoubtedly more.[1]

I have just outlined the development of mental testing from its inception, with the work of Galton in

---

1. Some of the group tests most commonly used in the classification of children are the following: Dearborn I and II; The Haggerty Delta I and II; The National Intelligence Tests; The Otis Group Tests; The Pressey Group Tests; The Whipple Group Test.

individual testing, and its subsequent growth up through the recent large-scale testing of intelligence in the army and later in the public school systems of this country. Following this development there has been a movement to question somewhat the claims made for the tests by their various proponents, and the development of the more specialized attempts to observe and isolate factors which affect the mental growth of children.[1] (I am speaking here of the verbal types of tests and not the performance tests.)

---

1. This phase is relatively unrelated to our problem but I think reference to some of the contributions that attempt to explain causes affecting mental development are important enough to mention here. Woodrow with his concept of anatomical age and its important relation to mental age, together with his recommendation of making a correction when using the formula $I.Q. = \frac{M.A.}{C.A.}$ adds immensely to our understanding of the effect of physiological development on mental development, and offers a better means of classifying children than is possible when we consider only chronological age.

Along the same line is the Harvard study now being conducted under the direction of Professor Walter F. Dearborn. This study is to cover a period of twelve years, and hopes to trace the mental and physical development of several hundred children throughout this entire period.

A great number of studies have been conducted both by school systems and cities, and the Federal Government itself, to determine the effect of malnutrition on both physical and mental development. The literature on this subject is too vast to be listed here. Following the same line of attack, that is, condition of physical health and its relation to mental development, is the work of E.K. Strong on the "Effects of Hookworm Disease on the Mental and Physical Development of Children" published under the above title in Publication 3 of the International Health Commission Reports. Much work has been carried on by various investigators concerning the effect of (cont'd)

The early opinion of the individuals concerned in the development of the tests was that innate intelligence, independent of any acquired learning or experience, was being measured. Binet was more uncertain of this, and less dogmatic in his claims of what the tests actually measure, than were some of the psychologists that followed him. Terman, I think, believes that he is actually able by means of the Stanford Revision to test out and measure native intelligence, not school knowledge or home-training. Stern thinks that "the tests actually reach and discover the general developmental conditions of intelligence, and not mere fragments of knowledge and attainments acquired by chance". I am inclined to think that with the passage of time, a slightly different attitude is being developed regarding the tests, and what they measure. It seems to me that it is impossible to submit innate intelligence to measurement. All we can do is to measure acquired intelli-

---

physical defects, more particularly tonsils and adenoids, on mental development in children.

The whole study of the pathological and unstable child represents another branching off, and an attempt at explaining differences in mental development of children.

Besides many studies similar to those indicated above are those which seek an explanation of differences in mental development from the"environmental" side. The article by Bridges and Coler in the Psychological Review for January 1917, entitled "Intelligence and Social Status", summarizes very well these attempts and some of the findings. A small pamphlet which seems to me to present the crux of the situation for the environmentalists is one, strange to say, published by Lewis M. Terman entitled "An Experiment in Infant Education."

gence, but we can infer from differences in this acquired intelligence, differences in native endowment. If you submit to a test a group of individuals who have had common experiences, i.e., environment or nurture having been the same for all, and then find that they differ in acquired intelligence, we have grounds for believing that they differ in innate intelligence or ability, Colvin would go even farther than to say that the group must have common experiences before we can infer native intelligence from differences in acquired intelligence. He would say that there must be common interests as well as common experiences within the group. He makes this point more clear by referring to the Stenquist Mechanical Ingenuity Test that he gave to two high school groups, one of boys, the other of girls. The difference in results showed the boys vastly superior. It might be inferred that the boys had more innate mechanical intelligence, but Colvin believes that the explanation for the difference lies in the fact that although the girls might have had experience with mechanical things, -- opportunity to learn, so to speak, -- the incentive or interest was lacking, and was the important factor in determining their lower scores.

This whole newer attitude toward the question of innate ability and its possible measurement is a very

"healthy" one, and is one that is very essential in the measuring and classifying of children. Where the home environment and early training of members of the group differ widely, it is practically impossible to arrive at any true measure of native ability or intelligence by measuring differences in acquired intelligence. It is only when these experiences and interests are similar for the entire group that we can infer innate differences from the differences in the attainment of these individuals in the tests.

This leads up to the question of measuring the intelligence of foreign children. In their case, not only does home training and environment differ from that of American children so as to make it difficult to measure intelligence satisfactorily by means of tests standardized on American children, but the whole question is further complicated by the language factor.

Objections to the Binet Test and its revisions were made, you remember, on the ground that the tests were too verbal, that success in them was too dependent on ability to use words, and that one would be led to take differences in verbal ability to be differences in innate intelligence. Furthermore, these objections were made on the basis of work with children who were not foreign or of foreign parentage.

If criticisms of this nature are applicable to the verbal type of tests when administered to English speaking children, they are even more applicable and justifiable when applied to the measuring of the intelligence of foreign children. In the case of this latter group, the language difficulty offers a handicap which cannot be surmounted, and which makes it impossible, it seems to me, to measure adequately the intelligence of the foreign child.

When foreign children are tested by means of the Stanford Revision, or the more common group tests of the verbal type, they fail to score as high as do American children. More data have been collected on the Italian group in this country than on any other, and the findings of the different investigators confirm this statement. But must it be inferred from such evidence that the Italian group or any other foreign group is necessarily inferior to an American group in intelligence? I think not. I think, rather, that the lower scores of the one group are due in the main to the inability on the part of the foreign children to understand directions and questions in the first instance, and in the second place to express themselves in a tongue which is not their own and which they have not as yet mastered. This conclusion is in accord with the results of studies that have been made.

Brown[1] finds that in the particular Italian group with which he worked, these children rated much higher then tested in their own tongue.

Another study similar to the one just given is that made on a Polish group by Martha Mazurowska[2], Principal of School No. 13 in this city. This, as far as I can observe, was carried out very scientifically, and the findings bear out the fact that foreign children when tested in their native language obtain higher scores than when tested in English, this gain being relatively greater, as we might expect, in those tests that involve more language ability.

Measuring the foreign-speaking child on a scale translated into a foreign language, then, offers one means of securing a better measure of his intelligence[3] than is now being obtained. The other means lies in the use of the performance tests, where the verbal factor is greatly minimized and it is the findings concerning certain of these performance tests and their use in intelligence testing that I now wish to present.

---

(1) "Intelligence as Related to Nationality." Journal of Educational Research, Vol. V. No. 4 pp. 324-327.

(2) "An experiment with Otis Group Intelligence Scale" published in the School Magazine (Buffalo), Vol. 4, June 1922, No. 10.

(3) Pintner thinks that tests involving language when translated into various languages, can never form an adequate comparison and that the non-verbal tests seem to offer the only possibility -- from Intelligence Testing pp. 357-358.

SECTION III

ADMINISTRATION OF THE TESTS

Being interested primarily in the question of more adequate methods of testing and classifying foreign children in the public school system, and the relative value of different performance tests in accomplishing this end, my attention was naturally drawn to the choosing of a foreign group with which to work. Largely by accident I came upon a group of children in School No. 11 which was classified as definitely retarded, being the fourth or lowest division of the first grade. Although my interest was not primarily with retarded children, on being informed that this retarded group was almost entirely composed of Polish children, I realized that here undoubtedly was an excellent field for study. What was the connection or relationship, if any, between the question of retardation and the fact that a large percentage of these children were of foreign parentage? Had these children been adequately classified and grouped? Would the various performance tests give scores comparable with those in the more verbal type of test, such as the Stanford Revision of the Binet Test, and the Pressey Primary Test, which I found had been used in the testing and subsequent sectioning of these children?

Then came the question of carrying out a similar

study of the various performance tests in a group that was neither retarded, nor of foreign parentage. Here, perhaps it would have been wiser if I had used the group of first graders classified as average, instead of choosing, as I did, the bright division of the first grade. So I may say then that this study has been carried out on selected groups of individuals -- in the one case a bright division of children, and in the other a definitely retarded division.

I am of the opinion, though, that the group which I have referred to here, and will continue to refer to throughout the study, as the retarded group, in reality represents the group of foreign speaking children that we find in most of our first grades in foreign sections of the city, and in the great majority of cases in the lowest division of such a grade. I feel, therefore, that this retarded group in which I studied the various performance tests and their value is a rather typical grouping of foreign children which can be easily duplicated throughout the city school system.

On second thought, I am not so sure but that the using of the bright division of children as a group of subjects had its own value. Here certainly was offered the opportunity of testing our performance tests at the other extreme, and finding out if the same results held in the

one group as in the other. If results were found that appeared applicable to both groups, would not such findings add more value to the results than if we had found the same results existing between our retarded group and the average group in that particular first grade?

Let me illustrate. Suppose it should happen that results in the retarded group would show that certain performance tests were more dependent than others for success on cultural and language factors. If similar results were obtained in the bright division -- a division working under so few language and environmental handicaps -- it seems to me our first results would be even more truly diagnostic than if we had compared results with those of the average group, though perhaps in making comparisons of a different sort the average group might have proved superior.

The Subjects.

Let us now turn to a more definite study of the two groups of subjects. The retarded group was composed, as I have said, almost entirely of Polish children. Out of a total group of twenty-six individuals, thirteen girls and thirteen boys, there were twenty children of Polish parentage in whose homes Polish was the language spoken, and six children of American parentage -- in this latter group the mother of one of the children was also Polish.

I was informed by the teacher that many of the

Polish children on entering her class in September 1924 could not speak or understand a word of English, and that in many cases the only English heard throughout the day was that spoken in the schoolroom during the three hours in the morning or in the afternoon.

The median chronological age for this group is 7 years - 1 month, while the median mental age, as computed from the Binet Test, is 5 years - 11 months. The median Binet Intelligence Quotient (I.Q.) is 80.

A large number of the children in this group were in need of a medical examination and medical treatment. Eye defects, tonsil and adenoid cases could not help being detected even on the part of an inexperienced observer. I found on weighing and taking height measurements that 46.15% of the group were underweight, i.e., were 10% less than normal weight for children of their age and height.

On going over the records to ascertain the number having had Kindergarten training, I found that only two out of the entire group of twenty-six had received or had had the advantages of this training.

A questionnaire was made out in which was incorporated among other items the occupation of father, and such data concerning home conditions as could be inferred

from the ownership of an automobile, telephone, piano, bath-tub, books, papers, etc. These answers, together with a visit made to each home during the Easter vacation, gave me rather definite information as to the social status of the homes from which this group came. According to Taussig's division into the five non-competing groups -- Professional, Semi-Professional and Higher Business, Skilled, Semi-Skilled and Unskilled -- I found that most of the fathers of these children fitted into the Semi-Skilled and Skilled Group. In other words, the social status of this group, if we consider the Professional and Semi-Professional and Higher Business Groups as superior, the Skilled Group as average, and the Semi-Skilled and Unskilled as inferior, would be low average with a decided bent toward inferior.

The bright division was composed of thirty-seven children, twenty-three boys and fourteen girls. In this group there were four children of Polish parentage. The question will probably be asked how these particular Polish children came to be in the bright division, and does not the fact that they are so grouped disprove any assertions or claims that the verbal type of tests, i.e., the Binet and Pressey Tests, do not classify foreign children satisfactorily. First, I think the very size of this group of foreign children makes it rather impossible for use as a means of comparison, and secondly, there is an adequate explanation for their presence in this group.

One of these children, Dewitt Z., age 7 years - 2 months, came from a definitely superior Polish home, that is, superior in a cultural sense. His mother is a cultured woman, having been educated in Lithuania and Poland before coming to this country. His father had been in this country for thirty-five years and is a manual training teacher in one of the schools in this city. Some Polish is spoken in the home but the boy hears more English, and the home environment is one that is conducive to what we may call school success.

Two of the other three children, Clara G. and Harold B., are older, having chronological ages of 8 years - 4 months and 8 years - 5 months respectively. This age factor undoubtedly had its influence or weight in the classifying of these two children in the bright division of the first grade. The system of sectioning children on the basis of chronological age has enrooted itself very deeply in our school system and often teachers realize, as in this case, that the individuals could not do the work in the grade in which their age would normally place them, yet they feel that this same age factor ought to place the child in the most advanced section of the lower grade.

The remaining child in the small Polish grouping, Helen M., is approximately of the median age of the group,

7 years - 3 months, but she is a repeater of Section B of the first grade, and would naturally under this circumstance be placed in the advanced division of the first grade for the following year, i.e., that division of bright children with which this study is concerned.

The median chronological age for this group is one month higher than the median for the retarded group, it being 7 years - 2 months. Quite in contrast to this slight difference in median chronological ages is the difference existing between median mental ages in the two groups, the retarded group having, as we have said, a median mental age of 5 years - 11 months as compared with a median mental age in the advanced group of 7 years - 9 months. The median I.Q. rating for the bright group is 107 with two I.Q.'s as high as 122, and fourteen above 110.

Of the children in the bright division 27.02% were underweight, and this in one sense seems to offset any comments as to differences in the physical appearances of both groups, but still the difference was so decidedly noticeable that it is impossible to refrain from mentioning it at this time. The retarded group on the whole appeared under nourished, anaemic, and lacking in that energy, interest and enthusiasm that seemed to radiate from the faces of the bright division, and manifest itself in their activities. Sometimes it seemed almost impossible for me to believe that such different appear-

ing children could be found existing across a corridor from one another.

The data showing physical defects, undernourishment, lack of proper kinds of food, and insufficient hours of sleep in the retarded group can perhaps explain the fundamental reason for the retardation, if such really exists, which appears in this group, but in this present study we have not time for a digression of this sort.

In the bright group there were only six children who had not had Kindergarten training against the thirty-one who had had such training. These figures certainly stand out in contrast to those secured in the retarded group where only two children out of a total of twenty-six had attended Kindergarten. I was interested in getting the data on Kindergarten training in order to see if such training had any effect on the results in the Binet Test ratings or in the scores secured in the performance tests. However, the unevenness in the divisions to be studied in each group rendered such an attempt on the whole practically valueless.

Again, as in the case of the retarded group, I used the questionnaire and visited the homes of all the children. In this group every child except one came from a home where the father is at least a Skilled worker; many

of the children have fathers engaged in the Semi-Professional and higher business occupations, and also some children have fathers engaged in Professional work. According to the classification applied to the other division, this group would rank as average but we may consider it decidedly weighted in the direction of the superior division.

The Tests and Methods Used.

In selecting the performance tests to be used I was influenced in my choice by the Short Performance Test Scale that had been prepared by Pintner and Paterson.[1] The Mare and Foal Test, the Seguin Form Board and the Knox Cube Test were those which they had included in their scale and were three of the tests that I used. In addition, I used the Healy A Form Board (a test included by Pintner and Paterson in the longer scale of fifteen tests), the Kohs Block Test, and the Porteus Maze Test.

Before giving a detailed account of each of these tests and the methods used in administering them, I wish to say a word in regard to the Binet testing. Most of the children in the bright division had been given the Stanford Revision of the Binet Test, while only three individuals in the retarded group had been given the test, the rest in

---

1. Rudolf Pintner, Intelligence Testing, pp. 122-125.

that group having been classified on the basis of the Pressey Primary Test. In order to use the Binet Test as a means of comparison with our performance tests it became necessary to administer it to the remaining twenty-three individuals in the retarded group and to fifteen of the children in the bright division. This testing was accomplished and I thus obtained the mental age and Binet I.Q. ratings of each child in the two groups.

It is also important, I think, to state that all of the testing in both groups was done at similar times. Both of these groups were in session for only half a day, one month in the morning and the following month in the afternoon; hence if the Kohs Block Test was given in the morning to one group it was also given at a later date in the morning to the other group. This arrangement held for all the tests so that the item of fatigue was equalized as far as possible in both groups.

The room used in administering the individual tests was the same in all cases -- a large room, well lighted, quiet, containing little furniture except a large table in the center and several chairs, with book cases along the wall. On the whole there was very little interruption or disturbance caused by people entering. When a test was administered to a group the regular classroom was used for that purpose.

Before giving any of the tests I visited the groups a number of times, observing the children in their school work and in their play activities, learning their names and finding out the individual characteristics and peculiarities of many of them. Most of the work of weighing them and of filling out the questionnaires from talking with them was completed before the testing began so that I did not appear as a stranger, and I think for this reason the response from both groups in the tests, though especially that from the retarded group, was better than if I had merely given the tests without first establishing the contacts. Perhaps under such circumstances, an examiner is apt to do some coaching and make allowances, but on the other hand it is very possible for the examiner to be perfectly objective in his attitude and observations without losing sight of the personal side of the relationship.

The Mare and Foal Test. (Individual Test) This test, devided by Healy and Fernald, consists of a board 11-1/2" x 9" upon which is mounted a country scene. This scene, showing a mare and foal grazing, chickens, sheep and a country farm house, is so arranged that pieces come out and have to be replaced in their proper position before the scene or the objects themselves are complete. Eleven pieces in this puzzle are removable. Two refer to definite

objects; four pieces are parts of animals; one piece is just an odd shaped piece of green which completes the meadow; while the remaining four are sky pieces. In two of the latter it is essential to match clouds and trees. There then remain two pieces to be fitted together in order to complete the puzzle. (◿◺ △)

The subject was shown the scene with the parts spread out at the side and told to "put the pieces in so as to make a picture." The time was recorded both when the two triangular pieces were fitted together to complete the picture and the time when this fitting was not done, the rest of the test having been completed. A time limit of five minutes was allowed for the completion of the test.[1]

Previous to giving the performance test to these particular groups of children I had tried them out on five other first-grade youngsters of similar ages, noticing at this time that the fitting of the triangular parts together was largely a matter of chance. It was for this reason that I distinguished between time necessary for

---

1. Pintner and Paterson recorded time and errors in their use of the test, but the relation between these factors is so direct (time increasing when errors increase) that the time was the only record taken into account.

completion with triangles and time necessary for completion without triangles, with a view toward taking the factor of chance somewhat into account when computing the final score for each individual. This was done by working out an arbitrary formula for the time: $T = A + 1/2\ B$, in which T represents the time, A the time to complete the test without triangles, and B the time required to complete the test with the triangles. If a child failed in the five-minute period to complete the test, failure being due to inability to place the triangles, a score of 10 minutes (10') was recorded for B. If, on the other hand, the five-minute period elapsed with only two pieces in on the whole board, a score of 10' was recorded for B and the score estimated for A. In this case A would be 22'30".

The following table shows the median score and the first and third quartiles for this test in both the bright and retarded groups:

Bright Division

Median - 2'49.2"
Q_1 - 2'9.6"
Q_3 - 5'15"

Retarded Division

Median - 6'24"
Q_1 - 4'37.5"
Q_3 - 9'

The Seguin Form Board. (Individual Test) This form board as designed by Seguin consists of a large board 16" x 22" in area and 1-1/2" thick. Ten variously shaped pieces one inch thick are set into sockets one-half inch deep. The individual was shown the form board, the blocks to be fitted in having been placed on the table beside the board, and then told to put the pieces in their places as fast as he could. This procedure was repeated and the time for three successive trials noted, but in using the results only the time of the shortest trial was taken.

The following table shows the median score and the first and third quartiles for this test in both the bright and retarded groups:

Bright Division

Median - 21.3"
Q_1 - 19.08"
Q_3 - 24.37"

Retarded Division

Median - 27.5"
Q_1 - 20.5"
Q_3 - 46.25"

The Knox Cube Test. (Individual Test) This is a performance test devised and used by Howard A. Knox primarily for the mental classification of immigrants at Ellis Island.[1] Four one-inch cubes are fastened about two

---

1. Journal of the American Medical Association, Vol. LXII, pp. 741-747.

inches apart to a piece of thin boarding. The experimenter holds the board in one hand and with the other taps the cubes in a certain order and at a certain definite rate (about one tap per second), always commencing with the cube to the right of the subject. The subject is placed in front of the experimenter at a distance of about two feet and is given the following direction: "Do that."[1]

The relative unimportance of the language factor for success in this particular test stood out as an important item in influencing my choice, for here I felt that the foreign child would be better able to compete with his fellows, with the handicaps of language at least partially removed.

---

1. Pintner does not believe that even this is necessary and he is sure that without the command exactly the same results would be obtained in the test. This was borne out in his work by some of the subjects who did not understand English and by deaf children. In these cases he found that all that was necessary was to make some gesture indicating that the pointer was to be picked up and the blocks were to be touched. He also found that the first line of the test (1, 2, 3, 4) was so simple that even though the subject does not know while watching the examiner that he will be required to do the same thing, he can easily imitate what has been done and then understands what is expected further of him, without any directions being given during the entire procedure.

I used eight combinations: (a) 1 2 3 4; (b) 1 2 3 4 3; (c) 1 3 2 4; (d) 1 4 3 2; (e) 1 4 2 3; (f) 1 4 4 1; (g) 1 3 2 4 1; (h) 1 4 4 2 2 1.[1] Five of these lines were identical with Pintner's; line (g) differing only in the fact that he had 3 following the last 1 (1 3 2 4 1 (3)), while lines (f) and (h) were entirely different.

The number of combinations or lines correctly imitated is recorded. In this case, then, the maximum possible score obtainable was eight. One- half a point was allowed if the subject on the second trial completed the line correctly.

The following table shows the median score and the first and third quartiles for this test in both the bright and retarded groups:

Bright Division

Median - 6.18
Q_1 - 5.25
Q_3 - 6.84

Retarded Division

Median - 5.25
Q_1 - 2.9
Q_3 - 6.08

---

1. Knox in standardizing the test used only five combinations or lines; (a) 1 2 3 4; (b) 1 2 3 4 3; (c) 1 2 3 4 2; (d) 1 3 2 4; (e) 1 3 4 2 3 1 (number one always referring to the block on the right hand side of the child). Pintner in his further work on this particular test increased the number of lines to twelve.

The Healy A Form Board. (Individual Test) This test was devised by Healy in his work on delinquents. It consists of a frame 4" x 5" in which five (5) blocks can be made to fit exactly. The subject was shown the frame with all of the blocks in place. The frame was then turned upside down on the table, the blocks falling out, and the following directions given while pointing to the blocks: "Put those in as they were before." Pintner in recording the score counts time and moves, but as in the case of the Mare and Foal Puzzle Test the increase in errors shows itself in the increased time scores, so that the latter scores were the only ones recorded. Two trials were given, a five minute time limit being allowed for each. The final score used was the total time for the two trials.

If a child failed in the allotted time to get the blocks in he was given a score of 10' and shown how they could be made to fit in. The frame was again turned upside down on the table and the subject instructed as before. The time necessary for the second trial was then added to the 10' for the total score. In case an individual failed after being shown, his score would be indicated by 20'.

The following table shows the median score and the first and third quartiles for this test in both the

bright and retarded groups:

Bright Division

Median - 1'22.5"
Q_1 - 32.6"
Q_3 - 4'52.5"

Retarded Division

Median - 2'
Q_1 - 1'
Q_3 - 10'15"

The Kohs Block Test. (Group Test) This test, which was devised by S.C. Kohs,[1] can be used as an individual test or as a group test. We used it as a group test. Seventeen sheets of card board, carrying designs that increase in difficulty and in the number of blocks necessary for their completion, are exhibited one by one for a definite period of time (these time periods having been worked out by Kohs) and the subjects are asked to make with the set of blocks they have (sixteen blocks in all) the design that they see before them. Prior to the beginning of the test, the examiner takes one of the blocks in his hand and by questioning different children in the class as to the different colored sides of the block, he ascertains whether or not they know and recognize the different colors. Besides using this precaution in both groups I drew a design on the board, colored it, and asked the

---

1. Samuel C. Kohs, Intelligence Measurement: a psychological and statistical study based upon the block design tests.

children to make it for me with the blocks they had. I then walked around the room, examining the work of each child, correcting where necessary, and commenting if the design was made correctly. After this had been accomplished, the test proper was commenced.

The teacher of each division helped me in administering the tests by keeping the time. Using Kohs' direction, a minute and a half was allowed for Tests 1, 2 and 3; two minutes for Tests 4, 5, 6, 7, 8, and 9; three minutes for Test 10; three and one-half minutes for Tests 11, 12, 13 and 14; and four minutes for the remaining three tests.

In scoring, two points were given for a correct arrangement of the blocks, the design being exactly like that exhibited on the chart before the class. One and one-half points were given if the elements of the design were correct, but some slight error had occurred in the accuracy of the plan. One point was given if some of the elements in the general plan were correct. While a score of 0 was given when the design was entirely wrong, or when no attempt was made on the part of the subject to perform the test.

Allowance was also made in this test for speed. An additional credit of one was added to an individual's

score every time he correctly completed any of the first ten tests in less than one minute, or the tenth one in less than a minute and a half. Additional credit would have been allowed if any of the tests beyond the tenth had been completed in the shorter time interval, but this was not done as no subject in either division succeeded with any of the tests beyond the tenth.

This exception was made, however, to the above. Before being given the additional credit the subject had to do at least two tests in the minimum period of time. This was done to counteract somewhat the simplicity of the first two tests in which most of the individuals, especially in the bright group, were finished before a minute's time had elapsed.

The following table shows the median score and the first and third quartiles for this test in both the bright and retarded groups:

Bright Division

Median - 10.88
Q_1 - 8.75
Q_3 - 14.45

Retarded Division

Median - 7.71
Q_1 - 4.75
Q_3 - 10.5

Porteus Mazes. (Group Test) In the type of test represented by the mazes, according to Porteus, we are getting at or testing "prudent and preconsidered action", a type of response which he believes is lacking in the other types of tests, more particularly in the Binet. Furthermore, it had been held that this type of test was also very suitable for testing non-English speaking children, as no language responses are required. For these reasons I was very anxious to try out these mazes on the two groups of children.

All the tests are carried out by tracing with a pencil on a printed maze. In the mazes the child is faced by a number of blind alleys and he must trace his way through without entering these, the blind alleys really penalizing the subject for his lack of forethought. I used the mazes that had been standardized for the fifth, sixth and seventh year levels.

Before distributing the three mazes I drew on the blackboard in each room a typical maze illustrating, as was done in the printed mazes, the starting point by an arrow, and a letter S, and the finish by another arrow. I then explained what I was going to do -- find my way through the maze -- and with the chalk I drew the path that I would take. Two or three times I purposely ran into one of the blind alleys and then asked what was wrong.

The responses at this time were such as to make me think the purpose of the test was clear in the minds of the subjects.

Included also in the instructions were the directions that three mazes were to be given out and as soon as work was completed on one, the next was to be turned right side up and completed, the same procedure to be followed for the third maze. At the end of two minutes all work was stopped and the mazes collected.

The scores were expressed by total number of errors made on the three mazes.

The mazes were given first to the bright division and on looking over the scores I found that only nine individuals out of a total of thirty-seven could be classed as successful on the three mazes, and in all of these cases there was recorded at least one error for maze V. This led to the conclusion that either the directions had not been understood or that some factor which I did not know about had entered in to make this test practically valueless as far as the bright division was concerned.

The maze devised for the fifth year level differed decidedly in arrangement from the other two mazes, besides differing from the maze I had drawn on the blackboard. Not only was the arrangement different, but the arrow repre-

senting the finishing point was missing and as a result of this last difference many of the children were apparently at a loss as to what to do when they reached what actually was the end of the maze. On the other hand many of the children when they came to maze V, or if this maze happened to be encountered first, spent most of their time with it, or apparently not knowing what to do on account of its different arrangement, stopped working entirely.

It looks very much as if I had been at fault in not showing a maze of the type represented by V or at least in not explaining that in the case of V no arrow showing the ending would be found. Furthermore, I should have had maze V so arranged in the group that all of the children would have encountered it at the same time. It seems to me that those people who encountered V last had the advantage of those who attempted it at the first or at the second trial.

After realizing that results in the bright division were practically valueless, and that I would not be including the Porteus Mazes in my study of performance tests in both groups, I decided nevertheless to try out the tests in the retarded group, with some variations, however.[1]

---

1. Eleven children in the retarded group were given the tests with the same directions as those outlined above, and with maze V occurring in any order with respect to the other mazes. (cont'd)

It appears, from my work at least, that this type of test presents difficulties enough in a division of bright English speaking children, and that the language factor involved in giving directions is insurmountable for the non-English speaking child and renders the test impractical for use in a foreign group. It did happen that when the tests were given individually the Polish children improved, but in this case they had the advantage of reacting to one maze (the practice maze) and hence of understanding what was wanted and expected of them in terms of their experience instead of by a method of verbal explanation.

Of this group no one completed the three tests in the two minute time limit or was successful in all of the tests. As a matter of fact two was the highest number of mazes completed and these were not done without a great number of errors.

I then varied the administration of the test to the remaining fifteen individuals in the class. To one group of four I arranged it so that maze V would be encountered first, and illustrated on a large sheet of paper a typical maze, incorporating also some of the difficulties which would be met with in maze V, the children standing around me. To another group of seven I gave the test under the conditions described above, with the exception that directions were given also in Polish. This latter innovation, which I thought at first would be very successful, was not productive of any better results than when the test was given in English. I think the explanation for this lay in the fact that the children were not accustomed to hearing Polish spoken in school. They themselves had been repeatedly reprimanded for using Polish at recess time or when they were playing in little groups. When directions were given by a teacher in their language they were even more confused, - in fact, their reaction appeared to be one of astonishment, followed by a tendency on their part to laugh, thinking someone was joking with them. To the (cont'd)

Speaking of the group of performance tests used in general, I have already commented on the fact that the Knox Test does not require language for its success. I think the same applies to the Healy Form Board Test. In both of these the subject watches the examiner for instructions really, and the language factor is relegated to the background. In the Mare and Foal Test and Kohs Block Test, it seems to me the language factor is more important again as is also an acquaintance with a rather typical American environment. In other words, we might classify these latter tests more as "culture" tests. Up to this point this is, of course, purely supposition and shall be verified later on in this study, Part IV. This portion has been given over merely to descriptions of the tests, before treating of them within the two groups in a more analytical way.

The Method of Correlation. All of the correlations except those comparing the teacher's estimate with the different performance tests were computed by using Spearman's method of expressing correlations in terms of ranks or position. $\rho = 1 - \frac{6 \times \Sigma D^2}{n(n^2-1)}$, where ρ' = the coefficient of

---

remaining four individuals of the class I gave the test with maze V coming first, but with directions being given individually. Furthermore, in addition to giving the directions individually, I had each child find his way through the practice maze. This method, as we might naturally expect, gave the best results.

, ρ later being converted to r.

correlation, D = the difference in ranking of an individual in the two series being compared, and n = the number of cases or individuals involved. In the correlations concerning teacher's estimate and various tests I used Pearson's product-moment correlation formula, $r = \frac{\Sigma xy}{n(\sigma x \sigma y)}$, where r = the coefficient of correlation, σ = standard deviation, and Σxy = product-sum of the deviations.

The best measure of the value of any correlation is called the probable error (P.E.). The P.E. of a correlation represents that distance from the correlation index (r) within which one-half of similar correlations will be expected to fall. The formula commonly used is: P.E. $= .6745 \frac{(1-r^2)}{\sqrt{n}}$, where r = the coefficient of correlation, and n, the number of cases. For our correlations the P.E. is in the neighborhood of .1 ranging from .07 to .14, which means that our correlations must be entirely suggestive of significant differences. As was stated previously, it is not claimed that the results are more than to warrant further investigation in this field.

SECTION IV

RESULTS

After stating the purposes of the study, and describing the subjects and tests, it seems only necessary to add the table of correlations.

Table A.

		Mare and Foal	Seguin Form Board	Healy Form Board	Knox Cube Test	Kohs Block Test	Combined tests (5)	Selected Tests Kohs and Mare and Foal
I. Q. Ratings	Bright	-.15	-.28	.23	-.06	.06	-.05	
	Retarded	.39	.20	.16	.20	.18	.27	
Mental Age	Bright	.25	.02	.16	.15	.40	.29	.40
	Retarded	.65	.23	.46	.28	.60	.64	.66
Chronological Age	Bright	.37	.26	-.19	.15	.17	.21	
	Retarded	.31	.09	.29	.15	.52	.42	
Teacher's Estimate	Bright	-.06	-.14	.39	.07	.18	.16	
	Retarded	-.05	.14	.35	.54	.28	.35	

In considering these coefficients of correlation, attention is called to the fact that the correlations in the retarded group are on the whole higher than similar correlations obtained in the group of bright children. This is due to the fact that the magnitude of a coefficient of correlation usually depends upon the range of the measures dealt with or, otherwise stated, upon the degree of hetero-

geneity of the group. If the group is homogeneous, like the bright division, all of the individuals will have about the same ability. One individual will rank the highest on one test, perhaps lowest on the next, and average on the one following, so that when correlations are obtained between the tests, these tend on the whole to be low. On the other hand, in a heterogeneous group the range of ability is greater. These individuals who score high in one test will be more likely to score high on the other tests, and as a result, similar correlations in such a group tend to be higher.

In the bright group the Healy A Form Board gives the best correlation with Binet Intelligence Quotient Ratings;[1] that is, this test stands out in this group as being a good brightness or alertness test.[2] The fact that this test also gives a negative correlation with chronological age - a correlation of -.19, which, though not high enough to be of much value, shows at least a tendency - also points to the Healy Form Board as a brightness test.

---

1. The references throughout which are made to Binet mental age, Binet Intelligence Quotient ratings, Binet chronological age refer to the Intelligence Quotients, mental ages and chronological ages as determined by the Stanford Revision of the Binet Tests.

2. For this and subsequent correlations through page 75 refer to table A on page 55.

Mental age does not tell us how bright a child is. Two children might have the same mental age and yet differ enormously in brightness. For example, a boy ten years old chronologically might have a mental age of eight years, while another child six years old might have a mental age of eight. Both have the same mental age, yet we would speak of the former as being retarded, even dull; while the latter we would classify as superior in intelligence. Obviously, the meaning of a mental age depends upon the chronological age that goes with it. Thus when we consider mental age with respect to chronological age and deal with an individual's Intelligence Quotient, which represents his mental age divided by his chronological age, we have a measure of his brightness. And it was in connection with these I. Q. ratings or brightness indices that the Healy Test stood out as the best.

The Healy Form Board is also a test of abilities and capacities in which the teacher is undoubtedly interested, for in the bright group we find a positive correlation of .39 existing between this test and teacher's estimate of intelligence.

The Seguin Form Board Test in the bright group gives a negative correlation of -.28 with Binet I.Q. ratings and a positive correlation of .26 with chronological

age. This leads to the conclusion that this test is largely a maturity test, success in which is based to a great extent on physical development and muscular coordination.

It is interesting to note that the Mare and Foal Puzzle, apparently a culture test, although not showing any appreciable negative correlation with Binet I. Q. ratings, shows a comparatively high positive correlation with chronological age ($r = .37$), thus showing that the age factor is also important for success in this test. In this latter test, however, the age factor counts not in the sense of better muscular coordination and adjustment, but, rather, in affording the child a longer opportunity to become familiar with, and to have experienced and dealt with those elements that are a part of the test. In other words, the older child has had more of an opportunity for reacting to horses, chickens, and a "country situation" in general.

The other correlations in the bright group with respect to I. Q. ratings are too small to be of any significance.

When it comes to the correlations existing in the retarded group between Binet I. Q. ratings and performance tests, the Mare and Foal Puzzle gives the highest correla-

tion (r = .39). That is, the test which is largely cultural stands out in this group as the best brightness test.

The question arises as to how this test, which showed a negative relationship with Binet I. Q. ratings and a positive correlation with chronological age in the case of the bright group, - a relationship we expect to find existing between an index of brightness and chronological age - can in the retarded group show a positive correlation with both Binet I. Q. ratings and chronological age. I think the answer perhaps lies in the fact that the retarded group, as before mentioned, was far more heterogeneous in character, and as a consequence possessed a much wider range of abilities. Some of the younger children did not succeed any better with the test than did the older ones, and on the whole a positive relation seems to exist in this group between mental age and chronological age, i.e., those who were the older mentally on the whole were the older chronologically. This is corroborated by a positive correlation between Binet mental age and Binet chronological age of .24. On the other hand a similar correlation between mental age and chronological age in the bright group gives a small negative correlation. This correlation of -.18 is almost negligible, except in showing that the negative relationship which we normally expect to find

existing between mental and chronological age tends to exist in the bright division.

Another rather interesting result is the correlation of .27 found to exist between Binet I. Q. ratings and the combined tests. A combination of performance tests apparently would serve as a better measure of brightness in a retarded group than any one test, with the exception of the Mare and Foal.

Let us now turn our attention to the relationship existing between our various performance tests and mental age, for on the whole for school purposes we are more interested in tests which measure mental age than those which give us measurements of relative intelligence or brightness.

In the bright division the Kohs Block Test and the Mare and Foal Puzzle Test with the positive correlations of .40 and .25 respectively, stand out as the best measures of mental age. When the Kohs and the Mare and Foal Tests are combined and correlated with mental age, we find again a correlation of .40. This is slightly lower than we should have been inclined to expect from the correlations found when these two tests were run separately with mental age.

In the retarded group the Kohs Block Test and the Mare and Foal Puzzle are also found to correlate the highest with mental age, though in the reverse order from that obtained in the bright group. The Kohs Test gives a positive correlation of .60, and the Mare and Foal a positive correlation of .65.

The Mare and Foal Test tends in the retarded group to give a fairly high correlation with chronological age, a result similar to that found in the bright group with this test, and the explanation offered in that connection is again applicable here.

It was rather unexpected to find the Kohs Block Test, which in the bright group correlates so poorly with chronological age, showing an even higher correlation with chronological age in the retarded group than the Mare and Foal Puzzle, the latter giving a positive correlation of .31 as compared to .52 for the former. Perhaps in a group of foreign children the elements necessary for success in the puzzle, (i.e., experiences with chickens, horses and sheep), are encountered earlier in the child's life. Even here age is an important factor, as we have already seen, While in the situation more typical perhaps of American middle class homes, represented by the blocks, and their grouping into various geometrical designs, the age factor plays

an even more important role.

It is a very interesting and important fact for us that the highest correlations with mental age found in either group are those given by the Kohs Block Test and the Mare and Foal Puzzle Test. These two, which we have spoken of as culture tests, for we believe they are more dependent than the others on an acquaintance with a rather typical American environment, are the performance tests which show the best relationship with mental age as obtained by the Binet Test.

Thus our assumption that the Mare and Foal Puzzle and the Kohs Block Tests are culture tests is substantiated by the results found on running the correlations between them and Binet mental age, for it is a matter of common observation on the part of most psychologists that the Binet Test is very largely a culture test, and hence the mental age derived from this test would also be influenced by the factor of culture.

I believe there is another factor operative especially in the retarded group in regard to these two tests, (the Mare and Foal and the Kohs Block Test) that is brought out by the high correlations found between them and Binet mental age. These tests apparently call out or upon verbal and linguistic abilities which we know are necessary

for success in the Binet Test. That is, the understanding of the directions given in English necessary for success in these particular tests, tends on the whole to offer to a foreign group, though not to the same extent, the same difficulties met with in the Binet or in any other type of verbal test.

In a group composed almost entirely of foreign children it looks as though it were possible to administer either the Kohs Block Test or the Mare and Foal Test to the group and get results that agree as closely with Binet mental age as do the results obtained by means of the Pressey Test.[1] A combination of Kohs and the Mare and Foal Puzzle would give an even greater agreement with Binet mental age. As the objection on the part of departments of education to the Binet Test is the time consumed in testing each individual, this combination of Kohs and Mare and Foal, or of Kohs alone, (it being a group test) might advantageously be substituted for the Binet in place of the Pressey which is the most frequent substitute at present for the Binet.

---

1. Correlation between Pressey ratings and Binet I. Q. ratings +.56.
Correlation between Pressey ratings and Binet mental age +.61.

If, on the other hand, we look at the other tests in the retarded group we find the Healy Form Board and the Knox Cube Test standing out as tests worthy of consideration. They show good correlations with mental age, and also with teacher's estimate, and yet because their correlations with mental age are lower than the culture tests, they prove themselves less dependent on the environment of the middle class American home and it seems to me for this very reason that they are more suitable in classifying foreign children.

It is also significant that the Healy Form Board and Knox Cube Tests require in their administration very little explanation as to what is wanted, which in turn means that the foreign child is not confronted by the language difficulties that he is in tests such as the Mare and Foal and Kohs and this factor may also account for their lower correlation with Binet mental age.

These four tests have their values. The Mare and Foal and Kohs classify the children in both groups more in accord with the Binet Test, and may even in some cases be advantageously substituted for it. The other two, the Knox Cube Test and the Healy Form Board, especially in a foreign group, show themselves to be of value in classifying the children because of their very independence

of language training and general "American" background. In other words success in these tests does not depend on acquiring a strictly American culture.

I have already mentioned the fact that in the bright group the Healy Form Board correlated well with the teacher's estimate. In the retarded group we see that the kind of abilities - alertness, attentiveness, etc. - which are tested in the Healy and Knox Tests, are also factors in which the teacher is interested and takes into account in estimating the child's intelligence. Further mention will be made concerning teacher's estimate and the tests later in the study.

Interested still further in these various performance tests I prepared some tables for both the bright and the retarded groups to show the effects, if any, of Kindergarten training and of the language spoken in the home on the scores obtained by the children in the tests.

The divisions within each group representing the children without Kindergarten training and those with it, were very uneven, and perhaps little value can be attached to the tables for this reason. Yet the results tend to show that in neither group did Kindergarten training nor the absence of such training have any appreciable

effect on the performance test results. However, the results obtained when I. Q. ratings were used showed that in both groups those without Kindergarten training tended on the whole to be lower in their scores than those children who had had the advantages of Kindergarten training.

When we look at the factor represented by the language spoken in the home, we are again met by the unevenness in the divisions within each group. In the bright division there were only four individuals who came from homes where English was not spoken; in the retarded group the divisions were slightly more even, however, two-thirds of the children coming from homes where Polish was the language used and nearly one-third from homes where English was spoken.

In both groups, though the significance is questionable on account of the small number in the bright group, the distribution of I. Q. ratings on the basis of language spoken in the home shows that the child from the home where a foreign language is spoken is decidedly lower. This is the result expected, for it is known that the Binet is a culture test, dependent upon a typical or average American environment, and furthermore that its success is based largely on linguistic abilities. Hence it

would be only natural for the foreign child, reared in a home where the language so necessary for success was never spoken or at best spoken or heard only occasionally, and where the environment within the home, at least, is so different from the environment in an average home, to be at a decided disadvantage in this test, and to make a lower score than that of the American child.

In the bright division, for the performance tests, there is apparently very little relationship between the language spoken in the home and test scores, but here again it must be remembered that the number of children coming from non-English speaking homes was so small that the prediction of any relationship between language in the home and test results is impossible.

Especially in the retarded group we might have been led to expect that the scores in the Mare and Foal and the Kohs Block Tests would also be affected by the language spoken in the home. The Mare and Foal Test does show the children in English speaking homes scoring higher, but the Kohs Block Test shows the group represented by the foreign speaking children as being superior to the group representing children in whose homes English is spoken.

I have been under the impression that the Eng-

lish speaking children in the retarded group were definitely inferior to the majority of the Polish children, but that the latter scored lower on the Binet Test largely on account of language difficulties. I have also thought that language difficulties entered into the Mare and Foal Puzzle and the Kohs Block Tests, thus giving the English speaking portion of the retarded group advantages here also. This latter opinion was confirmed by the study within the group in regard to the Mare and Foal Test. Here the children from English speaking homes scored higher. The Kohs Test, however, did not show the expected relationship between English and non-English speaking children.

It is in connection with this study of the tests within the group that the Healy Form Board, and especially the Knox Cube Test again prove themselves good tests for classifying a group of foreign children.

In both of these tests the language difficulty is removed to a much greater extent than in the Kohs or the Mare and Foal, and as a result the foreign speaking children are found standing out as definitely superior to the English speaking third of the group - standing out in what I believe is their true position within the group. The median for Polish speaking children in the Healy is

1'48.8", for the English speaking group 3'23.5". In the Knox the median scores are 5.25 and 2.5 respectively.

The Binet Test or the Mare and Foal Test never could give this relation within the group, for in these tests the child speaking the foreign tongue is put at an initial disadvantage, which he is never able to overcome, while on the other hand children who are actually inferior to him in ability are able in the more verbal-and-culture type of test to appear more intelligent. Or, if we speak of the classification of foreign speaking children in general, they are at a disadvantage in the verbal type of test. Yet the Binet or Pressey Tests are those usually employed for classification. More frequently than not, the Pressey, being a group test, is the one that is chosen, and this latter test is even more dependent than Binet on linguistic ability. Being at this disadvantage their abilities and capacities are never adequately measured. From I. Q. ratings these children appear retarded, dull, even feeble minded, if we accept Terman's classification of an I. Q. below 75 as designating feeble mindedness, and are grouped with English speaking children who, under no language handicaps or handicaps of a partially foreign environment, test low and are definitely dull or borderline.

Here it would seem is where a grave injustice

is done the foreign speaking child, and it is toward the solution of this problem that the second part of this study is devoted. Can certain of these performance tests be advantageously used or combined with the Stanford Revision of the Binet Test so as to classify children more satisfactorily than they are now being classified in the public school system? This question concerns the foreign speaking child more particularly though as before in studying the various performance tests the study has also been carried on in the division of bright children.

When we look at the graphs[1] showing the distribution of the I. Q. ratings, and the scores on the various performance tests in both groups, we notice the fact that in the distribution of performance test scores there is far more overlapping than is found in the curves showing distribution of I. Q. ratings in both groups. The overlapping again is greater in the distribution of Healy and Knox scores than in the other performance tests. These findings, together with the data presented above showing the relation of the language spoken in the home to test scores, impress the fact upon us more emphatically that in fairness to the foreign speaking child different methods, or at least correction of the methods now in use, are necessary for the more adequate and just classification

---

1. Inserted at the end of the study.

of children in the grades.

In securing for better classification the best performance tests, or the best combination of performance tests with the Binet Test, we have used as the criterion, teacher's estimate. The teacher in both sections was asked to rank all the children in her class as to their general intelligence. She was asked to include under this heading mainly the kind of school work done by each child, his abilities and capacities for leadership as manifested in play, and his general alertness, understanding and interest in the "school situation". A scale ranking from 1 - 5 was used: 1 being very inferior; 2, inferior; 3, average; 4, superior; 5, very superior. After the children had been ranked in this way correlations were obtained between these estimates and Binet tests, the various performance tests separately, and with combinations of these.

The question will undoubtedly be raised as to the value to be attached to teacher's estimates, and by some this criterion may be classed as valueless, but it seems that at the present time at least this is the best criterion by which we have to judge the tests. The most adequate method of judging the tests, of course, would be to observe these children, and twenty or thirty years

from now see if they have been successful in their adaptations to their environment. Were those who scored high in the Binet Test and performance tests the most successful in life; or were those classified as average in the tests more successful in adapting; or perhaps, were those ranked low in the tests better able to meet and cope with the exigencies of life?

Records of test scores and records of success in later life will eventually, we hope, be tabulated and we then will know definitely the value of our tests, but until then the estimates of an understanding teacher who has been in close association with the children within the group throughout an entire year seems, I think, to stand out as a criterion or "touchstone" of considerable value in the way of checking or testing the tests.

Example after example might be cited of the checking up and the evaluating of various kinds of tests by teachers' estimates concerning the individuals tested. Cyril Burt, eminent English psychologist and worker in psychological testing, adopts as his criterion for testing the tests the judgments of the class teacher or principal, and in justifying this procedure states: "There is no standard of comparison which can surpass or supersede the considerate estimate of an observant teacher working

daily with the individual children over a period of several months or years."[1]

There is one check on the estimating of general intelligence which I did have on the teacher in the retarded group, and which is worthy of comment. This teacher knew the I. Q. ratings as based on the Pressey Test that had already been given. I had, therefore, assumed that her estimates of the general intelligence of her children would be biased, or at least influenced by this knowledge, and yet on running the correlations between her estimate and Pressey ratings, and her estimate and Binet ratings - the latter being ratings about which she knew nothing - I found a positive correlation of .31 between her estimate and Pressey ratings as compared with a positive correlation of .44 with Binet ratings. This, it seems to me, would tend to show that at least in this particular teacher's case her judgment not only was unbiased but was definitely valuable. Undoubtedly the Binet Test is a better measure of intelligence than the Pressey, and more particularly so in a foreign group, and the judgment of this teacher was more in accord with

---

1. Mental and Scholastic Tests. p. 199.

the Binet measure of intelligence than with the Pressey ratings she already knew.[1]

When speaking of the different tests in both groups, we have mentioned those which measured abilities in which the teacher was interested, and though at that time the point was not stressed, it must be realized that if we accept the teacher's estimate as a criterion by which to judge the tests, those performance tests which show higher correlations with teacher's estimate are on the whole the most valuable.

It will be seen that in both groups there is a similarity between the correlations of the various performance tests and teacher's estimate, if in the bright group we exclude the correlations with the Knox Cube Test and with the combined tests. It is probable that the Knox Cube Test when given to a group of bright children is too easy for them and hence does not bring into play any of the capacities that a teacher is interested in, whereas if given to a group of foreign children, the demands made on alertness and attentiveness for success do call out,

---

1. This teacher is also a college graduate and has specialized in educational work.

or upon, abilities that a teacher is interested in and considers in her estimation of intelligence, in fact, as is shown in the tables, the Knox Cube Test in the retarded group gives a better correlation with teacher's estimate than does the Binet Test. It seems possible to give this single test to a foreign group and get a better classification of children than we are now securing by using the Binet Test alone, though this procedure would not hold at all in a bright group of English speaking children.

The difference in the correlations .16 and .35, when the combined tests are compared with teacher's estimates in the bright and retarded groups respectively, is more difficult to explain. In the bright division the combination of the five performance tests has very little in common with the Binet Test, as is shown by the negative correlation of -.05 with the I. Q. ratings and the relatively low positive correlation of .29 with Binet mental age. But on the other hand we find the teacher's estimate does correlate well with Binet I. Q. ratings, i.e., the teacher is interested in those capacities and abilities which the Binet Test measures. As we have found that very little relationship exists between this test and the performance tests in this group, it might be expected that

the teacher's estimate would correlate poorly with the performance tests. This was found actually to be true.

In the retarded group the relation of performance tests to the Binet Test is reversed, though the relationship between teacher's estimates and Binet ratings still remains. The various performance tests, more especially the Mare and Foal Puzzle and the Kohs Block Test, test or measure the same sort of abilities in a foreign speaking group, that are measured by the Binet Test. This is clearly shown in the positive correlation of .27 between combined tests and I. Q. ratings, and the still higher correlation of .64 between the combined tests and Binet mental age. So with this relation between performance tests and the Binet Test, and the previously mentioned relationship between the Binet ratings and teacher's estimate, we would expect to find in the retarded group a higher correlation between teacher's estimate and the combined tests.

Table B.		I.Q. Ratings.	Combined Tests (5)	Binet and Combined Tests (5)	Binet and Healy Form Bd. and Knox Cube Test
Teachers Estimate	Bright	.46	.16	.50	.51
	Retarded	.44	.35	.41	.56

In the retarded group, we find a positive correlation of .44 existing between teacher's estimate and Binet I. Q. ratings. Realizing that the Binet Test is largely a linguistic and verbal type of test, I conceived the idea that perhaps I could get a higher correlation with teacher's estimate, and hence a better classification of the child, if I combined the five performance tests with the Binet Test and then ran the correlations with teacher's estimate. I did this and found a positive correlation of .41. This lower correlation was rather surprising at first, and then upon thinking the matter over I decided from my investigation and previous knowledge of the various performance tests themselves that certain of these tests, especially the Mare and Foal and Kohs, were too much like the Binet Test to contribute anything by being combined with it. In order to test this out, I took only the Healy Form Board and the Knox Cube Tests, those which I had found in the study were less dependent on language and environmental factors, and combined them with the Binet Test, and then ran the correlation with teacher's estimate. I found a positive correlation of .56. From this it may reasonably be concluded that we have here a better means of classification for foreign children than that given by the Binet Test alone, or by the combination of the Binet Test and unselected

performance tests.

A similar grouping of tests and subsequent correlations in the bright division add to the value of the results found in the retarded group. The correlation between teacher's estimate and Binet I. Q. ratings is .46. In combining all of the performance tests with the Binet Test a positive correlation of .50 is obtained. This increase, though contrary to the results found in the retarded group, actually proves the contentions that some of the performance tests in a foreign group are too similar to the Binet Test. In the bright division no language difficulty is encountered and the combined performance tests have little in common with Binet ratings. Certain abilities not brought out in the Binet Test are called forth in the performance tests and as a result when the five performance tests are combined with Binet ratings, we get a higher correlation with teacher's estimate. But even within the bright group, if we select those performance tests which we know are less similar to the Binet Test than others, namely, the Healy Form Board and Knox Cube Test, and combine them with the Binet ratings we get a slightly higher positive correlation of .51. We may, therefore, conclude that even in a superior group of American born children a combination of the Binet Test and Healy

and Knox Tests would not be without value and would give us a better estimate of the child's ability and a better classification of children into groups than the Binet Test can give alone, though in such a group the difference between Binet alone and Binet plus Healy and Knox is not so great, nor is the necessity for the combination so urgent as in a foreign group where we find the coefficient of correlation increased from .44 to .56 by making use of the above combination of tests.

SECTION V

CONCLUSIONS.

From our study of a group of performance tests, it is apparent that they differ markedly.

The Mare and Foal Puzzle and the Kohs Block Test are tests which classify children more in accord with the verbal type of test such as the Binet. This makes these two tests less valuable in classifying or sectioning a foreign group of children than are the Healy Form Board and the Knox Cube Test. These tests, especially the latter, depend very little on linguistic abilities or cultural background for success, and hence are of great value in measuring the capacities and abilities of foreign-speaking children.

Any type of verbal test, such as the Binet, is unfair when used to classify foreign children. Can certain performance tests be combined with the Binet Test to measure more adequately the abilities of such a group? Our results warrant the conclusion that a grouping of unselected performance tests with the Binet Test does not accomplish this end, but that the use of the Binet Test plus the Knox Cube Test and the Healy Form Board does give a more satisfactory classification of the foreign child than

is obtained by using the Binet Test alone.

In a division of bright English-speaking children the combination of performance tests with the Binet Test is also valuable. In this group, however, it is not so essential that the performance tests be selected, but even here the combination of Binet plus the Knox Cube Test and the Healy Form Board apparently furnishes the best means for classifying the group.

This contradicts the attitude of men like Terman who consider the Binet Test sufficient for any child born in the United States, even those of foreign parentage and seems to indicate that such tests as those devised by Pintner, Knox and Healy have a necessary part in our classification of children.

PROBABLE ERROR FOR CORRELATIONS

(Bright Division)

r	P.E.	r	P.E.	r	P.E.	r	P.E.
.51	.08	.29	.10	.19	.11	.07	.11
.50	.08	.28	.10	.18	.11	.06	.11
.46	.09	.26	.10	.17	.11	.05	.11
.40	.09	.25	.10	.16	.11	.02	.11
.39	.09	.23	.10	.15	.11		
.37	.09	.21	.11	.14	.11		

(Retarded Division)

r	P.E.	r	P.E.	r	P.E.	r	P.E.
.66	.07	.52	.10	.29	.12	.16	.13
.65	.07	.46	.10	.28	.12	.15	.13
.64	.07	.42	.11	.27	.12	.14	.13
.61	.08	.41	.11	.24	.12	.09	.13
.60	.08	.39	.11	.23	.12	.05	.14
.56	.09	.35	.12	.20	.13		
.54	.09	.31	.12	.18	.13		

BIBLIOGRAPHY.

The following articles and books were used in the preparation of the study:

Arlitt, A. H.; "On the Need for Caution in Establishing Race Norms", Journal of Applied Psychology, Vol. V, No. 2, pp. 179-183.

Bridges, J. W. and Coler, L. E.; "The Relation of Intelligence to Social Status", The Psychological Review, Vol. XXIV, No. 1, pp. 1-31.

Brown, G. L.; "Intelligence as Related to Nationality", Journal of Educational Research, Vol. V, No. 4, pp. 324-327.

Dearborn, W. F.; "Repeated Measurements of the Physical and Mental Development of Children", School and Society, Vol. XX, No. 513, pp. 515-518.

Gebhardt, J. C.; "Malnutrition and School Feeding", Bulletin No. 37, 1921, Department of Interior (Bureau of Education).

Herskovits, M. J.; "Brains and the Immigrant", The Nation Vol. 120, No. 3110, Feb. 11, 1925.

Knox, H. A.; "A Scale, Based on the Work at Ellis Island

for Estimating Mental Defect", The Journal of the American Medical Association, Vol. LXII, No. 10, pp. 741-747.

Kohs, S.C.; Intelligence Measurement; a psychological and statistical study based upon the block design tests. (1923).

Luckey, B.M.; "Racial Differences in Mental Ability", The Scientific Monthly, Vol. 20, pp. 245-248.

Mazurowska, M.; "An Experiment with Otis Group Intelligence Scale", The School Magazine (Buffalo) Vol. 4, No. 10.

Pintner, R.; "The Standardization of Knox's Cube Test", The Psychological Review, Vol. XXII, No. 5, pp. 377-401.

Porteus, S. D.; "Mental Tests for Feebleminded; A New Series", Journal of Psycho-Asthenics, Vol. XIX, No. 4, pp. 200-213.

Porteus, S. D.; "Mental Tests with Delinquents and Australian Aboriginal Children", The Psychological Review, Vol. XXIV, No. 1, pp. 32-42.

Schmitt, C.; "Standardization of Tests for Defective Children", The Psychological Review Monographs, No. 83, July 1915.

Strong, E. K.; "Effects of Hook-Worm Disease on the Mental

and Physical Development of Children", Publication No. 3, International Health Commission.

Terman, L. M.; The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence. (Educational Psychology Monographs No. 18, edited by G. M. Whipple).

Binet, A. and Simon, T.; A Method of Measuring the Development of the Intelligence of Young Children. (1915) (A translation by Clara H. Town).

Binet, A. and Simon, T.; The Development of Intelligence in Children (1916). (A translation by Elizabeth S. Kite).

Brigham, C.; A Study of American Intelligence. (1923).

Gesell, A.; The Pre School Child. (1923)

Goddard, H.; Feeblemindedness; its causes and consequences (1920).

Kellogg, V.; Mind and Heredity (1923).

Matter, F.; The Unstable Child. (1924).

Miller, H.; Races, Nations and Classes (1924).

Pintner, R.; Intelligence Testing (1923).

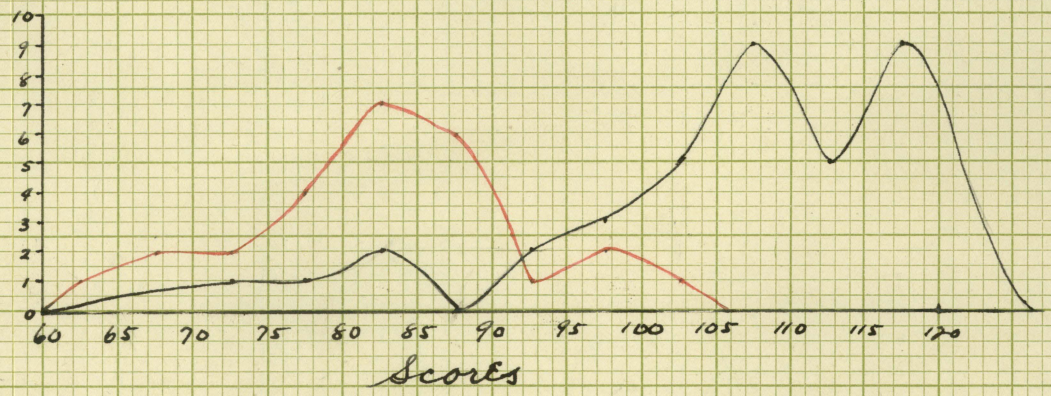
Spearman, C.; The Nature of Intelligence and the Principles of Cognition. (1923)

Terman, L.; The Measurement of Intelligence. (1916)

Terman, L.; The Intelligence of School Children. (1916)

Thurstone, L.; The Nature of Intelligence. (1924)

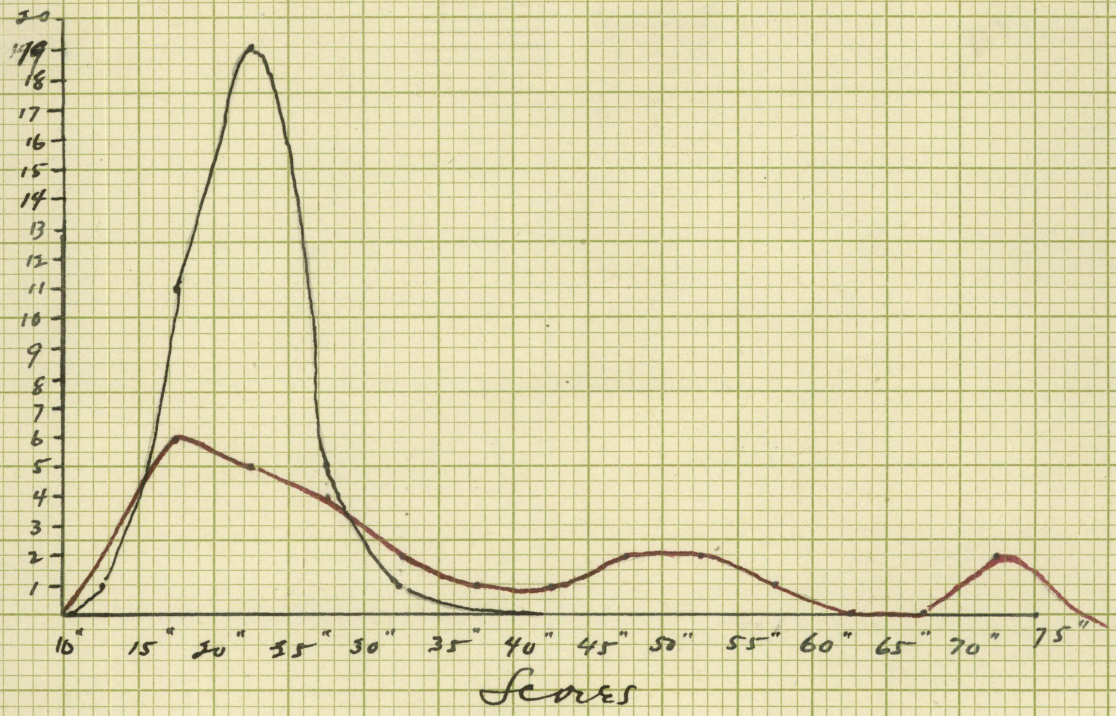
Woodrow, H.; Brightness and Dullness in Children. (1919)



Graphs Showing distribution of I.Q. ratings in Advanced and Retarded Sections

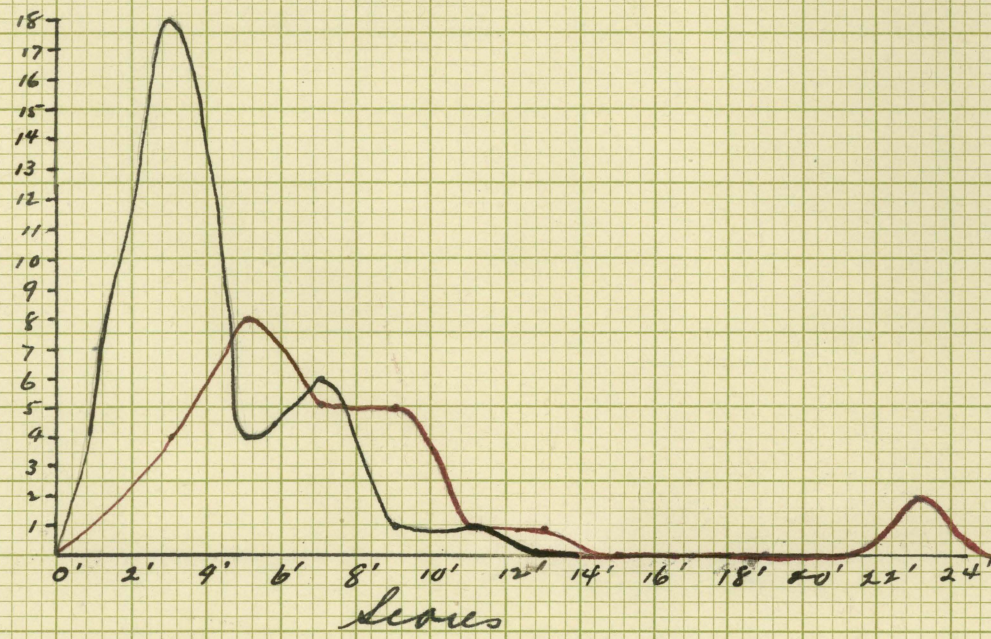
— = Advanced Group

— = Retarded Group.



Graph Showing Distribution of Pagini Foden Board Scores
— = Advanced Group
— = Retarded Group

Lowest Limit Quintiles = Highest Score

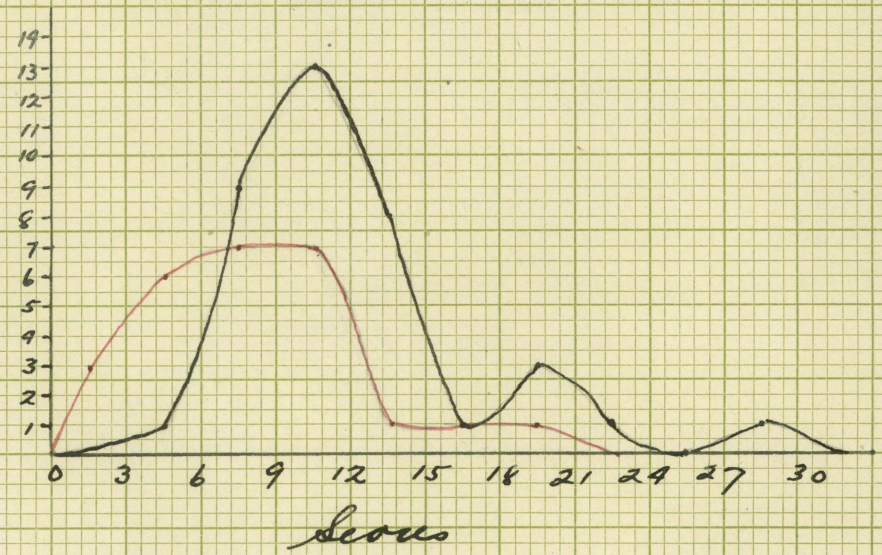


Graphs Showing Distribution of "Mark & Goal" Puzzle Scores

— = Advanced Groups.

– – = Retarded Groups.

Lowest Time Intervals = Highest Score

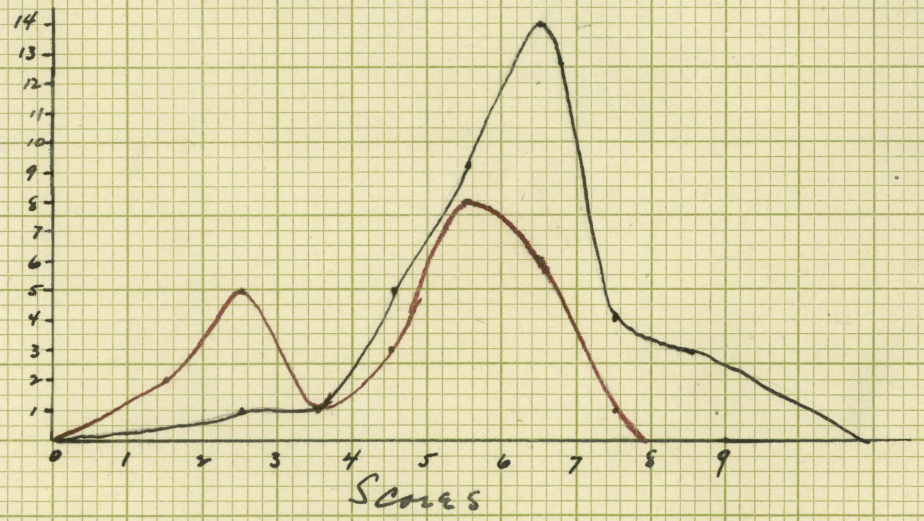


Graphs Showing Distribution of Kohs Block Test Scores

— = Advanced Group

— = Retarded Group

High Score numerically = best score



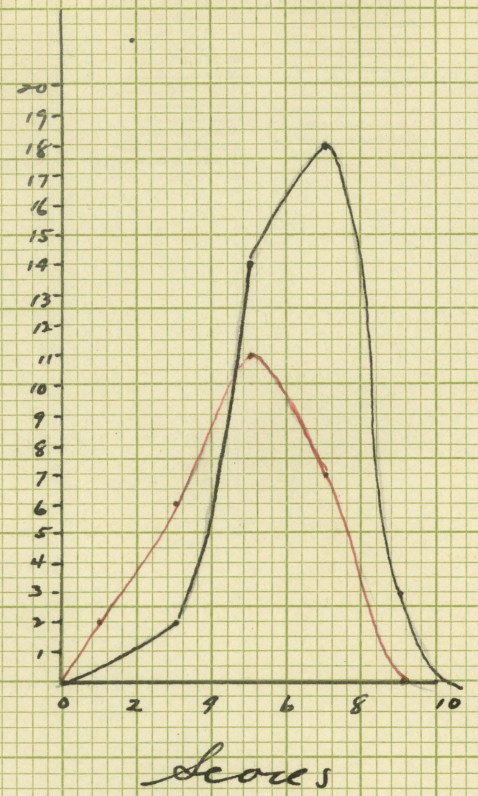
Graphs Showing Distribution of Kurr Cook Test Scores

— = Advanced Groups

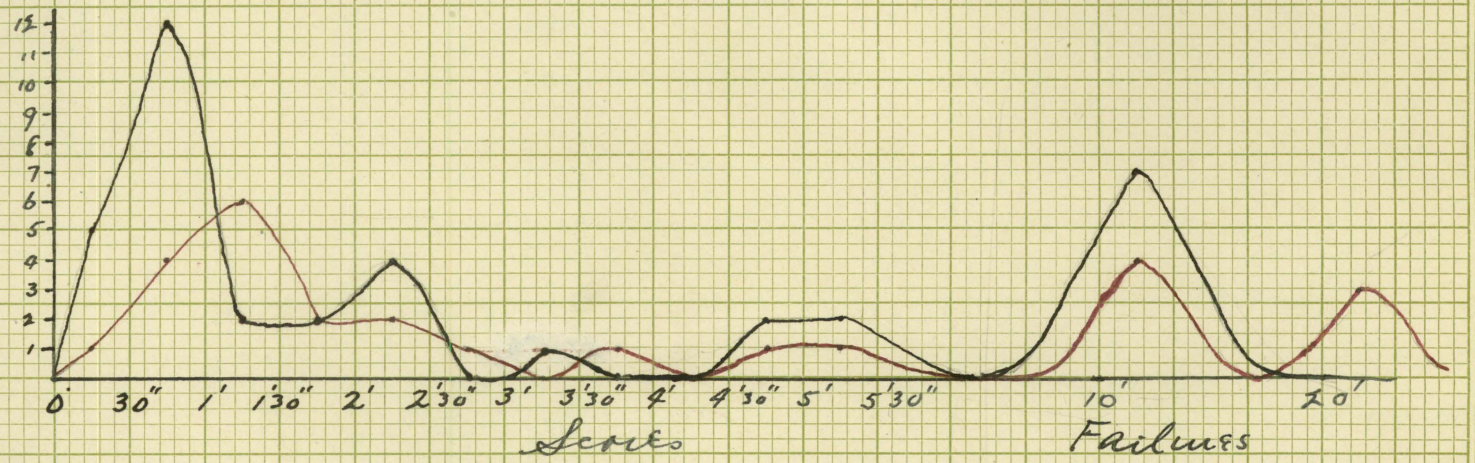
— = Retarded Groups.

(1)

High Score numerically = best score



Graphs Showing Distribution
of Knox Cube Test Scores
— = Advanced Group
— = Retarded Group

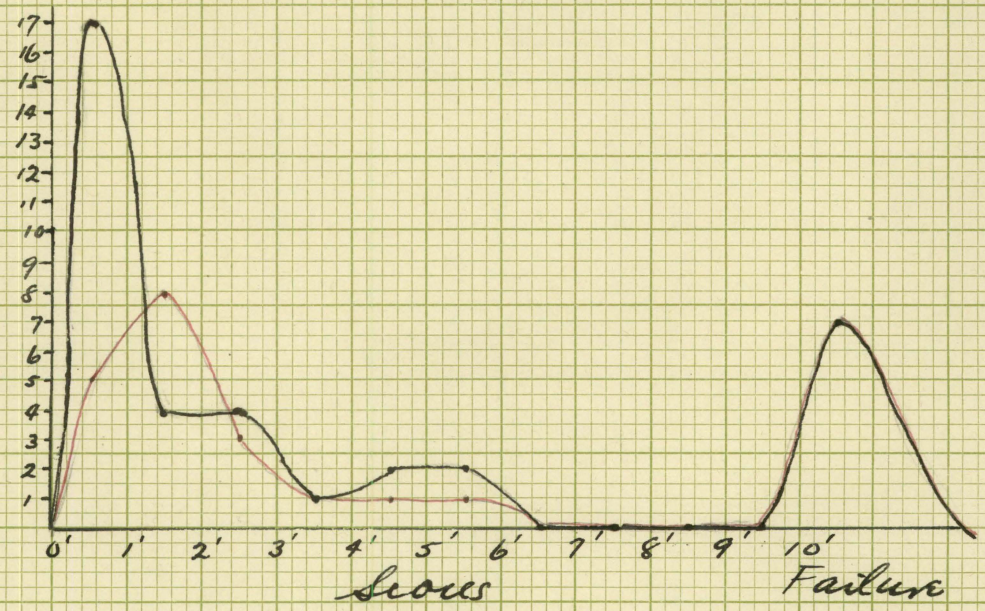


Graphs Showing Distribution of Healy "A" Form Board Scores.

— = Advanced Group

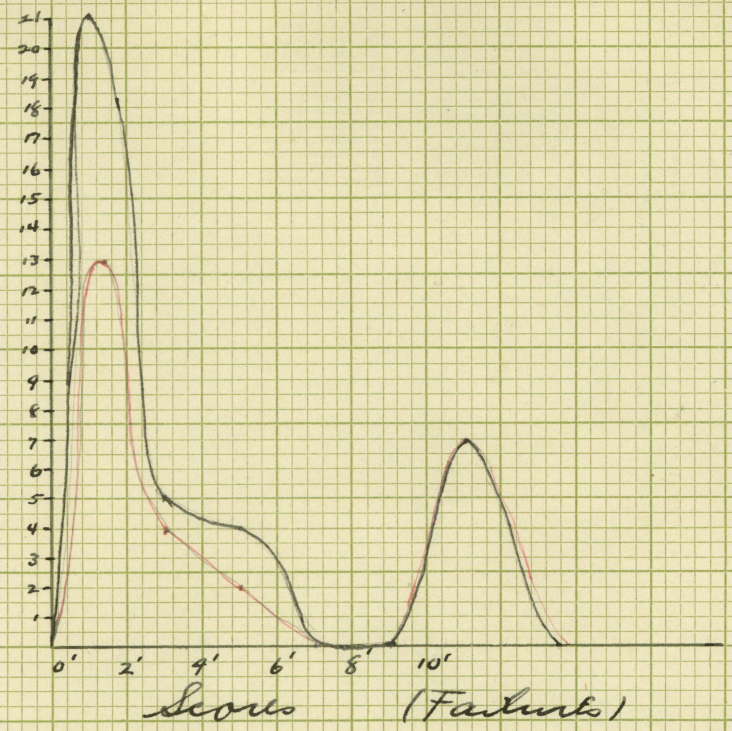
— = Retarded Group

Lowest Score Intervals = Highest Score



Graphs Showing Distribution
of Healy "A" Form Board
Scores
— = Advanced Group
— = Related Group.

(2)



Graphs Showing Distribution
of Healy Form Board Scores
– = Advanced Group
– = Retarded Group

(3)