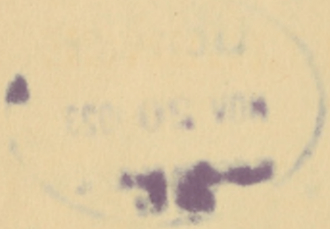
DE EDUCATIONAL MONOGRAPHS

# MEASURING INTELLIGENCE

HINES



Riverside Educational Monographs

EDITED BY HENRY SUZZALLO

PRESIDENT OF THE UNIVERSITY OF WASHINGTON, SEATTLE, WASHINGTON

# MEASURING INTELLIGENCE

BY

HARLAN CAMERON HINES

ASSOCIATE PROFESSOR OF EDUCATION

THE UNIVERSITY OF WASHINGTON

AUTHOR, A GUIDE TO EDUCATIONAL MEASUREMENTS



HOUGHTON MIFFLIN COMPANY

BOSTON NEW YORK CHICAGO SAN FRANCISCO

The Riverside Press Cambridge

The Riverside Press
CAMBRIDGE · MASSACHUSETTS
PRINTED IN THE U.S.A.

# CONTENTS

# EDITOR'S INTRODUCTION

In the period just prior to our generation, educational reform was not a vigorous concern of school teachers. Reconstruction was slow. Such modifications as were under way, were largely the product of a common sense, noting divergencies in current practices and choosing what seemed the better. Tradition was powerful, hence innovations were not sufficiently numerous to provoke much criticism, reflection, and change. It was a time of relatively stable teaching practices. There was little discontent with existing procedures and slight disposition to experiment.

Not until late in the nineteenth century did extensive and startling experimentation begin to appear. It was directly stimulated by new philosophies of educational procedure, largely brought to American soil by educational leaders who had been trained abroad. These novel doctrines ushered in a new era in American thinking about education, one characterized by changed assumptions as to the aims of education and the qualities of human nature. Many new methods

of teaching were produced by the detailed applications of these general theories.

This period of philosophic reform continued vigorously through the first decades of the present century. Foreign philosophies of education were considerably modified and indigenous theories were promulgated. Their influence is still marked with us and will continue to be. The whole movement, from first to last, was philosophic rather than scientific in its method. Such reconstructions of procedure as ensued were the products of deductive rather than inductive thinking, a reasoning down from accepted generalizations to particular methods which were to be verified by actual trial, rather than a reasoning up from concrete facts to principles and their applications. Great richness and variety in educational practice followed, stirring the minds of schoolmasters with a curiosity to know which methods were better or worse and upon what accurate and tangible grounds. The way was being paved for the coming of scientific method in education.

Scientific method had been entering the profession in a small, wedge-like way for fifteen years. Not until the last half dozen years, however, has its presence been felt with such force as

to give it a position of high respectability and wide acceptance. We may safely say that we are now at the beginning of a distinctly new period by its advent. Common sense interpretations of personal and professional experiences will still bring about modifications, and new philosophies will continue to suggest changes, but these modes of reform will be felt to be inferior to that which has the method of scientific inquiry as its basis. It is this change in professional attitude which marks off the time we are in from those which have preceded it. A new method of thinking and acting has the position of highest prestige even if as yet it influences a very small part of the field of actual practice. Faith in scientifically examined fact supplants enthusiasm for plausible philosophic assumptions. Science has arrived at the door of the public school and is walking within.

In our professional scholarship the field of the history of education was the first to respond to accurate methods of inquiry. But the history of education did not offer an influential critique of educational procedure, for few practitioners had a deep interest in this field of study. Educational administration was the next professional field to become inductive in spirit and method of inquiry. It proceeded to base its reform proposals on a

careful study of actual American conditions and at once its conclusions exercised a wide influence on every phase of school organization, administration, and supervision. Here the managers of schools were gripped with the worth of inductive study, but the teachers were still left unawakened. That awakening came later, chiefly through the scientific work of the educational psychologists, but most of all through those creative studies of human ability and development which led to the formulation of mental tests and measurements, now such important diagnostic aids in the business of schooling.

There are still many unawakened teachers who are cynically critical of intelligence tests and standard measurements. But they are not as harmful as that smaller number of enthusiasts whose acceptance has blinded them to all defects and shortcomings. Yet, neither of these groups can long hamper the power of those workers in educational research who carry scientific caution with them at every step. The measurement of intelligence and its development is one of the most promising intellectual activities within the educational corps. Every teacher, every parent and every good citizen who senses our social reliance on education should acquaint himself with its

purposes, its methods, its results and its limitations. In this we would aid them by presenting for their use a volume of small dimensions and large compass which will tell what ought to be commonly known about the most important single movement in our current educational life.

# PREFACE

THE purpose of this volume is to present a brief résumé of the aims, principles, problems, and progress of the measurement of intelligence. The writer recognizes that these have been treated in other books and in periodicals, to which frequent references have been made, and that this is an ever-changing field; but he feels that there is need for a book which will sum up the developments and which will invite attention not only to the place of intelligence tests in modern school practices but to the discrepancies in testing which all too frequently have been overlooked.

In presenting such a discussion he has attempted to strike a balance between the enthusiasts on the one hand and the ultra-conservatives on the other. Much good has been accomplished through the use of intelligence tests; likewise, some harm has been done; in neither case has our educational program been revolutionized. The pendulum of education is gradually lengthening its arc, and these newer instruments are but a sign of progress — an aid to the fulfillment of the aims and purposes of modern education.

# PREFACE

An attempt has been made to observe a simple treatment throughout, in order that those unfamiliar with measurement problems may have an easy approach to their study. In this attempt the writer is indebted to Dr. Edgar D. Randolph, Professor of Educational Sociology at the University of Washington, who read the manuscript and whose valuable suggestions made possible greater clarity of expression.

H. C. H.

SEATTLE, WASHINGTON

# MEASURING INTELLIGENCE

## I

### THE MEANING OF INTELLIGENCE

In three consecutive issues of one of our well-known periodicals fourteen educational psychologists, without conference, attempted to define the meaning of the word "intelligence." There were almost as many definitions as writers and, of these definitions, some of which were direct and some implied, no two fully conveyed the same meaning. These psychologists had been called upon to enter a discussion as abstract as politics or religion and they reacted as all human beings must react when confronted with a like situation.

The reason for this is that no one knows precisely what intelligence is. No one even knows the exact nature of any one of its separate faculties or functions. Aside from the formal definitions expounded in recent years, it has been described as "understanding" or the "capacity to understand," "knowledge" or "information,"

"intellect," "mental acuteness," and so on. Ebbinghaus had called it "intellectual ability" and had said that it consists in the "elaboration of a whole into its worth and meaning by means of many-sided combination, correction, and completion of numerous kindred associations." To the person unfamiliar with psychological phraseology such a definition is not very clear, but it is interpreted to mean that intelligence has the general character of a combination activity, that every true instance or evidence of it may be reduced to an act of combining.

Spearman looked upon intelligence as a general function, the outward manifestation of a high correlation between each of the separate functions, while Thorndike has held that it is a multitude of functions "each of which involves content as well as form and is so related closely (as Spearman has suggested) to only a few of its fellows, to the others with greater and greater degrees of remoteness." Meumann denies that there exists a faculty known as general intelligence, while Binet describes intelligence as having three characteristics of the thought processes, namely, that it (1) tends to take and maintain a definite direction; (2) has a capacity to make adaptations for the purpose of attaining a desired end; and

(3) has the power of auto-criticism. Stern calls intelligence a general capacity which consciously adjusts the individual's thinking to new requirements — general mental adaptability to new problems and conditions of life.

The task of defining intelligence, in such a way that the definition will be acceptable to all minds, is as difficult as an attempt to define life, itself, or a living wage. It is not understanding alone nor the capacity to understand, for it includes the salient factor of being able to put understanding into action. It is not knowledge or information, for knowledge or information can be but one of the component parts of intelligence. It is not synonymous with "intellect," for, in the common usage of the terms, an individual may be highly intellectual (have a knowledge of many things and be able to discuss them in high-sounding phrases) and not be highly intelligent when he tries to apply in a practical manner those things which he knows.

Perhaps a better definition is "mental acuteness." This implies sharpness of the mind, a certain ability to discern, a sagaciousness, an ingeniousness, a power to make fine discriminations. In other words, it is, to use a synonymous term, "mental ability" — what Stern describes

as "general mental adaptability to new problems and conditions of life."

Those who choose to make "intelligence" synonymous with "common sense" may not be too far wrong. However, the scientist prefers to speak in scientific terms, and the modern psychologist, confronted with the necessity of making a choice between the scientific and popular terminology, probably would describe intelligence somewhat as follows: "What we mean by intelligence is intellectual ability. Ability is the product of experience acting on capacity. Capacity denotes native rather than acquired receptive powers, i.e., a capacity to acquire capacity. Intelligence, therefore, is the power to receive into, plus the power to perceive through, the intellect. Such a definition as 'common sense' is non-technical, unscientific, and empirical."

Yet it may be argued that this is but another way of defining "common sense." "Sense," used in a popular way to describe intelligence, is "perception through the intellect." "Common" implies that it is shared similarly by two or more powers, states, qualities, individuals, parts, or divisions. Hence, common sense is the outward evidence of the unity of all the sensations into the expression of a general sensation, ability to esti-

mate the strength of which gives the power to discriminate between kinds and degrees of intelligence.

But, before it is necessary to settle upon a definition to guide the discussion of the instruments for measurement, their uses, their abuses, and their values, it is well to pause to consider why it is we are concerned with the problem of intelligence. Dewey has shown us that "if a man's actions are not guided by thoughtful conclusions, then they are guided by inconsiderate impulse, unbalanced appetite, caprice, or the circumstances of the moment." The best evidence of intelligence is the evidence of thoughtful conclusions expressed in the written or spoken word. A high degree of intelligence has come to be thought of as being qualified by the amount, style, form, and quality of such expression. To the development of this high degree of intelligence our system of modern education has been directed. There is a desire to develop the thought processes. There is also a desire to develop the skills. The one is thought of as an accomplishment in the abstract, the other an achievement in the concrete. One is no more important than the other; it is best that they accompany each other.

It may be objected that it is possible, while dealing with some types or "levels" of mentality, to develop only skills, skills which are expressed through motor activities; that there are those whom we may describe, for want of a better term, as "motor-minded." Yet there is no scientific proof that the nerve-energy which has gone to build the "motor-mind" has lost caste. There has been too much of an inclination to base estimates of intelligence on what has been chosen to be called the evidences of abstract thinking. Unusual as it may seem, the expressions of intelligence which have been the easier to estimate have received less attention, are less certain, are less familiar in pedagogical studies than those which present a vague aspect and mysterious appeal.

One of the most popular conceptions of intelligence is that it represents amount of general information. It was stated above that information can be but one of the component parts of intelligence. In attempting to estimate the amount of general information possessed by individuals or groups, it is immediately evident that it is impossible to secure a result which may be definitely classified as a "degree." This condition results from the fact that no individual

knows exactly how much he does know and it is impossible for him, at any specified time, to exhibit all he knows, although at times such a thing may have seemed to have occurred.

There are those who have come to believe also that a high degree of intelligence represents a high degree of education, of school training. Take, for instance, the case of a certain father and son. The father is a farmer with only an eighth-grade education. The son is a successful administrator, a graduate of two higher institutions of learning, and writes his degrees in seven letters. In the amount of school training, it is evident, they are very unalike. In achievement, the father, through his ability to raise the finest hogs in his state, is as highly successful in his way as the son is in another way. Setting aside the academic verdict that the son has attained the greater goal, the situation provokes some timely questions. Is the father less intelligent than the son? Has he had an inferior training? Would the father have been more intelligent had he had the advantages of the son's schooling? Would the son have been less intelligent had he chosen to be a farmer?

These are the types of questions which it will be necessary to answer if an acceptable definition

of intelligence is to be formulated. The definitions offered have been prepared by those who, it is unanimously conceded, have reached the full meanings of their definitions. But the criteria on which such definitions have been based have been established too frequently within the shadow of school walls. With such a premise, some attention should be given to the more recent definitions of intelligence referred to at the beginning of this chapter.

Stating that he realizes that definitions and distinctions are pragmatic, Thorndike defines "intellect in general as the power of good responses from the point of view of truth and fact. The power of good responses to abstract qualities and relations rather than gross total facts and to ideas rather than direct experiences may be called the more intellectual variety of intellect." Terman says that " an individual is intelligent in proportion as he is able to carry on abstract thinking." Colvin further extends Terman's definition by saying that " an individual possesses intelligence in so far as he has learned, or can learn, to adjust himself to his environment." Pintner modifies Stern's definition to read: "Intelligence is the ability of the individual to adapt himself adequately to relatively new situa-

tions in life." Henmon limits the definition to three words: "Intellect plus knowledge," while Woodrow states that it is "the capacity to acquire capacity." Haggerty differentiates between "high" and "low" intelligence by saying that "low intelligence means simple occupations and crude civilization" and "high intelligence means the possibility of efficiency and leadership in the more complex problems of civilized society."

In elaborating upon the definitions, most of the psychologists quoted above are inclined toward the belief that intelligence is determined more largely by training than by native ability. In that conclusion there is widespread agreement; but there is an ever-recurring question as to what is meant by "training." Are there two types of education — one which may be achieved in school and one which may be attained outside of school? Or must all of life be thought of as a school in which individuals and groups are trained to become working units of society?

It is not assumed that Haggerty meant to infer that simple occupations are always performed by persons of "low" intelligence, nor that positions of leadership are always held by persons of "high" intelligence. And it is necessary to

understand that Terman recognizes various degrees of abstract thinking, produced, no doubt, by the life experiences of the individuals representing those degrees. If it is admitted that the farmer would have been more intelligent had he received more school training and the son would have been less intelligent had he chosen to drop school for farming, intelligence cannot be native endowment alone but must and does include school training. Such questions, answered in the negative, tend to limit intelligence to inherent ability.

Whether intelligence is a general capacity or whether it can only be thought of as a multitude of minor abilities (native or acquired), it does exist. The difficulty in defining it satisfactorily doubtless lies in the failure to accept the evidences of its "spread" from ideas to direct experiences, from the abstract to the concrete. It will make a difference, also, whether the definition is to include the intelligence of society in general or whether it shall be limited to the intelligence of the school-trained. Since our school system is organized to train pupils to become worthy members of society, it seems that we should have a definition which will approximate the inclusion of all types and conditions of human mentality.

## THE MEANING OF INTELLIGENCE

Taking all the definitions which have been submitted, it should be possible to formulate a single one which will be acceptable to the majority of those interested in the problem. In order that a working hypothesis may be established in as simple terms as possible, it may be stated that *intelligence is the capacity* (native ability plus training) *of an individual to adapt himself to a new situation.* This is but another way of saying what Stern, Terman, Colvin, Pintner, Haggerty, and other psychologists have already said. It does not leave out "intellectual ability" or "common sense." It includes the farmer with his eighth-grade diploma as well as the administrator with his university certificates. It is equally applicable to all types and conditions of mentality.

The common tendency to think of "training" as school training only cannot be other than objectionable. There is a type of training that is frequently overlooked, especially by those psychologists who subscribe to the theory that children do not show marked differences before reaching the age of adolescence. That training, in various descriptions, is secured in the home, particularly during the time between birth and the first attendance at school. How large a factor

such a period of training may be in the development of intelligence has not been, and may not be, determined. It is emphasized, however, by the studies which have been made of the influence of social factors on general mental ability.

The term "degree" has been used a number of times in this discussion. By speaking of "degree of intelligence" it is implied that intelligence can be measured; it exists, and the psychologist maintains that whatever exists at all exists in some amount. It follows that whatever exists in some amount can be measured. This contention is open to question. Take, for instance, the part of intelligence known as "general information." Because it is impossible for an individual to exhibit at any one time all of the general information in his possession, any measure of it would not be complete. Is a measure which is not complete a scientific measure? That question may well be carried through the chapters which follow. But it will perhaps be necessary to amend the previous statement to read: "whatever exists in some amount can be measured in some amount."

Going back to Dewey's thesis that a man's actions should be guided by thoughtful conclusions, and to the statement that these thoughtful

conclusions expressed in the written or spoken word are the best evidence of intelligence, it will readily be seen that psychologists have been afforded an excellent opportunity to measure, if possible, what has been done toward the development of thoughtful conclusions, or "abstract thinking." Their researches have been largely carried on in connection with the work of the public schools, with some special excursions into other fields. What they have achieved is small only in comparison with that which remains to be done. The measuring instruments which they have evolved in the past few years are worthy of careful consideration by all those who are interested in the welfare and future of the children in our schools.

REFERENCES

DEWEY, J., *How We Think*, chap. 5.

GREGORY, C. A., *Fundamentals of Educational Measurement*, chap. 3.

MCCALL, W. A., *How to Measure in Education*, chap. 1.

TERMAN, L. M., *The Measurement of Intelligence*, chap. 3.

STERN, W., *Psychological Methods of Testing Intelligence*, chap. 1.

THORNDIKE, E. L., and thirteen others. "Intelligence and Its Measurement," *The Journal of Educational Psychology*, March, April, May, 1921.

## II

### THE INSTRUMENTS FOR MEASURING INTELLIGENCE

ALTHOUGH experimentation with rather crude tests to determine innate abilities had been carried on by psychologists throughout the last half of the past century, the first practical test of intelligence did not appear until 1905 when Professor Alfred Binet, a French psychologist, and Dr. Th. Simon, a French physician, devised a scale known as the Binet-Simon Tests and intended to be used in separating the subnormal from the normal children in the public schools of France.

The Binet Scale, as these tests have been popularly called, included tests of familiarity with common objects, a test for memory span, and tests involving comparison and judgment. The first draft of the scale contained thirty tests, followed by a revision in 1908 in which twenty-six new tests had been added and the tests arranged for the ages from three to fourteen, and followed by another revision in 1911.

For use in America a number of revisions of

the Binet Scale have been made. Goddard, while still at Vineland, New Jersey, made some minor changes; Kuhlmann revised the tests for work with subnormal children; Yerkes devised a point scale based on the Binet; and Herring has published a new revision with the purpose of shortening the time and further simplifying the procedure of giving. The best known of the American revisions, however, is that made by Terman at Stanford University, who undertook the task in 1913 and published the results in 1916. In the Stanford-Binet there are ninety tests (thirty-six more than in the 1911 Binet Scale), six for each age level from three to ten, eight for the age of twelve, six for the age of fourteen, six for average adult level, six for superior adult level, and sixteen alternative tests to be used when a test has been rendered undesirable by coaching or otherwise.

Binet had held that intelligence tends to take and maintain a definite direction and the score he was able to secure by his tests was expressed in terms of "mental age." For instance, a child whose actual age is ten might, when tested by the Binet Scale, show a mental age of twelve, or two years accelerated. Expressed simply, the mental age of a child is his score on the test divided by

the number of months in a calendar year. To classify a child according to his test score the difference between his chronological age and his mental age was recorded, but this resulted in a fictitious classification of older children.

Terman, in order to obviate this difficulty, transcribes the test result in terms of the intelligence quotient (I.Q.). This is obtained by dividing mental age by chronological age. In the case of the child of ten who showed a mental age of twelve, the I.Q. would be 1.20 or, as it is more commonly designated, 120. This makes it possible to classify a child on the basis of his mental ability irrespective of his actual or chronological age. Terman is inclined to the belief that the I.Q. remains constant from year to year and is testing children through consecutive years in order to verify such belief.

In order to test the mental ability of non-English-speaking children, children with a limited English vocabulary, and children with speech defects, a type of test much different from the Binet has been devised. It is known as the "performance test" and has taken several forms. Perhaps the best known of the performance tests is that devised by Seguin and modified by Goddard and others. It is called the form-board and

tests the ability of the child to fit wooden blocks of various shapes into forms cut to receive them. A variation of this test is the Healy manikin puzzle in which the parts of a figure must be fitted together to complete the manikin. Still another type of performance test is the picture puzzle test, similar to the manikin puzzle in that the parts must be fitted together to make the original. The maze test, taken from the laboratory of animal psychology, has been widely used, the best example of which is the Porteus Maze Test. In the maze test the subject, sitting before a printed form of the maze, is required to trace with pencil the correct way of going through and getting out of the maze.

All of the tests mentioned thus far have been designed to test the intelligence of individuals, and no group test of intelligence had been published prior to 1917. During the years just preceding America's entrance into the World War, Otis, who had been trained in testing by Terman, had been attempting to devise a test which would make it possible to test several or many individuals at one time. The manuscript he had prepared was taken over by the American Psychological Association and adapted to the needs of group testing in the Army. It under-

went several revisions but was finally issued under the name of Army Alpha. There were eight separate divisions, some of which were based upon the principles of the Binet Scale, designed to be used as follows:

(1) to test the ability to carry out commands
(2) to test the ability to work reasoning problems in arithmetic
(3) to test the ability to select the best reason for a statement
(4) to test the ability to detect likenesses or differences
(5) to test the ability to reorganize disarranged sentences and to indicate whether a sentence was true or false
(6) to test the ability to complete series of numbers
(7) to test the ability to select by analogy
(8) to test the range of general information

Complete directions for giving and scoring the tests were supplied to examiners and the results of the tests were intended to be used in the classification of literate and English-speaking soldiers in the American Army.

Another group test which made its appearance at that time was Army Beta. It was of the non-verbal, performance type, and was given to all soldiers who had not completed the third grade in school, who could not speak or understand English, or who were illiterate. It was also given

to those who failed to make a satisfactory showing on Army Alpha. If a subject could not make a required score on either Alpha or Beta, he was given an individual examination such as the Stanford-Binet, the Point Scale, or the Performance Scale.

These army tests were the forerunners of a number of group tests, some verbal and some nonverbal, prepared for use in schools and colleges. A number of the men who had helped to devise the army tests turned their attention to the composition of similar tests which might find practical use in the various school divisions. Among these men were Otis, Whipple, Haggerty, Pintner, Thurstone, Terman, and Thorndike. Others interested in psychological research turned their attention to the improvisation and standardization of group tests, and there are, at the time of this writing, thirty-six such tests, most of which are intended for school use.

Among the notable examples of these tests of intelligence are the Detroit First Grade Intelligence Test, Haggerty's Intelligence Examination Delta 1 for the primary grades, Otis's Group Intelligence Scale Primary and Advanced, the National Intelligence Tests, Terman's Group Test of Mental Ability, Otis's General Intelli-

gence Examination, Otis's Higher Examination for high schools and colleges, Thurstone's Psychological Examination for college freshmen and high-school seniors, and Thorndike's Intelligence Examination for High-School Graduates. Most of these tests were either modeled after Army Alpha with certain revisions to make them applicable to school conditions, or were constructed in such a way that the faults of the army test might be eliminated.

The Detroit First Grade Intelligence Test, devised by Engel, consists of ten separate tests covering the following subjects: (1) information; (2) similarities; (3) memory; (4) absurdities; (5) comparisons; (6) relationships; (7) symmetries; (8) designs; (9) counting; and (10) directions. It is intended to include the testing of all the types of abilities exhibited by first-grade children.

Haggerty's Intelligence Examination Delta 1 consists of six tests with a fore-exercise for each test. The six tests are designed to measure: (1) ability to follow directions; (2) ability to copy designs; (3) ability to complete pictures; (4) ability to compare pictures as to likenesses or differences; (5) ability to accompany symbols with certain digits; and (6) ability to compare words as to sameness or oppositeness in meaning.

The fore-exercise is included to equalize, through practice, the experience of the children with the test.

The Otis Primary Examination consists of eight tests which do not involve the ability to read. In the Advanced Examination there are ten tests given the following titles: (1) following directions; (2) opposites; (3) disarranged sentences; (4) proverbs; (5) arithmetic; (6) geometric figures; (7) analogies; (8) similarities; (9) narrative completion; and (10) memory. The Otis General Examination and the Otis Higher Examination are "scrambled" tests, covering the same abilities tested by Army Alpha, but in which the test divisions do not appear.

The National Intelligence Tests, prepared by Haggerty, Terman, Thorndike, Whipple, and Yerkes, consist of two groups of tests, each with five test divisions. Scale A is composed of tests of arithmetic reasoning, sentence completion, logical selection, ability to compare words as to sameness or oppositeness, and ability to accompany symbols with certain digits. Scale B is composed of tests of computation, range of information, vocabulary, selection by analogy, and comparison. Like Haggerty's Delta 1, each scale includes a fore-exercise for each test division.

The Terman Group Test of Mental Ability contains ten separate tests classified as follows: (1) information; (2) best answer; (3) word meaning; (4) logical selection; (5) arithmetic; (6) sentence meaning; (7) analogies; (8) mixed sentences; (9) classification; and (10) number series completion. This test represents an effort to produce, in so far as possible, a group test the results of which will be comparable to the results secured by the Binet Scale.

The Thurstone Psychological Examination is also a "scrambled" test covering somewhat the same abilities tested by Army Alpha and similar in form to the Otis General Examination but quite different in content. The Thorndike Intelligence Examination is comprised of two examinations of the type of Army Alpha, but extended and made more difficult. In both of these tests an attempt is made to measure the abilities expected to be displayed by high-school seniors or college freshmen.

In these group tests of intelligence attempts have been made to measure twenty-eight different types of mental abilities. The test divisions which appear most frequently are those covering general information, similarities, carrying out directions, reasoning problems, and analogies.

These appear in four of the tests discussed. Next in order are the tests covering logical selection and the reorganization of mixed sentences, each of these types of tests appearing in three of the series. Those abilities which are tested in at least two of these tests of general intelligence are: memory, ability to compare objects, ability to trace designs, ability to complete series of numbers, ability to accompany symbols with certain digits, ability to detect sameness and oppositeness, and the ability to complete pictures.

The completion test, adapted from Ebbinghaus, appears in four different forms in five different tests. Ebbinghaus, in 1905, devised a test which consisted of a paragraph in which words with syllables omitted were presented to the subject, who was required to fill in the omissions. In Army Alpha and the Terman Group Test this takes the form of number series completion; in Haggerty's Delta 1 it is the picture completion; in the National Intelligence Tests it is sentence completion; and in the Otis Advanced Examination it is narrative completion. The Completion-Test Language Scales, worked out by Trabue and reorganized into a single test by Kelley, represent another adaptation of the Ebbinghaus measure

especially well standardized to test the ability to complete sentences.

In nearly all of these tests there has been exhibited a desire to measure, if possible, the *general* intelligence of individuals or groups of individuals. In the effort to do that instruments of various kinds have been devised, ranging from those which test roughly but one ability to those designed to test as many as ten abilities, and frequently differentiated in their powers of analysis by the limitations of certain age or group levels. Beyond this, there is a type of test, such as the Myers Mental Measure, which is definitely set out to measure intelligence at any age or in any progression of training.

There are certain other tests which cannot be classified as tests of intelligence but which are designed to measure certain fundamental human traits closely allied with intelligence as conceived by the test-maker. One of these is the Downey Will-Temperament Test, issued for use both with individuals and with groups. This test is largely a study of variations of handwriting under controlled conditions, and it is said to measure such factors as speed of decision, coördination of impulses under the mental set of both speed and accuracy, ability to maintain high speed, ability

to inhibit a motor impulse, ability to disguise and imitate handwriting, care in details, assurance, amount of motor impulsion, resistance to opposition, and perseverance. The total score on the test is supposed to represent the will tendencies of the individual.

Some psychologists have turned their attention to the measurement of character traits. Interest in this subject is doubtless the result of investigations along two different lines: (1) the need for a further measure of character to accompany the measurement of intelligence; and (2) the need for a standardized test which will supply a satisfactory substitute for the many rating scales now being discarded. Among the tests devised to measure character are the Voelker Test and the Liao Tests.

The Voelker Test was used to measure: the tendency to exaggerate, suggestibility, willingness to accept help in the solution of a problem, punctuality in returning a borrowed object, honesty in money matters, willingness to accept a "tip," and truthfulness under various conditions. The Liao Moral Judgment Scale requires the subject to indicate the best reason of five reasons given for the truth of a statement. One of the five reasons is a moral reason, the other four

being of a general or personal character. Liao also included a vocabulary test of fifty words to be checked if familiar enough to be used in a sentence or to define, and the last ten words checked were to be defined.

Still another type of test used in connection with intelligence tests is represented by the Stenquist Mechanical Aptitude Test. It is composed of two tests, one containing pictures of common mechanical objects, with questions about relationships, and the other containing pictures of machines, with similar questions. Largely a subject-matter test, it has been used to measure the mental abilities of those who are sometimes spoken of as "mechanically minded." The other extreme of this type of test is represented by the Chapman Trade Tests, tests of finished ability.

All of the tests described have been prepared with great care. Some of them represent the fruits of two or three years of constant labor. Many of them have been used very widely. Terman estimates that the Binet tests are being given in the United States at the rate of a quarter of a million a year and that for the year 1920–21 probably not less than two millions of children were given group tests of intelligence. In view of these facts, it is well to turn our attention to

what is being done with the results of testing, what is really being tested, and what, if anything, can be done to make the tests more reliable.

REFERENCES

COLVIN, S. S., "Principles Underlying the Construction and Use of Intelligence Tests"; *Twenty-First Yearbook*, chap. 2.

GREGORY, C. A., *Fundamentals of Educational Measurement*, chap. 4.

HINES, H. C., *A Guide to Educational Measurements*, Part II.

PRESSEY, S. L. and L. C., *Introduction to the Use of Standard Tests*, chap. 11.

TERMAN, L. M., *The Measurement of Intelligence*, chap. 3.

# III

## THE USE OF THE INSTRUMENTS

SINCE Binet had standardized his series of tests by recording the reactions of individual children in the public schools of France, it is natural to expect that the American revisions of his scale would be founded upon the same principles and practices. The public school affords an excellent laboratory for experimentation of this nature. Backed by a disciplinary control that makes performance obligatory, when once the testing method has been accepted into the general program of procedure, it has not been especially difficult to secure the kind of responses needed to throw some light on the nature of general intelligence.

The way to a program of general testing has not always been clear, however. Administrators have been slow to accept the theory that intelligence could be measured, especially in the sense that certain physical qualities are measured by the graduated standard and the anthropometric scales. Teachers, also, frequently have been averse to the methods presented by psychologists

and have clung tenaciously to the time-worn prerogative that teachers, as such, are the persons best qualified to pass upon the intelligence of school children and that personal estimates are sufficient to meet all the exigencies arising in that field.

It is doubtful whether psychologists would have been able to secure an early entrance into such a laboratory had it not been for the small school of educational psychologists who, through their training in the two fields of psychology and pedagogy, were enabled to put into practice the theories advanced by those more narrowly confined to one of the two fields. These educational psychologists interested themselves, as Binet had done before them, in the segregation of the subnormal from the normal. In doing that, they did not eliminate the teacher from the problem but, quite to the contrary, drew her into it by requesting that she designate those children whose school work was of such poor quality that they might be suspected of mental deficiency. Binet had gone about this problem the other way by asking teachers to point out those whom they considered to be the most intelligent. The difference in procedure is greater than would at first be suspected. No great difficulty is encoun-

tered in detecting the mentally deficient through personal observation, allowing, of course, for those exceptions which arise through failure to discriminate between physical and mental backwardness. However, the teacher frequently misses the mark in the selection of bright children because brightness and performance of school tasks often do not go hand in hand.

It is assumed that the most important problem in intelligence testing is the location of the feeble-minded, i.e., those individuals who are not capable of normal mental development. Teachers recognized the presence of such children in the classroom many years before intelligence tests made their appearance. Not knowing the degree of defect, however, nothing was done toward the modification of content or method necessary to the training of such children. They were accepted as part of the teaching burden; in the early years of their training they sank or swam according to their strength, but usually they sank. To keep them from sinking, so long as they are capable of accepting training of any kind, is doubtless within the province of the intelligence examiner.

To test for feeble-mindedness in its varying degrees requires the use of an intensive individual

examination. In the moral background of testing, in order to remove, if possible, the stigma which often attaches to mental deficiency, no stone must be left unturned in the effort to classify the subject in the division where he rightfully belongs. The enormity of this problem is not usually understood by those who oppose intelligence testing or by those who have superficially skimmed the surface. Nor is it generally understood how fairly the classification is made by examiners of long experience.

It has been pointed out that the Stanford-Binet has been accepted as the most reliable revision of the Binet Scale for American use. The main reason for this is that it is an intensive individual examination. It is more nearly equally accurate at all points than any individual test yet produced. The problems it includes have been designed to "test native intelligence" primarily and "not school knowledge or home training. How much the child has learned is of significance only in so far as it throws light on his ability to learn more." To secure a measure of that general ability, the scale provides for testing differences in memory, differences in the power to reason, ability to compare, power of comprehension, time orientation, facility in the

use of number concepts, power to combine ideas into a meaningful whole, the maturity of apperception, wealth of ideas, knowledge of common objects, extent of vocabulary, etc.

As was explained, the results are interpreted by the use of the I.Q. It represents the subject's "general native ability." It is argued that should the I.Q. be below 80, the child should, if possible, be given special instruction in a class organized for such purposes or, at least, given a type of instruction best fitted to his ability to learn. This practice has been followed in nearly all school systems where the Stanford-Binet has been given. By separating the subnormal from the normal or superior, the burden of the teacher has not only been lessened but the child has been given an opportunity to learn all he is capable of learning while a member of the school community.

Upon the introduction of the Stanford-Binet it became possible for the school child to make a much higher score on an individual test than had previously been possible. As a result, intelligence quotients of 140 and above have been recorded by many examiners. A quotient between 120 and 140 is an indication of "very superior intelligence," while those making above 140 are classi-

fied as in the "genius or near genius" group. Terman has succeeded in arousing a great interest among educators in this type of child and has been engaged in a special investigation outlined to search out the especially gifted and make suggestions for their training. In some cities special classes have been established to care for those of "superior intelligence" or above.

It is of interest in passing to note the types of data which Terman and his four full-time assistants are attempting to secure about those who are classified as "genius or near genius." They fall into eleven divisions as follows: (1) the results of at least two intelligence tests; (2) the results of achievement tests in all the main school subjects, involving altogether three hours of testing; (3) a general information test of about an hour's duration; (4) a two-hour test of certain moral and emotional traits; (5) about twenty anthropometric measurements; (6) a record of all books read during two months, together with the subject's rating of each book; (7) a test of interest in and knowledge of ninety typical plays, games, and amusements permitting the computation of deviations from age and sex standards; (8) ratings from parents and teachers on twenty-five traits, "by a much better method than formerly used";

(9) data to be supplied by the parents in a sixteen-page Home Information Blank, by the teachers in an eight-page School Information Blank, and by the subject himself in a four-page Interest Blank; (10) home and neighborhood ratings on the Whittier Scale; and (11) of a small number of highest cases more intensive studies will be made, especially with reference to heredity.

It will be seen that the problem has grown from the segregation of the subnormal to the segregation of both the subnormal and supernormal. The results of such attempts are only reliable when the individual examination is used, and then always qualifiedly so. To give individual examinations, however, to all school children is a task examiners have not been prepared to meet. From twenty-five to ninety minutes is consumed in giving the Stanford-Binet to an individual child, depending upon chronological age and mental ability. Allowing an average of one half hour for each child, and working at top speed, not more than twenty children can be examined by one person in one day. Hence, psychologists saw the necessity of devising a group test which, they felt, would approximate the results of the individual examination in a much shorter time. Army Alpha was the out-

growth of such a necessity, as were the tests which have been patterned after it.

That intelligence, however it may be defined, takes on a different character with increasing chronological age, is emphasized by the different types of group tests it has been necessary to prepare. But it is recognized by workers in the field that the group test has not been an altogether satisfactory substitute for the individual examination, largely because it is not so intensive and precise in its applications. As Thorndike has pointed out, the group test only measures the ability or abilities of the group, and its results cannot be so refined as those which are secured through the use of the individual test. The difference in the application of the two types of tests is somewhat the same difference that exists when the distance of one hundred yards on the ground is measured both by a measuring tape and by "stepping it off." Because of a lack of time and sufficiently reliable instruments, "stepping off" of group intelligence has become necessary.

Two types of group tests have been employed in the examination of children in the primary grades. Pictorial tests, requiring comprehension of oral language, have been used extensively. A notable example of the pictorial test is the

Detroit First-Grade Intelligence Test previously described. That test was given to ten thousand first-grade pupils entering the Detroit schools during the school year 1921–22. Berry reports that, as a result of the findings, the ten thousand children were classified in three groups, X, Y, and Z. He made a thorough study of the intelligence rankings in relation to promotion, attendance, teachers' judgments, and nationality, and offers the following conclusions: (1) the pupils were classified with sufficient accuracy to be of marked help to the teachers; (2) the classification of pupils by means of intelligence ratings greatly increased the teacher's interest in the individual pupil; and (3) it is essential that different methods of instruction be developed for the bright and dull pupils. Frasier, in 1922, directed the giving of the Detroit, Dearborn, and Cole tests to three thousand Denver first-grade pupils. He reports that "the teachers like the Detroit Test best of the three," but that "it failed to distribute the upper ranges of intelligence. The Dearborn distributed the upper ranges better than the lower," while the Cole "distributed both ranges in quite a satisfactory manner."

The other type of group test for primary children is intended to test linguistic capacity.

Many of that type have been devised, but it has been found that children in the first grade do not perform so well when confronted with words as when confronted with pictures. Directions are more difficult to follow and the marks that children make on the test papers are not easily interpreted. Haggerty's Delta 1 is a combination of the pictorial and linguistic tests, and has been used widely in the classification of children in the first three grades, especially in the Virginia Survey, and in the cities of St. Louis, Cincinnati, and Minneapolis.

The plan of procedure following the giving of such tests is to divide the children of a given grade into three groups, the slow, normal, and fast, and to develop methods of instruction for the bright and dull groups. In Detroit the supervisory staff of the elementary grades have developed a little book, *Toys and Plays*, which is used in teaching the "Z," or slow, group of pupils. Certain indirect results of testing primary children should not be overlooked, however. The increase of the teacher's interest in the individual pupil is given as a strong argument for testing. Again, the score a pupil makes on an intelligence test has served as an aid in determining whether or not he shall be promoted.

Further, it is claimed, some light has been thrown on the nature of habit formation.

While testing for intelligence among primary children is a difficult task, the problem becomes simpler as the tests are used in the more advanced grades of the elementary school. There the emphasis is not placed so strongly on testing for "native intelligence" as upon testing for the amount of increase in learning capacity. This problem is approached by using scales which are designed to measure "what the child has learned." Examples of such scales are the Otis Advanced Examination, Haggerty's Delta 2, and the National Intelligence Tests. Dickson and Norton tested 1043 eighth-grade pupils in twenty-nine schools in Oakland by the use of the Otis Test and reported that the individual scores ranged from 14 to 152 points, and that the average scores for the twenty-nine different schools ranged from 48 to 109. They state that the "mental ability of the best eighth grades was as good as that of the average ninth grade, and the mental ability of the lowest eighth grades equalled only that of the average sixth grade." By the use of the same scale, Colvin reports a range of 27 to 143 points for seventh-grade pupils, and a range of 47 to 171 points for eighth-grade pupils.

Similar results have been secured by the use of the other group tests cited. Especially have they found a place in the comparatively new school division, the junior high school. In 1921 the writer gave the Terman Group Test to 548 pupils in the 9-A classes in eight Los Angeles junior high schools. The range in individual scores extended from 27 to 198, and the average score for the entire group was 107. The average scores for the various schools ranged from 84 to 132. In each school the boys tested higher than the girls, but in two schools girls registered the highest individual scores. The average for boys was 115, with a range of 34 to 198, while the average for girls was 100, with a range of 27 to 187. The ages of these pupils ranged from twelve to nineteen years, but the thirteen-year-olds achieved the highest average score (118). The boy who scored 198 was but thirteen years of age.

The birthplace and previous schooling of each pupil were checked in an attempt to ascertain whether pupils who had received some of their training elsewhere would score higher than children who had received all of their training in Los Angeles. The average score for each of the two groups was found to be 95. This was followed by averaging the scores for each of the

ten test divisions and it was found that these 548 pupils were weakest in those tests covering arithmetic, word meaning, sentence meaning, and mixed sentences. By checking the school marks of these same pupils and the percentage of failures recorded against them, and by correlating each with the intelligence scores, it was possible to recommend a revision of the marking system in those school departments where the degree of correlation was low.

Such a study shows how these tests may be used not only to secure scores which will represent amount and kind of intelligence but to bring out individual differences as they are influenced by social conditions, sex, previous training, and school standing. Within the past two or three years many such studies have been made in American high schools. Madsen used Army Alpha in the high schools of Omaha to show the relationship between intelligence scores and success in high-school work. Proctor, at Stanford, used the same test in conjunction with the Binet Scale to show the necessity of making new adjustments in high-school courses to meet the different types of intelligence displayed by high-school pupils. Miller, at the University of Minnesota, after extensive application of a test which he

devised, concludes that the classification of high-school pupils on the basis of the results secured by mental tests gives certain advantages: (1) it makes possible an adaptation of the technique of instruction to the needs of the group; (2) it makes possible, but does not insure, an adaptation of the materials of instruction to the needs of the group; and (3) it may make competition operative as an incentive.

In colleges and universities the problem is much the same. Army Alpha, the Thorndike Intelligence Examination, and the Thurstone Psychological Examination have been given to thousands of college students for the purpose of predicting school success. Usually given to entering freshmen, an attempt has been made to show how the test scores are indicative of the amount of success the beginning student can be expected to attain in his academic work. Thus, at Ohio State, Michigan, and Brown universities the results are utilized in guiding and counseling students in the selection of courses and choice of their life work. This is the same use advocated by Proctor as a result of his experimentation with intelligence tests in high schools. In a few institutions a student is admitted to entrance upon a satisfactory showing on an intelligence

test, plus certain other ordinary requirements. In other institutions the testing has been confined to students on probation, or has been extended, for statistical purposes, to the entire student body.

Thus, it may be seen that intelligence tests, even though varying greatly in difficulty and in amount and kind of material presented, are used in much the same way throughout the various school divisions. Pressey has shown that individual examinations have been used for at least the three following purposes: (1) to ascertain what pupils are exceptionally dull or exceptionally bright; (2) to test those children to whom group tests cannot be given; and (3) to analyze the difficulties of the emotionally unstable. Group tests have been utilized to: (1) help in the classification of pupils in section or grade; (2) serve as one guide in educational and vocational guidance; and (3) show the differences in the problems of instruction encountered in the training of pupils of different group levels of intelligence. It is to be understood, of course, that both individual and group tests have been given in conjunction with standardized subject tests, and that such practice is being extended.

## REFERENCES

BERRY, C. S., "The Classification by Tests of Intelligence of Ten Thousand First-Grade Pupils"; *Journal of Educational Research*, October, 1922.

BOOK, W. F., *The Intelligence of High School Seniors*, chap. 14.

BROOKS, S. S., *Improving Schools by Standardized Tests*, chap. 9.

COLVIN, S. S., "Some Recent Results Obtained from the Otis Group Intelligence Scale"; *Journal of Educational Research*, January, 1921.

DICKSON, V. E., and NORTON, J. K., "The Otis Group Intelligence Scale Applied to the Elementary School Graduating Classes of Oakland, California"; *Journal of Educational Research*, February, 1921.

HINES, H. C., "A Program for Lowering the Percentage of Failures"; *School and Society*, May 14, 1921.

MADSEN, I. N., "Group Intelligence Tests as a Means of Prognosis in High School"; *Journal of Educational Research*, January, 1921.

MURCHISON, C., "Criminals and College Students" *School and Society*, July 3, 1920.

PRESSEY, S. L., and L. C., *Introduction to the Use of Standard Tests*, chap. 12.

PROCTOR, W. M., *Psychological Tests and the Guidance of High School Pupils.*

TERMAN, L. M., "A New Approach to the Study of Genius"; *Psychological Review*, July, 1922.

TERMAN, L. M., *The Intelligence of School Children*, ch. 13.

TUPPER, C. R., "The Use of Intelligence Tests in the Schools of a Small City"; in Terman's *Intelligence Tests and School Reorganization*, chap. 5.

WHIPPLE, G. M., and eleven others, "Intelligence Tests and Their Use"; *Twenty-First Yearbook.*

WHIPPLE, G. M., "The National Intelligence Tests"; *Journal of Educational Research,* June, 1021.

WILLARD, D. W., "Native and Acquired Mental Ability as Measured by the Terman Group Test of Mental Ability"; *School and Society,* December 30, 1922.

# IV

## WHAT IS TESTED AND WHAT IS NOT TESTED

A DOZEN years have passed since the writer pursued his first course in general psychology. The pursuit was furious, if not fast, for the terminology in which the subject was encased at that time was all but impenetrable. A truly scientific background was just beginning to be established and, although much had been done in the way of experimentation, clarity of expression was not the chief virtue. A large number of volumes had been written on the learning process, on thinking and reasoning, on memory and formal discipline, on the transfer of training, on emotion, and on will, and there was a dawning interest in the possible definition and measurement of "intelligence" which, whatever it might be, was agreed to exist.

Binet had completed his second revision of his own intelligence scale and Terman had just begun work on his American revision. If, at that time, there was any thought of testing groups of human beings for the purpose of ascertaining amounts, kinds, or degrees of intelligence, it was disguised under other names. The schools had

always provided a crude method of testing and segregating, of measuring "brightness" through achievement in school subjects, and a child was said to be "bright" or "intelligent" who was able to gather to him a large number of the much coveted "A's" and "B's." Since only those who received high ratings in school achievement were likely to go through to graduation, a satisfactory completion of the grammar grades, or the high school, or the college, came to indicate a certain level of brightness, about which nothing much was known but much assumed.

With the introduction of scientific methods into psychology and education, the manner of arriving at the ratings distributed to school pupils was examined and, by many authorities, found to be wanting. The method of measurement was not "standardized," i.e., no two examiners would give the same rating for the same piece of work. An English composition graded "A" by one teacher might, and probably would, be marked "C" by another. It followed, then, that a child's degree of brightness depended somewhat upon his good fortune in being able to submit his work to teachers who would give it a high rating. But the child who met the opposite condition, what could be done for him?

So far as a particular school subject was concerned, the remedy, it would appear, was comparatively easy to find. If the means of measuring work in that subject were not standardized, it was only necessary to provide a series of standardized tests which, when given and scored, would yield the same results from day to day and when given by different persons. The scheme was tried and met with enough success to encourage expansion and, as a result, to-day there are more than two hundred such tests for general circulation.

The standardized test in school subject-matter is designed to measure graduations of school training. It secures, with more or less accuracy, a measure of progress in the subject studied. It gives, with much greater accuracy than was achieved by earlier methods, an indication of the degree of "brightness" the child or the group has attained, i.e., "brightness" as it was measured empirically before the advent of intelligence tests.

The apparent success achieved in measuring the work of the schoolroom hastened the improvisation of tests of intelligence, particularly those designed to measure the intelligence of groups. Within recent years the two types of tests have been evolved conjointly. The situation arouses

in the minds of those who are concerned with these problems a curiosity which will not down. What is tested by tests of intelligence? In the scientific consideration of intelligence, how does it differ from the so-called "brightness" of a decade ago? What does the intelligence test discover that is not discovered by the subject test?

In a previous volume by the author,[1] Terman's reply to the first question is quoted. The various test divisions within a group, he says, test just what they test, just the ability to do *that*. He meant by this answer that the division confined to "problems in arithmetic" tests only the ability to work the problems given, that the division confined to "classification" tests only the ability to classify the words given, and that the division confined to "general information" tests only the ability to make the sentences true by marking the right words. He does not even infer, in this instance, that certain more or less hidden abilities are tested indirectly.

As has been pointed out (Chapter II), if all the group tests of intelligence which have been prepared for school use are considered together, the test divisions which appear most frequently are those covering general information, similarities,

[1] Hines, H. C., *A Guide to Educational Measurements*, p. 97.

carrying out directions, reasoning problems, and analogies. If to these are added tests covering logical selection and reorganization of mixed sentences, we have a composite test fairly typical of all group tests of intelligence. In attempting to answer the first question, we need go no further than has Terman. A test for general information may involve such traits as memory and imagination plus a breadth of experience, but these are of such fleeting and intangible character that if it were the purpose to measure them indirectly, they could not be fairly qualified with a numerical score. The same is true with the tests for similarities, directions, problems, analogies, selection, and reorganization. Indirectly, it may be, qualities inherent and acquired through previous training are measured but their values are not known. If, therefore, instead of trying to read into these tests characteristics which they do not possess, we simply state that the test divisions test what is evident, and nothing more, the atmosphere surrounding them will be greatly clarified.

It may be objected, however, that research workers have called these instruments tests of *general* intelligence and that, granting that they measure intelligence after a fashion, they still do

not include tests of enough abilities to be given the descriptive term "general." In considering this objection, it may be well to repeat our definition of intelligence. In a previous chapter (Chapter I) intelligence was defined as "the capacity of an individual to adapt himself to a new situation." It may be thought of as specific or general. A general capacity may make it possible to meet a specific situation wisely; a general capacity may be made up of specific capacities. A general intelligence, then, is the total of all specific abilities.

Now the composite test mentioned above includes the testing of but seven abilities. Abilities not tested by such a test are innumerable. Each is as much a part of general intelligence as are those which appear in the test. If they are not tested, *general* intelligence is not tested. Has the test, then, tested intelligence at all? It has, in two ways: (1) it has tested intelligence by testing integral parts of it; and (2) it has tested intelligence by presenting new situations and testing the capacity of the individual to adapt himself to them in such a way that the amount of adaptation possible and accomplished may be measured.

It would seem, then, that there is little, if any, difference between the trait psychologists have

chosen to call "intelligence" and the trait formerly called "brightness." If the two traits are not identical, it is because they were evidenced in different ways. In securing the earlier description the school child employed many media, such as personality, attitude, effort (often simulated), deportment, constancy in habits, the background of family, social environment, etc. In securing the later description the standardized test has been utilized. Because the standardized test has been given and scored each time in as nearly the same way as has been humanly possible, there has been little opportunity for the child to employ means other than actual performance under "controlled" conditions. The difference, then, seems to be that "brightness" is a term based on one or a series of personal judgments, while "intelligence" is a term founded on the results secured by "standardized" measuring instruments.

Carefully analyzed, these standardized measuring instruments, intelligence tests, have many characteristics in common with standardized subject tests. The computation problems in the National Tests and the reasoning problems in the Terman Group Test bear a close resemblance to those included in some of the better-known standardized arithmetic tests. The tests of gen-

eral information are usually composed of items taken from the subjects of history, household arts, manual training, nature study, literature, geography, music, arithmetic, agriculture, physiology, and physical culture, or the more abstract phases of these. Word-meaning tests and sentence-meaning tests are adapted from the subjects of reading and language. The ability to follow directions is involved in all standardized subject tests, while several subject tests involve the ability to make logical selections. Even the ability to reorganize mixed sentences is tested by an instrument for measuring ability in language.

The intelligence tests for groups as we now have them, and as we are likely more and more to have them, seem to represent a combination of a number of standardized subject tests. The latter, it was pointed out, are designed to measure graduations of school training or amount of progress in the various subjects studied. "Mental tests," says Colvin, "are possible when based on elements involving the common experiences of those tested." The same is true of subject tests, and it is possible to go just as far in inferring differences in native endowment when differences in the attainment of individuals are noted. The intelligence test, as it has been organized, certainly

does measure amount of school experience. This is witnessed on all sides, and was especially evident in the results of Army Alpha in which university and college students registered the highest scores. This does not argue, however, that separate subject tests may be substituted for group tests of intelligence.

What, then, is the chief difference between an intelligence test and a subject test? It seems to lie in the fact that the former seems to secure, as Ebbinghaus stated, a "combination activity." It tests, at one time, a series of abilities developed in and through the learning process, represented by the amount and kind of knowledge and fixed habits. It may, it is true, test a number of traits indirectly but these are not represented in the numerical score made by the individual. The score represents what the child can do with tests of arithmetic, reading, language, association, and general information. These measures can be secured one at a time through the use of subject tests but they are seldom, if ever, combined in the same way as the scores from intelligence test divisions. The total score made by an individual on a group test is the combined measure of a number of his specific abilities. It does not include all of the abilities it is possible for him to ex-

hibit; it does not even include all of the abilities developed in the classroom.

This discussion leads to an examination of those abilities, traits, or characteristics which have not been tested by intelligence tests. Thorndike has invited our attention to the fact that test scores do not indicate how well the subject "will respond in thinking about a machine that he tends, crops that he grows, merchandise that he buys and sells, and other concrete realities that he encounters in the laboratory, field, shop, and office . . . and how well he will succeed in thinking about people and their passions and in responding to these." In other words, intelligence tests do not measure amount or kind of aptitude. Stenquist was among the first to recognize this and his Mechanical Aptitude Test was devised to meet just such an emergency. Other tests of this character are the Wilkins Prognosis Test in Modern Languages and Thurstone's Vocational Guidance Tests, the latter series being designed to test for potential ability in engineering.

So far little has been accomplished in the measurement of adult intelligence and the types of intelligence developed without the help of classroom guidance. The farmer referred to in a previous chapter (Chapter I) is classified in both

divisions. Before judgment can be passed upon the amount and kind of his intelligence, he must be measured by an instrument which is not confined to the measurement of school training. The intelligence scores made by soldiers in the American Army were used in their classification by previous occupation, and certain conclusions were drawn which have since been found to be untenable. The scores did not represent, as was claimed, the types of relative intelligence exhibited by men in certain trades and occupations, but, to the contrary, gave only evidence of absolute intelligence as measured by a test of school training and its influence on mental ability. Murchison, in an excellent study of the intelligence of college students as compared with the intelligence of criminals, utilized Army Alpha in much the same way and, in both series of examinations, secured only measures of school training.

Terman has included in his Stanford revision of the Binet Scale a few tests of adult intelligence, but they have not been as extensive or as difficult as is desirable and he is engaged in extending them. He is also extending his scale in the other direction, in order that he may be able to secure a measure of the intelligence of children below the age of three years. By so doing, he feels that

"more reliable methods of diagnosis at this period would save a certain proportion of individuals from mental deficiency."

Since intelligence tests have either been confined to school examining or have measured only amounts of school training, little can be learned from the results about mental inheritance. It will be remembered that the Binet Scale is purported to "test native intelligence, not school knowledge or home training," but it can do so only by reflection and through inference. It is difficult to imagine a situation in which it would be possible to test for native ability in the absolute. Anthropometric measurements of the newborn babe can be made with a high degree of confidence. The babe will not change appreciably in stature while the measurements are being taken. But how much is his native mental ability altered by his environment? Who knows? While capacity to learn may be inborn, intelligence tests thus far have but measured what the individual has learned. That the Binet Scale either measures home training or that a child's native endowment may include a tendency either toward normality or deficiency may be inferred from Terman's proposal to devise a scale which will save some children from mental deficiency. Col-

vin states that "the brightest European child reared from birth by a group of African Pigmies would appear as a moron or worse if later transported to a highly civilized and cultured environment." If that is true, mental inheritance is far from having been measured. Many of our inferences, even, would have to be labeled as incorrect.

Looked upon in the light of ultimate success in measuring relative intelligence, it seems that the surface has but been scratched. Not much evidence has been produced to delineate racial differences, the effects of coaching, the effects of fatigue on performance, the influence of moral traits, the place of character qualities and emotion and will, the social adaptabilities of individuals, and those traits listed by Haggerty as "industry, perseverance, loyalty, and cheerfulness." These are all, to employ Freeman's terminology in a more extensive manner, "significant compoents of ability." They have not been tested by the instruments thus far produced, although beginnings have been made in that direction.

Even if we allow that the individual examination tests native intelligence only, it must be admitted that the group examinations test, in the absolute, nothing more than home training and

school training. The school child is given a definite task to perform in line with his previous experience. The degree to which he is able to perform accurately within the time allowed is expressed by the total score he registers. If it is understood what has been measured (a few or many of his abilities), the attempts to measure intelligence will be accepted in the proper light, without the necessity of resorting to the kinds of obscure verbosity referred to at the beginning of this chapter. Moreover, greater coöperation will be secured in overcoming the difficulties encountered in attempting to standardize the instruments.

REFERENCES

MONROE, W. S., DEVOSS, J. C., and KELLY, F. J., *Educational Tests and Measurements*, chap. 1.

STARCH, D., *Educational Measurements*, chap. 2.

TERMAN, L. M., and five others, *Intelligence Tests and School Reorganization*, p. 12.

THORNDIKE, E. L., and thirteen others, "Intelligence and Its Measurement"; *The Journal of Educational Psychology*, March, April, May, 1921.

# V

## THE DIFFICULTIES OF STANDARDIZATION

PERHAPS no phase of educational endeavor is more difficult to write about or more difficult to discuss in public utterance or in private conversation than the measurement of intelligence. Before we measure anything we must know what it is we are about to measure and be reasonably sure of the reasons for making such measurement. That has not been easy in dealing with intelligence because the term itself, representative of a combination of abilities possessed by human beings in varying degrees, is so difficult of standardization. In the possession of it no two human beings can be exactly alike and no one human being can have exactly the same amount at two different times.

Yet, the problem is with us; it has been started toward a solution; and it is doubtful whether so much public interest has been aroused since Darwin propounded his theories about the descent of man. Educational publicists, always keen to take advantage of a situation which presents a popular appeal, have not only maintained this in-

terest but have frequently added to the turmoil and confusion by seizing the opportunity to prepare new measuring instruments or to join the ever-growing vanguard of those who have all but solved this most intricate problem. For a time it seemed that the result would be that each of these individuals would solve the problem to his own advantage. Why not? Each one *has* intelligence: therefore, each one should know what it is.

It is no wonder, then, that we have had so many different types and kinds of intelligence tests. They range from one which tests an ability that does not exist to one which tests every ability found to be in any way commensurable. Naturally, in this rapid and extensive production, there would be many which are nothing short of ridiculous. There is a growing belief among educational psychologists that no less a period than three years should be occupied in the standardization of a group test of intelligence; yet, as Henmon has pointed out, "many of them are uncritically and hastily assembled, hurried into print without norms or standards . . . and often merely rearrangements of familiar material, selected on the basis of the author's opinion of the merits of the component tests."

In order to temper this headlong haste, the

Standardization Committee of the National Association of Directors of Educational Research has recommended that in the process of the standardization of a test certain steps be taken. Those steps, as they apply to the making of intelligence tests, may be listed as follows:

(1) *Preparation and selection of material.*

Before the material is prepared and selected, the problem of measurement must be exactly defined. It must be determined whether the test is to be limited to use in some one school division and, if so, what abilities are to be tested and what types of material have been found to be the best for testing those abilities.

(2) *Experimental organization of the test and the instructions for giving the test.*

If the test is to include the element of rate of work, the various items may be of equal difficulty, the rate at which pupils perform to be determined later. If it is to be a scale, an attempt should be made to arrange the items in order of difficulty, or to arrange the sections of the test, if there are several, in order of difficulty. The instructions under which the test is to be given must be clearly defined, but may be altered later if need arises.

(3) *Trial of tentative test to determine value of*

*elements, gross validity, reliability, and optimum conditions of giving, scoring, etc.*

Those items in the test which are correctly answered by every child and those which are answered by none ordinarily are eliminated, since they are not a fair test of what a child of a given age or grade can do. Validity is defined as the determination of what a test measures, and reliability as the consistency with which it measures at different times. Time limits may be approximated, as may the directions for giving and coring.

(4) *Final organization of the test.*

If it has been found that the test measures what it is intended to measure, it may be prepared for more extensive use. Usually two or more forms are prepared, equal in difficulty, so that the examiner may get a more reliable measure of intelligence than is possible with one form.

(5) *Final formulation of conditions under which the test is to be given, scored, tabulated, and interpreted.*

The directions must be precise and not too long; time limits, if any, must be definitely stated; scoring keys, in which nothing is left to personal judgment, must be supplied; and it is de-

sirable that some instructions be given as to how the test results may be interpreted.

It will readily be seen that the making of a test is no simple task. It is common practice, when the above steps have been taken, to submit the test and its accessories to experienced test makers or research workers for an official determination of its validity and reliability. Authors of tests frequently have endeavored to establish norms or standard scores by which the intelligence of pupils generally may be judged, but the tendency to use local norms, or averages, has become popular and there is less and less objection to this practice. A test becomes more reliable, or less reliable, as it is given to larger numbers of school pupils in various parts of the country. This accounts for the concentration of publication in the hands of a few test publishers who guarantee due publicity.

That standardization of a test may be facilitated some test makers and certain writers have emphasized the necessity of standardizing the procedure preliminary to giving. The examiner is warned to familiarize himself with the methods of examining and scoring, make certain that each test paper or booklet has been well printed and that the pages have been cut, be sure that each

child has been provided with a pencil, a stop watch has been provided, and that other conditions conducive to successful examining have been established. Further, the examiner should practice giving a test before he undertakes the examination of a group. Failure to do this sometimes leads to errors in inflection, enunciation, and proper emphasis, and may produce confusion in the minds of the children as to just what is wanted. His voice should be strong enough to carry to any part of the room, but should not be exercised in such loud tones that the children will be distracted rather than guided. Those children who hear poorly should be given seats near the examiner.

Since directions for giving and scoring are so explicit, there would seem to be little opportunity for securing other than a standard measure of the trait or ability examined. With a standardized instrument in the hands of an examiner whose methods of procedure have been standardized in advance, how could one but secure a result that would be exactly representative of the ability or abilities tested? Yet, it is known that there are factors, such as auditory and visual acuity of pupils, disturbances and interruptions, the attitude of the examiner, and certain inapparent physio-

logical states, that affect the performance of children and adults alike.

Thus it is evident that a standard score established from the results of a test which has been given in various parts of the country is of little value when considered from a scientific viewpoint. It is not a *standard* score at all, but simply an average which results from the statistical interpretation of the combined performance of many children in which the above factors are operating. Using a figure similar to the one employed earlier in this chapter, no two examiners can give the same test in exactly the same way, no two pupils can take the test under identical conditions, and no one pupil can register two total scores on the same test identical in what they represent.

Buckner has shown how extreme or erratic scores are often produced by group tests, but Pressey invites attention to factors which affect the interpretation of results. The latter states that some of these are: the age-grade situation, the time of year when the tests are given, the presence of a large foreign element, the exclusion of other sources of information, and the inaccuracy with which results are gathered and tabulated.

Perhaps the nearest it would be possible to come to the standardization of a test and its results would be for the test maker to administer his test to all groups examined by it, at the same hour on consecutive days, and to score all of the papers by a standard key without assistance. What factors might creep in then to defeat the purposes of standardization? The first doubtless would be the inability of the examiner to determine, without counsel, the relative values of test divisions, items, and instructions for giving. The second would be the dissimilarity in consecutive days in time relationships and atmospheric conditions. It would be impossible, of course, to keep groups of children mentally statical until all groups are examined, and it is to be admitted that meteorological conditions affect mental status. And the third, which is not so apparent, would be the inability of the examiner to score thousands of single papers without error; his own mental and physiological states are subject to change.

It may be objected that this is drawing too fine a line, that research workers have never hoped to reduce the operation of testing to such minute details. Nevertheless, there are those who have hoped for just such a refinement and, in a few instances, made it known that they are convinced

such a standardization has been arrived at. Taking the above argument seriously, there is nothing much left to do but despair, but it has not been presented with such a culmination in mind. These are simply points which must not be overlooked by those of us who subscribe to the use of intelligence tests in the solution of school problems.

To make the thesis clear, the writer may be permitted to draw from his own experiences. He was one of a large number of soldiers transferred from various army organizations during 1918 to a school for the training of psychological examiners. "Intelligence A," the first group test of intelligence, had been revised and shortened into *Army Alpha* and the officers in charge of the school, experienced psychologists, were engaged in training less experienced men in the art of giving intelligence examinations. Since this training extended over several months and since it was altered from time to time, the men who were graduated from the school and sent out to other camps at intervals to act as examiners did not receive what might have been designated as "standardized" training. This resulted in the army test being given in various ways to men from various walks in life and from various army

units and, it should be added, under very different conditions in different camps.

Having been selected to engage in a work the procedure of which had in no wise been standardized in the true sense of the term, the writer was called upon to act in all capacities from fire builder in the orderly room to chief examiner. In such activities he was able to observe some of the discrepancies which arose despite the effort of psychologists to keep them down. The examples are too many to enumerate here, but some of the things which made the results questionable may be listed. As indicated above, no two examiners administered the test in the same way. No two assistant examiners felt the same responsibility in helping to preserve order, supply new pencils, or cause the men to refrain from copying. The examination was seldom given without interruption or disturbance. Usually no attempt was made to locate men with impaired hearing or poor eyesight in order to give them an equal chance to make a good showing. Some men came to the examining rooms coached in what they would be expected to do; others came in fear that they would be given a "brain test" they might not be able to pass; still others came determined not to make a good showing in hope that they might

be discharged from the army. These, and many other factors, affected performance.

After the examination had been given, the papers were returned to headquarters, where they were scored by a detachment usually selected from a development battalion, men who were in no wise interested in the task before them and frequently working under the direction of a non-commissioned officer whose interest was not much greater. Training in scoring was not standardized; even the directions for scoring allowed for certain personal judgments. Naturally, there were many errors and much inaccuracy. Even the final reports sent in to the Washington headquarters were not the result of a standardized procedure, since methods of computing average scores and coefficients of correlation ranged from those invented by the local statisticians to the formulas evolved by Galton, Spearman, and Pearson.

No great number of these discrepancies were due to lack of vigilance on the part of psychologists. The speed with which the army was organized and moved caused many of them to arise. But the biggest discrepancy of all is the assumption that the test "measured general intelligence under controlled conditions." And

what may be said of the administration of the Army Intelligence Scale is applicable in a lesser degree to those tests which are mutations of it. The writer has seen many of the same factors creep in in public-school examining, and the enemies of intelligence testing have been quick to seize upon them as articles of confederation against the attempt to reorganize schools on the basis of their use.

Put into the hands of untrained examiners, intelligence tests are dangerous tools. With trained examiners, they fail to secure the standard results sometimes claimed. If Book had been able to administer personally all of the tests given to high-school seniors in Indiana, he would have come nearer to securing a measure of the abilities of a large group than any examiner who has worked in the field. The results he reports are among the most valuable yet accumulated, yet they are qualified, as he admits, by certain unavoidable factors. Until we have better ways of eliminating these factors, the recognition of their presence is always the better ground on which to take a stand.

If, as was stated at the beginning of this chapter, no one human being can have exactly the same amount of intelligence at two different

times, can the measurement of individual intelligence be so standardized that the intelligence quotient may be said to remain constant? That is, will the ratio of mental age to chronological age remain the same over a number of years? Terman has produced figures to show that it will. Sufficient evidence has not been produced to show that it will not. But in the attempt to prove or disprove, these unavoidable factors which have been mentioned, although present in a much lesser degree in individual testing, must not be overlooked.

Just as in the use of the word "intelligence" in testing for that which some writers have preferred to call "mental ability," the use of the term "standardized" is advisable only in so far as it defines a type of test resulting from scientific endeavor. It represents desire and not actual realization, but the steps that may be taken toward that realization lead on and on, as in the case of the straight line or the bell-shaped curve, to infinity. As has been stated, the work toward that end has but been begun. That which has been accomplished has its values. To ascertain what those values are, in their various applications, is a task which should not be shunned. It is the only way to search out scientific truth.

## REFERENCES

BUCKNER, C. A., *Educational Diagnosis of Individual Pupils.*

HENMON, V. A. C., "Intelligence and Its Measurement"; *Journal of Educational Psychology*, April, 1921, p. 197.

MONROE, W. S., *Introduction to the Theory of Educational Measurements*, chap. 2.

PRESSEY, S. L. and L. C., *Introduction to the Use of Standard Tests*, chap. 4.

TERMAN, L. M., "Mental Growth and the I.Q."; *Journal of Educational Psychology*, September, 1921.

# VI

## VALUES ACCRUING FROM THE MEASURING PROCESS

UNFORTUNATELY so much of the work that has been done in the measurement of intelligence has been of such impractical character that one is reminded of William James's blind man in a dark room searching for a black cat which is not there. If a certain American tendency to establish boards of censorship is carried over into educational psychology, the time may come when it will be at least theoretically possible to proscribe the highly unscientific efforts resulting from and attributable to unwise use of little learning and less experience. Such a possibility has been suggested by the Committee on Standardization of the National Association of Directors of Educational Research. Whether or not such action will ever be taken, certainly the time is not far distant when the measuring process will have to be carried on by those sufficiently trained in the work to understand its various ramifications.

Courtis, in his work in educational measure-

ments, has been forced constantly to combat the "forces of evil" which tear down what it has taken months of time and toil to build. In such work as he has done there is not so much danger from open criticism and antagonism as there is from having "the truth you've spoken twisted by knaves to make a trap for fools." This predicament is all the more noticeable in the measurement of intelligence. There, defeat is not met at the hands of those who are capable of making constructive criticism, but rather at the hands of those who, whether attempting to construct standard instruments or to use those previously devised, persist in misconstruing both theory and fact.

Adding these shortcomings to the difficulties of standardization, we may wonder why any attempt should be made to standardize tests and to measure intelligence. There must be arguments in favor of the development of the measuring process or the movement would have met the fate of other educational fads and fancies before it was well under way. Admitting that human curiosity may have served to extend the interest in testing for intelligence, a more influential factor is found in the desire to remould our school curricula and revitalize our school methods to

meet the needs arising from the discovery of individual differences. Important differences were known to exist before intelligence tests were invented, but the methods of estimating the "extent and frequency" of these differences were never satisfactory, since they possessed modifications and qualifications about which there could be no general agreement.

A measuring instrument which is made up of physical qualities or proportions, however limited, is of much greater value in securing a standard measure than a sum total of many different judgments. "Stepping off" one hundred yards is better than estimating the same distance; the human eye is subject to spatial or geometrical illusions. Thus, measuring amount and kind of intelligence by instruments which have been but partially standardized is better than estimating amount or kind by the exercise of personal judgments which can give us no standards by which to proceed. An important value accruing, then, from the measuring process is that which has resulted in the improvisation of partially standardized instruments and the attendant limitation of empirical judgment.

Contributing to the evidence that we shall be able some day to give a more adequate definition

to intelligence and that we shall be able to secure a measure which will be more easily acceptable, are the two correlative facts that it has already become possible to measure certain single abilities satisfactorily and that some of the recent group tests produced have combined the measurement of these abilities in such a way that the scores registered closely approximate those resulting from individual examinations. Perhaps no one factor served to make the Stanford-Binet a success more than that it differentiated mental abilities and has measured a large number of them. If the group tests are to possess such powers they must measure not only a large number of abilities but the largest number of abilities it is possible to include in tests which may be given in a reasonable time. What will constitute a "reasonable time" must be determined by the measurement of some of the factors mentioned in the foregoing chapter, such as physiological states, the will to do, fatigue, etc. Some of the more recent group tests do measure, with considerable resultant satisfaction, as many as ten abilities. The recognition that these abilities do not constitute the entirety of one's general intelligence, does not subtract from the value of what has been accomplished in their measurement.

Again, intelligence tests have made it possible to classify school children with more assurance and to bring about a higher degree of fairness to the individual child. There is no denying the importance and value of searching out those minds which are below normal and classifying them in such a way that they may be trained away from subnormality. No student of the measuring process would now contend that this cannot be done; it is being done daily, and to the benefit of all concerned. In the measurement of normal mentalities, in the discovery of individual differences, and in the detection of certain well-defined aptitudes, more light is being thrown on the nature of subject-matter previously offered and necessary to offer and on methods of instruction best fitted for average groups. What will be done in differentiating superior intelligence from normal intelligence depends to a great extent upon what it will be possible to do toward devising measuring instruments sufficiently difficult to test the higher levels. In that endeavor, however, much has already been accomplished, and it is now possible to point with reasonable certainty to those children who should be included in that classification.

The writer feels that if it should be agreed that

intelligence tests have had no lasting value save that they have invited the attention of teachers to individual differences in children, the tests will still have served a purpose far greater than any new method or process yet introduced into the life of the school child. The severest criticisms of public education largely have been the outgrowth of and have been directed toward the failure of teachers to discriminate between the causes which motivate for successful achievement and those which hinder and deter. Many children do not succeed in classroom work because the work is too difficult or because it is not difficult enough. It can hardly be said that the teacher has been at fault, unless, of course, she has failed to utilize all existing agencies to learn the problems of the individual child.

If the teacher could take the time, there is nothing discovered by tests of intelligence that she would not be able to learn about the children under her instruction. The same was true with army officers. Given six months or a year an officer in the American Army could, in acceptable terms, define and locate the different types of intelligence with which he has to deal; he would know which of his men had the amount and kind of intelligence needed to become leaders in army

service and which of them would either not be fitted to take part in responsible undertakings or might be trusted with only routine work. But, in the recent war, time could not be taken for such detailed study. Men were needed on the firing line, and in high positions of trust, and decisions were instantaneous. The army intelligence tests were intended to supply the officer with such information the day or the week the prospective soldiers were assigned to him. The school situation affords a striking parallel. In the modern system of education, a teacher who is endeavoring to instruct from twenty-five to fifty pupils needs a shorter method of ascertaining kinds and degrees of intelligence required to meet school situations than those she has been forced to employ. By the use of intelligence tests, she is now able to secure in a total of five hours spent in testing and scoring that which, without the tests, she was not able to secure in less than five months.

In some instances, intelligence tests have been used to discriminate between the general mental abilities of prospective teachers. If it is admitted that we have not yet been able to measure some of the factors that go to make up success, we should be very slow in making assumptions as to

the value of this procedure. So far, we do not have the instruments needed to determine which of these prospective teachers will learn to become instructors of average ability or above, although certain fundamental conclusions have been arrived at and agreed to concerning the cadet teacher who makes an exceptionally low score. The chief value in the practice, however, lies in the fact that prospective teachers are set to thinking about intelligence and its measurement.

Not only is the teacher caused to consider the differences in intelligence among the children she controls or will control, but all those in any way connected with or interested in learning and its outcomes have been set to thinking about intelligence. The instruments have attracted much attention and, as all new movements are open to ridicule, disinterested persons have found the measuring process an attractive subject for jest and cartoon. These attacks, contrary to that which might be expected, react in a favorable way by assisting the thought processes of those who might not have been reached in another manner. Once one has been started to thinking about intelligence, he is likely to become subject to a desire for a high degree of it. Thus, we may see values in that direction.

There are certain other indirect values. Among these is the value of having our educational system so closely inspected that we shall be compelled to erect new edifices of theory on foundations of concrete fact. From a personal introspection, educators have been forced to turn their thoughts of making education, as Suzzallo has worded it, "meet the social needs of a real world." This would have been possible, but not so probable, without the development of the measuring process. It bids fair to bring about the elimination of the non-essentials and the substitution of those essentials which have heretofore failed to find a place in the scheme of education.

Another indirect value, which at first may not seem so important, lies in the fact that the whole discussion has involved so many leaders in educational thought that we have been enabled to see more clearly which way to turn and whom to follow. That leader who seems to have the sanest views of this most modern of modern educational movements gains a sympathetic following. In that connection it is a matter of regret that certain educators have lost influence through their apparent inability to comprehend the magnitude of this movement and to be prepared to take a

stand. On the other hand, new leaders and new types of leadership have been developed who and which may carry us far along this newer road. In the readjustment, those who do not see the way clearly and those who are classified merely as obstructionists will be forced to stand aside.

A third indirect value accruing from the measuring process is the one which has been produced by the exposure of charlatans and pedants. The old popular methods practiced by phrenologists and "characterologists" are gradually losing favor, although the latter have a small following from among the illiterate, superstitious, and obdurate. There is more work to do, however, since psychological tests have fallen into the hands of those who prefer to make their living by advising others how and how not to conduct their lives. Much harm will be done by these, since the very magic of the word "psychology" will attract many of the less informed.

A fourth indirect value is seen in the marked tendency to think of all related educational problems in scientific terms. The four steps of scientific procedure employed by research workers have been followed in detail in the more serious attempts to measure intelligence. So far they

have been more applicable to individual examining where observation, hypothesis, deduction, and verification are followed in regular order. As these come to be more closely followed in group testing, through which larger numbers of professional workers are reached and influenced, other educational work, through example, will become more scientific.

In fact, intelligence testing has made the study of psychology more scientific in itself. Many unreliable and useless experiments of the laboratory, through the expansion of the latter to the public schools, have been cast aside, and the unintelligible terminology employed in describing them has practically disappeared. Psychologists are now dealing with concrete cases where once they all too frequently dealt with abstract theory. They are now better able to explain the fundamental rules on which they operate and are better prepared to withstand the attacks of the irreconcilables in education and religion. This is a value of much importance if great good is to be accomplished by and through intelligence tests. It is well that such opposition can be met with right thinking and clarity of expression.

Finally, the chief value, though still indirect, is the tendency to elevate the general level of

American intelligence. Thinking about intelligence, what it is, who has it, and in what ways it may be estimated or measured, cannot be harmful to the people. The case is somewhat analogous to the attempts to give disabled soldiers university training. Among these men are many who have not had the previous training which makes university work much easier of accomplishment. They do not learn easily and because of this fact they are frequently not held to the high standard required of other students with better advantages. Yet, they are encouraged to keep trying, for it is felt that they will leave the university with higher personal standards of living and will carry with them a social advantage which will make of them excellent citizens.

To get the people to thinking about intelligence and how it may be trained to higher levels is a task worthy of the best of publicity agents. Some of our research workers have been so busy, either carrying on their problems or devising newer instruments for the measurement of abilities yet unmeasured, that they seem to have overlooked a very important division of their labors, that of convincing the general public that the greatest asset of a nation is a high level of intelligence. Those persons interested in the

elimination of illiteracy cannot carry on this campaign alone; nor can those interested in moral education; they should have the assistance of all those who desire to see the nation reach the stage where each of its citizens is operating under an equitable expenditure of his individual mental ability. Such a campaign would necessitate some alterations in the statement of aims of American education. It seems now that those alterations are bound to come in any event.

Setting the values accruing from the measuring process against the disadvantage of being unable to test certain characteristics and qualities plus the difficulties of standardization, the second and third elements are outweighed by the first in a ratio of two to one. Although a group test of intelligence may be made up of test divisions which test only schoolroom abilities and which do not give us an absolute measure of native ability, and although certain traits and certain special aptitudes have not been measured and no completely satisfactory tests of adult intelligence have been devised, these in no wise compare in magnitude with the values of testing listed in this chapter. The same condition holds for the difficulties of standardization, many of which give promise of giving way as newer and better

methods are devised. Five years ago it would have been impossible to predict the enormous progress that would be made in testing in half a decade; nor is it now possible to predict with any certainty the great strides that are likely to be taken within the next five years. The very American tendency to rid ourselves of the mud and clay of theory and tradition and to get down to the hard pan of scientific fact, in which it is necessary to eliminate all those things which will be classed as non-essentials, will see the measurement movement through to a finish. What that finish will be will depend very much on the counsel and guidance of those to whom the making of our measuring instruments has been entrusted.

REFERENCES

BODE, B. H., *Fundamentals of Education*, chap. 7.
BAGLEY, W. C., *Educational Values*, chaps. 10, 11.
COLVIN, S. S., and BAGLEY, W. C., *Human Behavior*, chap. 18.
MCCALL, W. A., *How to Measure in Education*, chap. 5.
TERMAN, L. M., and five others, *Intelligence Tests and School Reorganization*, chap. 1.

# VII

## THE MEASUREMENT OF INTELLIGENCE AND THE AIMS OF EDUCATION

If intelligence is the capacity of an individual to adapt himself to a new situation and if we are trying to train school children gradually to become more intelligent, we shall have to take invoice of the aims of modern education in order to know whether they represent the criteria from which we can best proceed. Whether the degree of intelligence needed shall be evidenced in the "thoughtful conclusions" of Dewey or the "abstract thinking" of Terman, it will likely be judged on the basis of simpler manifestations. Doubtless the severest critics of public education to-day are those persons who have found cause to decry the apparent inability of teachers to train the school child to spell correctly, write legibly, speak articulately, and to perform the simple operations in the fundamentals of arithmetic. That these criticisms are not entirely without foundation has been recognized for a number of years and, although the indictment frequently has been too all-inclusive, the demand that something be done to improve the

situation, i.e., to teach children to be more intelligent in the exhibition of these abilities, has been so insistent that educators have cast about for a general yet utilitarian aim to which all teachers might be directed. Expressed in terms agreed upon, *the aim of modern education is to relate school training to life.*

There can be no quarrel with those who advocate this aim, for it can hardly be said to be new. It even dates back to primitive peoples, among whom the young were taught to care for themselves in the provision of food, clothing, and shelter, and in protection from attacking enemies. Their training was related to life in a very direct manner, and was passed on from generation to generation. Indeed, their training was life itself, and only those survived who were able to adapt themselves to new situations. With the advent of Oriental education, however, a higher level of intelligence was needed, for life had become more complex; labor had been divided and assigned, and education, somewhat more difficult to impart, was largely of a vocational character. As the centuries passed life became more complex still, and not only did schools for the training of the young appear, but books were introduced, and subject-matter which formerly had been trans-

mitted as a whole came to be divided into integral parts, in order that it might be made less susceptible to change "and ultimately to result in a body of knowledge more or less isolated from the world of active life and regarded by some as being more valuable on that account."

With the division of subject-matter, the chief aim of education in its simpler meaning disappeared, and other aims, either as mutations or expansions of it, arose to take its place. Life was no longer the comparatively simple problem of self-preservation, but the youth must be trained in religion, morality, and citizenship, that he might become an upright and useful citizen of the state. In all such training the state or church had jurisdiction or direct control, but, with the establishment of democratic forms of government, subject-matter became more varied and complex. In a modern democracy a certain fairly well-defined liberty is guaranteed the individual — liberty to help make the laws of the land, to share the rights of property, and, what perhaps is more important, to engage in whatsoever occupation he may choose. The right of choice, together with a further division of labor, has tended to confuse the issues involved in the methods and purposes of school training.

Not unlike theories in one respect, aims are frequently set up to explain why certain things have been, are being, and should be done. The general aim, as it has been expressed, is doubtless more properly used in an explanation of the thing that should be done. Worded in simple terminology and giving the effect of having covered the entire field of modern educational endeavor, it provides an easy answer to those critics who have questioned the ultimate purposes of educational training. Yet, for those who are more directly concerned with its application, it needs further analysis, and many attempts have been made to separate it into its parts and elements. This is very clearly done by McMurry when he divides the general aim into four factors "particularly worthy of acceptance as aims of school instruction." These four factors are as follows: (1) the teacher should inculcate objects in life or purpose on pupils; (2) the children should be taught to become constant students of the worth of things in order to prepare for their more immediate and future lives; (3) the children should be taught to organize or systematize their ideas; and (4) the pupil's ability to act as a leader, whether in his own affairs or in the affairs of others, should be developed.

These aims are all the explanation of things which should be done. They are, it will be seen, but an expansion, an interpretation, of the general aim. Yet they carry with them other meanings than those which are likely to be derived from a bald statement of the desire to relate education to life. Instead of confining the teaching of youth to the preservation of physical life, it must now include such factors as the acquisition of purpose, the weighing of values, the development and organization of ideas, and the cultivation of a spirit of leadership. The democratic purposes of education are defeated if the child is but taught to preserve his life by "making a living"; his conceptions of modern life outside the school, as Bagley has shown, will need to be broader than that.

It is quite likely that such aims are established with the conception of the average or typical school child in mind; but some difficulty is encountered in their application when individual differences are taken into account. A fact long suspected has been, in the main, established through the use of intelligence tests — there is a type of mentality wholly incapable of acquiring purposes, weighing values, organizing ideas, and cultivating a spirit of leadership by and through the ordinary

methods of the classroom. Such a type of mind presents a problem far removed from that presented by the normal or super-normal mind and a different set of aims will be necessary. The former common practice of attempting to make the same set of principles apply to each and every child is no longer acceptable; all children do not have the same potentiality, and the belief that *any* child can be trained to become a genius must be discarded.

When differences in endowment are considered, the fourth factor listed by McMurry is but the expression of an ideal. Much, if not too much, has been said about education for leadership. The expression should not be used without qualification. Leadership is as hard to define as intelligence, yet he who runs has a conception of it which is more or less common to us all. The reader may use his own conception, yet, if he can in any way rely upon the interpretation of the normal curve of distribution, not more than ten to fifteen percentage of any large unselected group will ever attain positions of leadership. What was found to be true through measurements in the physical sciences has been found to be true in the distribution of human abilities. Intelligence tests have proved this time and

again, and there is no reason to assume that the traits we have not yet been able to measure would be distributed in another manner.

It is possible to cultivate a spirit of leadership among school children in the sense that those who successfully complete a period of school training will, ordinarily, lead those who fail to carry on that far, but is a child given democratic freedom when he is trained to a leadership he is incapable of accepting? Studies of individual differences have brought out that a proportion (however much we do not know) of the ability to lead is inborn. To discover this innate tendency in the individual child and to apply the aim of leadership where it will best fit requires keen insight and delicacy on the part of the teacher. Although recent research has made clear that it is not altogether impossible, except in extreme cases of subnormality where mental ability is not sufficiently developed to meet the average problems of life, as yet there can be no surety which type of mind will succeed and which type of mind will fail. Until it becomes evident just what specific aid and direction should be given, it is important that each school child be impressed with the idea that he must go as far toward the top of the ladder as his capabilities will carry him. Having a positive

aspect, this will, at the same time, make more moderate and less sweeping the assumption that all children can and should be trained to become leaders.

"The child," says Snedden, "is educable along four lines sufficiently distinct to guide our choice of ways and means: (a) physical; (b) vocational; (c) social (or moral, religious, and civic); and (d) cultural (in a somewhat narrow sense of the word, embracing individual development along lines of pure æsthetic and intellectual interests for the ends of personal refinement and satisfaction)." The child who is physically able to attend the public school is capable of being taught the rudimentary habits of hygiene and the proper habits of play. In so far as the teacher is able to instruct him in these habits, just so far is she able to subscribe to the general aim of preparing him for life outside the school. Physical strength without proper development is practically useless from the standpoint of intelligent behavior, and it is not likely that too much care will be exercised in the child's physical preparation for the strenuous activities into which he will probably go.

The child who is physically and mentally capable of carrying on the work of the primary grades can be taught the simpler projects pro-

posed in modern school practices, but his vocational education will depend upon his capacity for growth which may or may not exhibit marked limitations before the period of adolescence. As individual differences have come to be more and more emphasized, the field of vocational guidance, and particularly that phase of vocational guidance known as pre-vocational training, has become more complicated, and it has been next to impossible to formulate definite standards of procedure for the vocational training of youth. Terman, on the basis of what he has learned through the use of intelligence tests, recommends that vocational training and vocational guidance should begin as early as the fifth grade of the elementary school, but that the latter should always be preceded by educational guidance. He states further that "mentality of eleven years is ample for ordinary kinds of unskilled labor, and many of the semi-skilled trades are within the reach of those who test a year or two higher." These deductions are made as the result of certain studies made in connection with the use of the Stanford-Binet and, while it may be admitted that the individual examination has certain predictive values, it must be remembered that group tests of intelligence have not been so refined that we can

predict with certainty the type of occupation to which the child may aspire with profit.

Social and cultural training are not so tangible as physical and vocational training. The child's physical condition is under daily observation, and his vocational proclivities from day to day may be discerned with a reasonable amount of accuracy. But what social inheritance will be of greatest benefit to him, and what attempts should be made to bring about his personal refinement, have not been generally agreed upon. Precept and example are still to be considered as important factors, but it is not without the bounds of prophecy to surmise that we shall have scales and tests for the measurement of social and character traits which will cause us to modify our aims once more.

In its outline of the aims of education as they may be applied to the secondary school, the Commission on the Reorganization of Secondary Education of the National Education Association includes as the second aim the "command of the fundamental processes," which are termed "tools of intelligence and culture." "Much of the energy of the elementary school," states the report, "is properly devoted to teaching certain fundamental processes, such as reading, writing, arith-

metical computations, and the elements of oral and written expression. The facility that a child of twelve or fourteen may acquire in the use of these tools is not sufficient for the needs of modern life. This is particularly true of the mother tongue." This report supports the writer's contention that so far as the school pupil is concerned classroom abilities are but subsidiary parts of general intelligence and that, by the use of intelligence tests, we have but measured in a scientific way the results of training, recognizing at the same time that native ability is present in all exhibitions of intelligence. Further, it shows that educators admit to serious consideration the criticisms offered by those outside the public school.

Considered collectively, the aims of education make evident the desire of educators to determine both the practicable and desirable in life. The aims of intelligence testing are directed toward the same ends. If either set were carried to fruition, a complete reorganization of instruction would be necessary and a course of study for each child would be the logical result. However, the variables encountered in the practical conduct of life make the desirable impossible of standardization. Further, the steady increase in enrollment resulting from the demand that education be sup-

plied to all the children of all the people makes such reorganization impractical pedagogically and administratively.

The teacher may find herself somewhere between these two sets of aims until they have been resolved into one. She will be told that individual differences must be sought out, but that individual instruction is not possible in a class of normal size. She will be told that she should train her pupils to become leaders, but that there are only a few who are capable of leadership. She will be told that all school children should be trained to become more intelligent gradually, but that there are those in whom acquisition of intelligence is impossible. Caught between theory and fact, she will be wise indeed if she is able to reconcile the two and resolve them into a program of action. But she it is to whom the "tools of intelligence and culture" have been entrusted; the use she makes of them will be reflected in the general level of mental ability of the American people.

With the introduction of scientific methods into education, the modifications which society works out are mirrored in the subject-matter and methods of the classroom. The demand that children acquire facility in the use of classroom

tools has made it imperative that certain tentative standards be set up to which instruction may be directed. There is now rather general agreement on the list of words the child will need to know how to spell, on the types of arithmetic problems he will need to know how to solve, on the quality of handwriting required of the typical citizen, and on the extent of vocabulary necessary to articulate intelligible speech. All this indicates progress; it is a definite and comprehensible approach to the general aim of education.

If the teacher is conscious of changing modes of living and changing conceptions of education, she, in a great measure, will be able to fulfill the desire to correct and extend the knowledge, habits, feelings, and attitudes of the children she is called upon to instruct. She will need to know, however, how children learn, how such learning is being, and can be, measured, and how it can best be developed. Her acceptance or rejection of some of the newer practices in education will depend very largely upon her knowledge of these things. Certainly her mind should not be closed to the newer practices, nor is there reason why she should accept them unwittingly and unprepared. If the facts secured through the measurement of intelligence are strong enough to stand

alone, they are likely to demand a part in the advancement of education. In that event the aims may be redirected and, since she must follow them if she is to be successful, she should have a voice in their restatement. It is she, after all, who will help most to determine values in the measuring process and, through a discriminating attitude, she will be able to recognize those things which are democratic and those which are undemocratic in modern education.

REFERENCES

BAGLEY, W. C., *Classroom Management*, Introduction. *The Educative Process*, chap. 3.

BLACKHURST, J. H., "The Normal Curve as Related to High School and College Grading"; *School and Society*, April 9, 1921.

CUBBERLEY, E. P., *Changing Conceptions of Education.*

EARHART, L. B., *Types of Teaching*, chap. 1.

GRAVES, F. P., *A Student's History of Education*, chap. 1.

HANUS, P. H., *Educational Aims and Educational Values*, chap. 4.

HENDERSON, E. N., *Principles of Education*, chap. 17.

KOOS, L. V., *The Junior High School*, chap. 2.

MCMURRY, F. M., *Elementary School Standards*, chap. 2.

O'SHEA, M. V., *Education as Adjustment*, chap. 6, 7.

SNEDDEN, DAVID, and DUTTON, S. T., *Administration of Public Education in the United States*, p. 326.

# VIII

## THE MEASUREMENT OF INTELLIGENCE AND DEMOCRACY IN EDUCATION

PART of our educational world seems to be fairly equally divided between those who believe that "scientific study and control of the child would not be right" and those who would put our complete educational program on a scientific basis. The latter class is again divided between the biologists and psychologists and their respective followers. The failure of these two schools of scientists to agree in such questions as the relative strength of hereditary and environmental forces has prevented them from securing additional support from those who would otherwise have flocked to their banners. But theirs has been a tremendous problem and they have been forced to carry on their investigations not only in the face of obstacles which arise in experimentation but against the barricades thrown up by those who are satisfied with existing conditions.

They who would measure intelligence are far from being satisfied with present conditions. "The world," says Steinmetz, "belongs to the

dissatisfied." If there were no obstacles the world would doubtless belong to the psychologists. But there are obstacles, and they not only are supplied by the forces mentioned above but by those who are affected directly or indirectly by the results of the tests administered. These latter are the teachers, parents, and laymen, who, through failure to comprehend the aims and purposes of mental measurement as they are set up by the saner thinkers in the field, have misconstrued the truths they have tried to utter and described the entire movement as a "tale told by an idiot, full of sound and fury."

No defense need be made of the exaggerated claims of those who, through novelty of experience and background of little training, have read into the results of intelligence tests not what is really there but rather what they have hoped would be there. The tendency of these persons to classify freely each exhibition of intelligence, regardless of the factors that have fashioned it, is not only defenseless but damaging to the more wholesome and justifiable conclusions. The writer has shown how much more difficult it is to measure mental ability than to measure physical property or dimension, but knowledge of this condition does not seem to have gained access to the

minds of some of the erstwhile enthusiasts who, though frequently long on verbiage, are a little short of perfection in abstract thinking.

Because of this tendency toward false interpretation, those who are none too warm in their praise of that which has been accomplished by research workers occasionally have felt it incumbent upon them to disseminate more or less vitriolic attacks upon the movement in general. Each time this is done by a leader in educational thought the measurement movement is set back several years. And this is not so much because there is no defense for the more worth while developments in the field, but rather because those who have been unalterably opposed are grounded in their unbelief deeper still and those who were almost convinced are caused to slide back into misgiving and doubt.

Although our American educational institutions almost from the very beginning have been founded upon democratic principles, not until the past ten years have we heard quite so much about democracy in education. It is a happy phrase and we, in the long evenings before the fireplace, are comfortable in the thought that our schools as well as our other social institutions make it possible for the individual, no matter from what lowly

origin, to secure an equal training with him who comes from a background of better advantages. In the midst of this thought it is disturbing, if not distracting, to be told that not *all* the children of all the people are capable of acquiring this much sought for education, that we have set up an ideal which is not only impossible of attainment but which will tend to undermine the foundations of our political and social institutions.

A few years ago, before intelligence tests had come to be such a factor in the determination of educational policies, there were other arguments proposed for the refutation of those offered in defense of democracy in education. It would not do, it was reasoned, to provide an education for each and every child brought into the world, for there are certain tasks which must be performed by the "uneducated." If every individual is to be trained so that he may accept and keep a "white collar" position, who will do the manual labor of the world? Certain evidence was produced to show that it would not be done by "educated" persons, and the statistics gathered, such as those relating to the preferences of high-school graduates, were, in the main, convincing.

The educational "democrats," however, met these arguments by demonstrating that we have

been and are gradually drifting away from the necessity of having manual labor performed by human beings, that science has always provided newer types of mechanical devices to be employed as substitutes for human labor, and that, when the substitutions are made, there has been no lack of employment of an even higher character. They contended that, through the education of every child, the race would not only be lifted to a much higher level of intelligence but that education itself would teach men to provide the ways and means of making the necessary substitutions. The resultant elevation of the race, it was argued, would perpetuate its life.

One may go on almost indefinitely arguing on one side of this question or the other and get no nearer to a decision as to whom an education should be offered than when he first began. Yet the preponderance of evidence has been on the side of those who subscribe to democratic practices. Because democracy guarantees a certain well-defined liberty, particularly the liberty of engaging in whatsoever occupation one may choose, it is easier to say and believe that the only way this ideal may ever be realized is to throw open the doors of the public schools to all comers. And not only shall the doors be thrown open but,

when the edifice is filled, no child shall be allowed to escape until he shall have reached an age beyond the limits of compulsory training. For we have so pledged ourselves to this ideal of education for the common good that we have placed laws upon the statute books making education mandatory, and we are now in the midst of a discussion concerning how far upward we shall place the upper limit.

There is much more to be said about this phase of democracy in education but it need not be said here, for there are other phases which demand attention. If we can judge by the surprise exhibited, the psychologists threw a bomb into the assemblage of educational "democrats" when they announced that by scientific procedure they could produce facts to show that all the children of all the people did not have the native ability to accept and carry this open-door education about to be thrust upon them. It was a startling statement. Why it should have been so is not clear, since teachers had known since schools began that there were those who met difficulty in learning even under the most skillful teaching. Perhaps they had not realized the import of that which they knew to be true, since they had frequently been told that all children could be

taught to learn if proper methods were employed. In fact, our educational system has passed through an evolution of educational method. More than that, it has made one complete revolution, for we are back to the problem of the individualization of instruction, where we were when schools were first established.

The apparent fallacy in the statement of the psychologists, however, is one which so frequently accompanies the announcement of a new discovery. Bagley has pointed it out by simply stating that "the facts were not all in." The implication has a familiarity about it that reminds one of other scientific discoveries, — those having to do with the poles of the earth and the relativity of the planets. But in the criticism of their findings, or the statements resulting from their findings, there is no desire to limit the operations of science or to prohibit an announcement of progress in experimentation. The misunderstandings which have grown out of the earlier delineation of the measuring process doubtless were due to unfortunate terminology. Intelligence, or general mental ability, had not been measured in its entirety, and the announcement that measuring instruments had been devised which would secure a complete measure

of *general* intelligence was, in the least, misleading.

That seems to have been the first important error. There have been many others, chief of which is the fact that psychologists, in their enthusiasm over the possibility of freeing the schools from the "educational lockstep," in some instances recommended and, in others, permitted persons unfamiliar with the instruments and their possibilities to employ them indiscriminately in testing for a function which some of them have thought to be a structure. This action is unfortunate and it has had a consequent reaction — untrained teachers have been told so frequently that little training is needed in preparation for the giving of group tests that those of us who see dangers in such widespread practice occasionally have been forced to admit to the untrained that they may proceed, knowing, at the same time, that they will proceed anyhow, and doing what little we can to direct them from a distance. The action has no analogy in the development of any other science. In the correlative field of medicine, for instance, it is difficult to imagine physicians and surgeons recommending to those who have not had sufficient training to recognize an ailment when it appears the indiscriminate use of instruments for diag-

nosing amount or kind of physical ailment. Not only is there great doubt as to whether it is wise to permit teachers untrained in psychological measurement to administer intelligence tests but there is great question as to the amount of training necessary to make the proper interpretations of results. One but needs to turn to the pages of some of our educational periodicals to realize what mistakes may be made in the name of psychological science, and these are sometimes the errors of persons who profess to leadership in these newer investigations.

It is little wonder, then, that those who adhere so strictly to the theories on which democratic practices in education are founded have been astounded. They met the earlier announcement with not a little nervousness, which was hastily changed to a feeling of dismay when later reports came in. So many of the "democrats" had felt themselves to be in the possession of at least average intelligence, yet they were confident that had intelligence tests been employed when they were passing through the elementary schools they would have been classed as imbeciles, or at best, high-class morons. They have passed from dismay to jesting tolerance, and finally to boundless anathematization.

The result has been that the argument has changed from a discussion over the necessity of "carrying on the world's work in the usual way" to a diatribe between the educational "democrats" and the psychological "aristocrats." One group stands for the generous training of all the children, the other for the generous training of a chosen few; one is for a loose construction of the educational constitution, the other for a strict construction. That seems to be the main difference and practically the only real difference between them. Both groups want the children to be educated, the former principally by methods now in use and the latter by methods so revised that they will care for the individual differences discovered by tests of intelligence. The first group agrees with the second that there are values in intelligence testing that must not be overlooked, but it refuses to stand quietly by and accept some of the disconcerting assertions made by the camp followers in the second group.

In order that we may understand this argument, it is only fair that both sides be presented. To do that we shall have to refer to a time before the argument took on serious proportions. In March, 1921, Haggerty, one of the leaders in the development and application of group intelli-

gence tests, in an address before the National Directors of Educational Research, said the following: "No claim is here made that our present tests are finally satisfactory. Sufficient evidence may be adduced to show that they are not, and that in behalf of scientific accuracy we must go on improving these tests, reducing the sources of error and determining the specific uses to which particular tests are best adapted. The direction which experimental work should take is fairly clear; certain of the important methods are known and important improvements in tests may be predicted." And a little further on he adds: "In noting the progressive use of intelligence examinations during the past year it should be pointed out, therefore, that there has also been a growing recognition of the limitations of intelligence examinations and even of intelligence itself as a basis for prognosis. At the same time there have been efforts to supplement the measurement of intelligence with estimates and measures of the non-intelligence factors contributing to success. There is first the effort to find tests for special aptitudes as a basis for predicting success in particular occupational fields."

If these remarks of Haggerty, coming at a time when we were passing from the extravagant

claims of some of the army psychologists to the extravagant claims of superficial workers in the educational field, had somehow found a home with all of his auditors and had been transmitted by them to all of their followers, there would have been far less need for the virulent attacks which have followed. It was a very modest claim he was making for intelligence tests; it is the type of claim which should have been made when intelligence tests were first devised. It had resulted from four years of careful deliberation upon and study of test problems. Terman has conceded that such deliberation and study is necessary in the announcement that a detailed treatment of the data accumulated in his investigation of superior children may occupy several years and the follow-up work an additional decade or two.

But all research workers are not so conservative, and there has been considerable "branding" of certain children as "inferior" or "superior" on the basis of the first returns of intelligence tests. Townsend, in February, 1922, registered objection to this practice in the following words:

> A generation or two ago under the influence of "instinct psychology" we should have said, "Something is wrong with the school." Now, under the influence of "mental test psychology" we say

"Something is wrong with the child." An unsatisfactory result of the test brands the person as inferior. An adjective which applies only to the result is applied to the person. An unsatisfactory result is translated without the slightest hesitation to read "an inferior child." This exposes what I have referred to above as the practical fallacy of our method. Failure is accounted for in terms of fixed degrees of ability and our minds are insulated from the concrete situation.

That this tendency to "brand" a child on the basis of an unsatisfactory score is not attributable to methods alone but may be accounted for in the types of instruments used, is inferred from the words of Thorndike who, in the same month in which Townsend's article appeared, took exception to certain nefarious practices in the following words:

In the elementary schools we now have many inadequate and even fantastic procedures parading behind the banner of educational science. Alleged measurements are reported and used which measure the fact in question about as well as the noise of the thunder measures the voltage of the lightning. To nobody are such more detestable than to the scientific worker with educational measurements.

It was Bagley, however, who, at this same time, fired the shot that was heard, if not 'round

the world, at least from the British Isles to the Hawaiian Islands. In an address before the Society of College Teachers of Education, February 27, 1922, Bagley attempted to show "that the sanction which mental measurements apparently give to educational determinism is based, not upon the facts that measurements reveal, but upon the hypotheses and assumptions that the development of the measures has involved; that these hypotheses and assumptions, while doubtless justified for certain purposes, are at basis questionable in the last degree; and that the present tendency to extend them *ad libitum* beyond a very restricted field is fraught with educational and social dangers of so serious and far-reaching a character as to cause the gravest concern" and that "even if the assumptions are granted, many of the fatalistic inferences drawn from the data in hand are not justified."

After quoting an account of a lecture delivered by a British educator, in which it was stated that seventy percentage of the children of England would never develop any more intelligence than that which should be possessed at the age of fourteen and consequently further education would be wasted on them, and from an editorial in an American educational periodical, in which intelli-

gence is classed as a possession as inherent as blue eyes or Roman noses, he proceeds to show that mentality is "not an anatomical structure, but a function," and that in using the I.Q. as a basis for predicting possibilities of vertical mental growth the possibilities for horizontal mental growth have been overlooked. "With the constancy of persistence of the I.Q. still in doubt," he states, "the edict has gone forth that, 'for all practical purposes,' it is safe to predict a child's future at the age of twelve. It is 'safe,' in other words, to stamp the twelve-year-old child with the brand of permanent inferiority. It is 'safe' to neglect the broader education of mediocre and dull children, to let them be satisfied with narrow specific training that will fit them only for routine work, and to reserve the higher privileges for the 'gifted' children. With his instruments of selection admittedly faulty, with his measures that measure something that no one yet has been able to define, the determinist proposes this policy and seeks to justify his proposal on the high grounds of social welfare and especially of social progress."

Concerning democracy in education Bagley has the following to say: "If the determinist is right, the ideal of democracy is wrong; the forces that have resulted in a democratic social order are

forces of social involution and not of social evolution; the educational aims and ideals that have gone hand in hand with this social involution and which have assumed the possibility of insuring a certain kind of equality among humankind, are a tragic, even a ghastly, mistake." Later on, he says:

> There is, however, a factor connected with this matter of "leadership" that merits the most serious attention. The qualities that make for democratic leadership, far from being exclusively intellectual qualities, are not even predominantly so. They are rather "human" qualities, such as sympathy, tact, humor, and sociability, and "moral" qualities, such as integrity, industry, persistence, courage, and loyalty. Men and women of average or below-average mentality may possess these qualities in such abundance that they become leaders inevitably. It is of the greatest importance that these men and women be prepared through education for the responsibilities that will devolve upon them.

Before some of the same auditors to which Bagley's address was delivered Holmes read a paper entitled, *The General Philosophy of Grading and Promotion in Relation to Intelligence Testing.* In it he presented six arguments which he himself summarized briefly as follows: (1) It may be granted that tests of intelligence distinguish

with reasonable accuracy between children with superior intellectual ability and their duller companions; (2) It may also be granted that segregation of gifted children is to their advantage by reason of the fact that they can be taught more effectively in a homogeneous group; (3) It may also be granted that such segregation does not necessarily result in "pushing" the bright children beyond their permanent powers of accomplishment; (4) The argument against such segregation on the ground of "democracy" may be dismissed; (5) But there is nothing in all this which proves the advisability of advancing gifted children more rapidly through the grades; and (6) rapid advancement may be justifiable for individuals, but it is questionable as an administrative policy. In explanation of the fourth argument Holmes adds:

> Democracy will be maintained on the playground and in the social activities of the school. The brighter children need not be told that they are gifted nor the others that they are not gifted. To teach bright children together is not to separate them from their companions who are of lower intellectual endowment.

Approaching the problem from its social rather than its intellectual implications, Counts, in April, 1922, wrote that

we must literally comb the elementary school for every scrap of superior talent to be found within its population, and then, utterly regardless of such extraneous considerations as accidents of birth, see to it that this talent receives the finest training that the richest society in human history can provide. Along with this special training must go the inculcation of powerful social ideals and the development of feelings of social obligation. Deficiency here must be regarded as sufficient cause for the cancelling of superior educational opportunities.

Returning to the arguments presented by Bagley — he was answered in a brief manner by Whipple in June, 1922, when he wrote:

We are perfectly willing to admit that we are trying to measure intelligence. We admit that we are not all agreed as to the precise nature of this capacity, but that is no reason why we may not watch its manifestations. No psychologist that I know of pretends that we are measuring intelligence directly. The physicist measures the phenomena of light and electricity without knowing their exact nature. We cheerfully admit that we measure merely certain aspects of behavior, and thence infer the existence of a certain degree of something that can be roughly designated as "general intelligence." We cheerfully admit that many other factors besides that of general intelligence do influence the progress of pupils in our schools. We do try to recognize and allow for the

operation of these supplementary factors. We contend, however, that our intelligence tests do measure, with a precision that is surprisingly satisfactory, a factor which is of the utmost significance for educational progress. . . . To say that the psychologists interested in intelligence testing want to or ever will forever condemn a pupil to a shortened educational career on the basis of a low I.Q. is a piece of rhetorical exaggeration.

In setting forth his creed, Whipple adds:

I believe that the real meaning of democracy *is* safeguarded in the notion of "equity of opportunity," and if any nation is destined to perish it is that one which fails to provide the best possible educational training for those of its rising generation that show promise of intellectual leadership.

Some of Terman's statements had been attacked in Bagley's address, and in June, 1922, Terman answered Bagley in the *Journal of Educational Research.* Among other things he said:

Dr. Bagley admits that vertical growth has its (predetermined?) limits, but thinks he has won his argument by showing that the possibilities of horizontal growth are limitless. In making this admission Dr. Bagley has given away his whole case. Presumably he is not aware of the fact that this very distinction was drawn by Binet fifteen years ago and that it has been especially insisted upon by all psy-

chologists in the field of individual differences. In reality it is one of the main arguments in favor of a differentiated curriculum. The differentiated curriculum allows all the vertical growth of which a particular child is capable and follows this with specialized training along horizontal lines. But hardly has the author admitted the limitations of vertical growth when, as if realizing that he had admitted too much, he retracts and argues that the common man in his daily business constantly grows in power to deal with abstract problems.... From Galton on down to Thorndike and Davenport the scientist has produced a considerable amount of evidence in support of the hypothesis that one's mental traits, especially his intellectual abilities, are pretty largely determined by native endowment. The hypothesis may or may not be modified when all the facts are in, but by the canons of scientific method it is entitled to stand until concrete evidence has been adduced to show that it is false.... The psychologist does not propose "to exclude everyone except the high I.Q.'s" from the privileges of secondary education, though he pleads guilty of demanding new types of secondary education which would be better suited to inferior intellects than is the typical college preparatory curriculum... the psychologist of individual differences believes that the one purpose of intelligence tests in the schools is to aid us in making the most of every child, the dull as well as the bright.... If the psychologist can escape the charge of being an enemy to democracy only by denying his facts, then he must

admit that he is an enemy of democracy. The writer, for one, would insist upon a less naïve definition of the term democracy. Certainly any definition to be acceptable will have to square with the demonstrable facts of biological and psychological science.

Bagley's rejoinder to Whipple appeared in August, 1922. In it he showed how that which he had really intended to say had been misinterpreted. He says:

I did not characterize the tests as "merely a set of highly artificial symbols, thrown together on hypotheses which are themselves debatable." I said that what the tests *directly* measure is not native ability but acquired ability; that the interpretation of the results of the tests in terms of native ability involves the assumption that educational and other environmental influences have operated with equal force upon all individuals compared; and that this assumption is always questionable except under rare conditions. . . . I did not say that the "differentiation of the curriculum and of the methods of instruction" for pupils with different degrees of learning capacity is "contrary to the best interests" either of "many individuals" or of "our national life." Refinement of the methods of instruction is, indeed, the greatest need in the education not only of the normal and subnormal, but also of the supernormal; and while, as I have pointed out repeatedly, the provision of a *per-*

*vasive common culture* is the prime function of democratic education, this does not necessarily mean an absolutely uniform curriculum, nor does it preclude an abundance of differentiation in what may be termed "non-essentials."

However, Bagley seems not to have been the only one whose statements had been misinterpreted. Dr. Rusk, the British educator referred to in an earlier paragraph, in October, 1922, called attention to the fact that he had been misquoted by the paper from which Bagley had quoted. He says:

Professor Bagley's attack on those whom for his own purposes he designates "determinists," is but one more chapter in the old educational quarrel between initial equality and initial inequality, between the views that the influence of education is limited and that it is unlimited. . . . His fear that the developments of mental testing threaten the entire theory and practice of democratic education argues a misunderstanding of mental testing or the ideals of democratic education or both. It will indeed be a sad day for democracy when democratic ideals are incompatible with scientific progress. It is not the democratic but the conservative like Plato with his caste system of education that lives in constant dread of innovations in education. . . . That there are differences in mental abilities and that these set limits to the influence of education, have always been recognized by

practical teachers. . . . All that the mental testers have done is to give precision to these differences by expressing them in percentages. By the use of mental tests, in addition to the usual scholastic tests, they have broadened the basis of selection of pupils for secondary education, thus affording children who have not had the best early educational opportunities a chance of displaying their abilities — surely not an undemocratic proceeding.

Bagley's address, well intentioned as it had been, seems to have brought forth a storm of criticism from those who recognized themselves as being among the group he termed "determinists." His answer to Whipple's attack has been noted; he has also answered Terman's reply; but, before we pass on to that, it is well to digress in order to read what President Cutten has to say about the measurement of intelligence and democracy in education. The reader may judge how seriously Dr. Cutten has considered the results of intelligence testing when an extract from his inaugural address (in October, 1922) is quoted.

It may be interesting to speculate concerning the effect of mental tests upon the problem of democracy. If the present hopes and expectations are realized they will result in a caste system as rigid as that of India, but on a rational and just basis. We are now examining children in the public schools, and find all

ranges of intelligence from imbecility to genius. We are told that the intelligence quotient of a child rarely changes, so that we are enabled to tell early in his life what the limit of intelligence of any person will be, and in a general way to what class of vocation he is best fitted, and, to a certain extent, destined. When the tests for vocational guidance are completed and developed, each boy and girl in school will be assigned to a vocation for which he is fitted, and, presuming that the tests are really efficient, he will in the future not attempt any work too advanced for his ability and hence make a failure of it, neither will he be found in an occupation too elementary for his ability and hence be dissatisfied. Economically, nothing could be more desirable. All differences in accomplishments or results from that which the intelligence quotients would indicate would be due to certain traits of character which intelligence tests do not measure, viz.: industry, perseverance, thoroughness, honesty.

One's intelligence quotient will eventually be known and persons will be classed thereby. Those of high intelligence will be directed into lines of occupation which call for leadership. Those persons will naturally be placed in the professions, and in leading positions in industry, commerce, and politics. Each person will then be directed on a scale of intelligence down to those whose work is of the most routine character of which an imbecile is capable. But what effect will this have on our so-called democracy? It must inevitably destroy universal adult suffrage, by

cutting off at least twenty-five percentage of the adults, those whose intelligence is so low as to be incapable of comprehending the significance of the ballot. On the other hand, it will throw the burden and responsibility of the government where it belongs, on those of high intelligence, and we come back again to the rule of the aristocracy — this time the real and total aristocracy. For its own salvation the state must assume the obligation and responsibility of selecting this intellectual aristocracy, and having selected it see that it is properly trained.

The above quotation is but one of a number which might have been cited from recent addresses of university and college administrators. That the "importance of the I.Q." has been well established in the minds of many of these leaders is plainly evident. But to get back to the discussion between Bagley and the "determinists." In December, 1922, Bagley endeavored to make a rejoinder to Terman. In this, he again attempted to throw more light upon the meaning of the terminology used in his much-discussed address. After clearing up certain technical differences in points of view, he concludes with these words:

If it is desirable that there should be more high I.Q.'s among our effective leaders, the best way to get them, I contend, is to educate the rank and file justly to evaluate and select them rather than to imbue

these gifted children with the notion that they were Heaven-sent to lead their dull fellows. Professor Terman accuses me of confusing "leadership" with "drivership." Well, whether the common man follows blindly or is blindly driven seems to me quite inconsequential. It is the tacit and nonchalant assumption of necessary and irremediable blindness that I am calling into question. It is my contention that the common mind of humanity has already demonstrated its ability to think for itself; that universal education can train it to think more clearly and in larger units — in any case, that universal education can give it a common stock of dependable ideas with which to do its collective thinking — and that the first and foremost task of education is to do this job passing well.

I should be remiss if I were to close this rejoinder without again stating my conviction that mental tests have made very important contributions to educational progress. It would be the height of folly for education not to avail itself of these instruments in detecting individual differences in learning capacity whether such differences be innate or acquired. My quarrel is not with the tests, but with the fatalistic assumptions which are part of their "heredity." They derive from Galton, and they are over-burdened with Galtonian tradition. The present-day determinists repudiate some of Galton's teachings. One in particular to them is anathema, although it seems to me far worthier of survival than others that they retain. I refer to Galton's theory that genius

is eminently well qualified to look after itself — that it is bound to come into its kingdom in spite of almost every obstacle short of premature death. I have a suspicion that Galton came pretty close to the truth here. On the other hand, there is the great rank and file of common or average intelligence. To endow the masses with genius is biologically impossible; but to endow the masses with the fruits of genius is both educationally possible and socially most profitable. The mental tests will help most if they aid the teacher in discharging this transcendent duty. They will render a gratuitous and disastrous disservice if they encourage in the teacher the conviction that the illumination of common minds is either an impossible or a relatively unimportant task. . . . My own criticisms of the movement were mild as compared with the appalling fallacies and inconsistencies pointed out by recognized authorities in the field itself.

That there is great difference of opinion among those who have taken part in this discussion is more apparent than real. It is unfortunate that, with all our clever phrases, we are unable to state opinions with sufficient clarity that they may be fully and fairly understood. Doubtless the sympathy of most of the writers quoted is with the school child as an individual. The differences, as they seem to the writer, are those of method. In his reply to Terman, Bagley quoted Bode in these words:

The right of the individual to share in the spiritual life of the race is really the issue. This is *the* important item in any properly constructed program of education.

What constitutes democracy in education, Bode has presented himself. He says:

A system is not democratic simply because it is made available to everybody or because it is administered without distinction of persons. In a Spartan scheme of education all are included and all are treated equally, but it is not democratic because the individual is subordinated, is made a means to an end. To be truly democratic, education must treat the individual himself as the end and set itself the task of preparing for him that intellectual and emotional sharing in the life and affairs of men which embodies the spirit of the Golden Rule. In proportion as common interests are permitted to outweigh special interests, the individual is becoming humanized and the successive adjustments of life will be made in the direction of democracy and in accordance with the needs of an expanding life.

There is nothing about this view which should arouse serious objection from the "determinists." It agrees with their concepts of "vertical" and "horizontal" growth. The special interests they would provide through the differentiated curriculum are not designed to outweigh the com-

mon interests. The special training suggested by Holmes and Counts is not proposed as a substitute for that previously designed to include the unfortunately dull. While Terman and Whipple and Haggerty maintain that they believe native ability has been measured by intelligence tests, they admit that it has not been measured directly. The difficulty, then, seems to lie in the decision as to whether or not certain assumptions and hypotheses are tenable. Townsend's attempt to show that branding with inferiority the child who does poorly on a test is a practical fallacy is acceptable to Terman, Whipple, and Haggerty if "all the facts are not in," and they have admitted that the non-intellectual factors which go to make for success have not been measured by their instruments.

Thorndike, himself an educational psychologist and a test-maker, has warned us against the many inadequate and even fantastic procedures parading behind the banner of educational science. Cattell, who indirectly helped to start this argument by devising some of the earlier tests but who has refrained from taking part in it, states in an unpublished annual report to the newly formed Psychological Corporation:

We can at present make certain standard determinations with the same degree of accuracy as the physician can diagnose a disease or a chemist analyse a water supply. The army intelligence tests have put psychology on the map of the United States, extending in some cases beyond these limits to fairyland. However little some of us may like newspaper and magazine exploitation of the assumption that psychologists have proved that a mental age of thirteen is prescribed by heredity for half the adult population, that ninety-five per cent of the people are below the average in ability, that clerks have been proved to be more intelligent than skilled mechanics, and the like, we may at least hope that this publicity will ultimately lead to an understanding of the proper use of psychological tests. Even the pretensions of ignoramuses and charlatans may be voices crying from the wilderness to make straight the way for psychology.

The assumptions listed by Cattell are among those which have been attacked by Bagley, and, in view of the fact that these assumptions are admitted by psychologists to be fallacious, it is easy to agree with him. On the other hand, it is not difficult to agree with Whipple and Terman in other phases of the argument which they present. In fact, in the attempt to recount the arguments presented by both "sides," the writer finds himself in agreement with practically all that has

been said. Such a state of conviction would hardly be possible if there were exceeding disparity in the views outlined. In the wholesale consumption of the "staff of life," it cannot make a great difference whether one prefers bread or rolls. The methods of preparing the school child for his future life are not so important if the same end is attained.

Certain statements have been quoted in this chapter, however, to which the writer does not wish to lend approbation. The inference made by Townsend that the mental test psychologist always concludes that "something is wrong with the child" when he registers an unsatisfactory score is not altogether fair. Many instances have been cited in which psychologists have used test results to show that there is "something wrong with the school," i.e., the school fails to provide a type of training best suited to inferior mentality and special classes or special schools are needed to deal with such inferiority. Something may be wrong with both the child and the school, and it is the psychologist's duty to inform us of these shortcomings. Townsend may have had in mind, however, the decrees of pseudo-psychologists, and that is quite another matter.

Again, whether Dr. Rusk did or did not make

the statement that seventy per cent of the children of England would never develop any more intelligence than that which should be possessed at the age of fourteen, such wide-sweeping assertions have been made in America, as Cattell has shown, and the evidence we have thus far been able to collect does not substantiate them. Even Terman, who was included by Bagley among those who have made assumptions not founded on fact, has been sixty-eight per cent more conservative than that. He says, in this regard: "Wherever intelligence tests have been made in considerable number in the schools, they have shown that not far from two per cent of the children enrolled have a grade of intelligence which, however long they live, will never develop beyond the level which is normal for the average child of eleven or twelve years." This may, as Bagley has intimated, make it possible to "stamp the child of twelve with the brand of permanent inferiority," but two per cent is not high, and it is assumed that Bagley will admit that the number of subnormals in our schools may reach such proportions.

It depends, however, on the meaning Bagley would attach to the word "subnormal." In the address which aroused so much consternation

among psychologists, he said: "Men and women of average or below-average mentality may possess these qualities (human and moral) in such abundance that they become leaders inevitably." If "average" has the same meaning as "normal," "below-average" may be thought of as synonymous with "subnormal," and Terman would not agree that those children who have been declared to be intellectually subnormal will ever become leaders. Again, the disagreement lies in the interpretation of meanings. Whipple asserts that a nation should provide "the best educational training for those of the rising generation that show promise of intellectual leadership." Bagley raises the question as to the method of procedure in locating this "promise of intellectual leadership"; if it is through intelligence tests alone, he will not subscribe to it, because of the other factors not measured. Terman does not appear to be in agreement with Bagley on this point, for in his answer to Bagley he does not make clear that he, for one, is considering other factors than absolute or even relative intelligence. In the survey in which Terman is now engaged, that of locating evidences of genius, he specifically states that he is employing many methods of finding superior talent other than the instruments under

discussion. In other words, he is not only measuring intelligence with the tools he has but is attempting to measure these other "human and moral qualities" listed by Bagley.

President Cutten offers a compromise between the "seventy per centers" and the "two per centers" by stating that the intelligence of twenty-five per cent of American adults is so low as to be incapable of comprehending the significance of the ballot. This is a strong statement, but the speaker does not stop there. He adds that "we are able to tell . . . in a general way to what class or vocation he (the child) is best fitted and, to a certain extent, destined. When the tests for vocational guidance are completed and developed, each boy and girl in school will be assigned to a vocation for which he is fitted." Nothing could be much farther from the intention of those psychologists who are constantly engaged in testing for the intelligence of school children. But it is the price that is paid for having spoken before the facts were all in. Many administrators, both in the higher institutions and the public schools, have expanded this idea of vocational guidance through the use of test results until, if it were generally accepted, we would have nothing short of a type of Prussian control. Bismarck, in his rule

of iron, did not exceed these expectations, nor Thomas More, in his conception of Utopia, these dreams. Vocational guidance, when interpreted by the best of psychological authorities, does not mean vocational compulsion. It is a fact that we should get straight.

Rusk may well have turned his attention to President Cutten's address as a sample of the Platonic attitude toward an educational caste system. He is not minimizing the strength of Bagley's arguments by inferring that he (Bagley) is a modern Plato in this one respect. For it was this type of assumption given widespread publicity by President Cutten that Bagley was attacking. Rusk seems to have reached over the head of Cutten to "get back" at Bagley. The three of them should be able to form a "vicious circle," which, however, is not the desire of those who want more light on this profound problem.

If we were to select one prominent point on which there exists a real difference of opinion, doubtless it would be the question as to whether intelligence tests measure native ability. Bagley contends "that what the tests *directly* measure is not native ability but acquired ability." The writer would amplify this statement to read that "the tests *directly* measure ability acquired in the

classroom." What they measure indirectly, such as native ability and home training, is not founded on fact but upon assumption. (The individual tests given to children younger than school age would, of course, measure ability not acquired in the classroom.) The words of Whipple and Terman may be interpreted to mean that they have merely assumed that native ability is measured, and Bagley would not deny them the right to assume. Rather, he has sent out a timely warning that the assumptions of psychological science must be sound if the psychologists of individual differences expect the support of leaders in other divisions of educational endeavor.

Which of the two schools of thought represent democracy in education in the best way need not be determined until a definition of democracy is agreed upon. There is some ground for the "determinist's" belief that the only real democracy in education consists in an "equity of opportunity," the opportunity for growth unlimited except by the fact that the possibilities for growth are known. The child, as Holmes suggests, need not be told what his limitations are. Bagley would give the child unlimited freedom to grow in intellectual power and grace, save for the limits the social order has already placed around

him, and at least until the tools employed to establish his limitations have become more perfectly standardized and universally acceptable.

A bird's-eye view of this absorbing problem is difficult to obtain. If intelligence, as Spearman asserts, is a function, and if it is the capacity of the individual to adapt himself to new situations, and this "general factor" has been subjected to measurement by instruments partially standardized and yielding limited values, are the plans proposed by psychologists for the reorganization of schools and the reclassification of school children justifiable? Within certain limitations, the writer holds that they are. They are when it is admitted that intelligence has been tested in so far as integral parts of it have been tested. They are when it is understood that all of a child's various mental abilities and some of his "human and moral qualities" have not been measured by these instruments, and when these qualities are included in establishment of standards for reclassification. They are when the instruments are administered by persons who understand them perfectly and who are able to interpret the results in light of the facts assembled. But they are not when the instruments are employed to produce evidence of correlation between elements

in which there is no relationship; and they are not when the instruments are employed to make deductions concerning traits over which the "controlled conditions" of examining hold no sway.

As an illustration of the last point, there has been a marked tendency to employ the results of intelligence tests in occupational analyses. Success on an intelligence test requiring performance with a pencil on paper is, in many instances, due to what seems to be transfer of training or to the recency with which the individuals have met similar situations, which may account for the hasty conclusion that "clerks have been proved to be more intelligent than skilled mechanics." In announcing the results of the army tests, the effects of practice were little considered, if at all. The median intelligence scores of various occupational groups present an interesting array of figures. The professional groups, with their superior amount of school training, examined by an instrument which tested classroom ability, would naturally rank much higher than skilled and unskilled laborers. Certain advantages to vocational guidance accrue from a knowledge of these things, but to determine definitely that, because a school child registers a score equivalent to the median score of unskilled laborers, he should be

directed to enter into unskilled labor, or the many counterparts of such determination, is the boldest attack on democracy in education America has yet seen. In all fairness it may be said that such proposals do not come from the psychologists quoted in this volume. Rather, they usually emanate from sources which are highly questionable, from two classes of individuals — those who have so little knowledge of the field in which they work that they are not competent to weigh values, and those who have set out willfully to undermine the conscientious efforts of scientific workers. Again, in all fairness, it may be said that such educational leaders as Bagley do not belong in either of these classifications.

If we can resolve the various proposals, arguments, and contentions, into a unified whole, democracy in education will necessarily take into account these two theses: (1) no child must be denied the right of *equality* of opportunity until it has been demonstrated beyond the shadow of a doubt that he is incapable of profiting by that opportunity; and (2) no child must be denied the right of *equity* of opportunity if it can be demonstrated that the administrative program necessary to that equity of opportunity is possible of achievement. Meanwhile scientists who are

seeking the truths that would make the dream of equity of opportunity a reality must face these truths when they find them and describe them to the world. There should be no subterfuge, and no desire to be the first to announce a discovery unless that discovery is founded upon undeniable and unquestionable facts. In such event, the intelligence movement will grow in power. Because of its wide appeal, teachers both trained and untrained in psychological measurement are likely to employ the instruments of testing. In that case what we need most is a clearing-house for interpretation, assumption, and hypothesis. Right-thinking and clarity of expression will do more to help us attain the democracy of equality and equity than confusion of thought and obscurity of meaning will accomplish in a thousand years.

REFERENCES

AYDELOTTE, F., "Better Training for Our Best Minds"; *School and Society*, November 5, 1921.

BAGLEY, W. C., "Educational Determinism; or Democracy and the I.Q."; *School and Society*, April 8, 1922.

"Educational Determinism Again: A Rejoinder to Professor Whipple's Reply"; *School and Society*, August 5, 1922.

"Professor Terman's Determinism: A Rejoinder"; *Journal of Educational Research*, December, 1922.

BEATTY, W. W., "Who Can Be Educated?"; *School and Society*, October 15, 1921.

BODE, B. H., *Fundamentals of Education*, p. 61.

COUNTS, G. S., "Education as an Individual Right"; *School and Society*, April 22, 1922.

CUTTEN, G. B., "The Reconstruction of Democracy"; *School and Society*, October 28, 1922.

DAVIDSON, P. E., "The Socialists on Equality of Educational Opportunity"; *School and Society*, April 2, 1921.

HAGGERTY, M. E., "Recent Developments in Measuring Human Capacities"; *Journal of Educational Research*, April, 1921.

HART, J. K., *Democracy in Education*, chaps. 36, 37.

HOLMES, H. W., "The General Philosophy of Grading and Promotion in Relation to Intelligence Testing"; *School and Society*, April 29, 1922.

KOHS, S. C., "The Educational Lock-Step and Teacher Conservation"; *School and Society*, February 12, 1921.

LIPPMANN, W., "A Defense of Education"; *The Century*, May, 1923.

MACDONALD, D. J., "Democracy in School Administration — Some Fundamental Principles"; *American School Board Journal*, September, 1921.

RUSK, R. R., "Dr. Bagley and Educational Determinism"; *School and Society*, October 14, 1922.

SNEDDEN, D., "Collegiate Education and Democracy"; *School and Society*, November 25, 1922.

SEASHORE, C. E., "Progressive Adjustment Versus Entrance Elimination in a State University"; *School and Society*, January 13, 1923.

TERMAN, L. M., *The Measurement of Intelligence*, p. 6.

"The Psychological Determinist; or Democracy and the I.Q."; *Journal of Educational Research*, June, 1922.

THORNDIKE, E. L., "Measurement in Education"; *Twenty-First Yearbook*, chap. 1.

TOWNSEND, H. G., "The Concept of Inferiority"; *School and Society*, February 4, 1922.

WHIPPLE, G. M., "Educational Determinism; a Discussion of Professor Bagley's Address at Chicago"; *School and Society*, June 3, 1922.

# OUTLINE

## I. THE MEANING OF INTELLIGENCE



## II. THE INSTRUMENTS FOR MEASURING INTELLIGENCE

## OUTLINE

### III. THE USE OF THE INSTRUMENTS



### IV. WHAT IS TESTED AND WHAT IS NOT TESTED



### V. THE DIFFICULTIES OF STANDARDIZATION

# OUTLINE

## VI. VALUES ACCRUING FROM THE MEASURING PROCESS



## VII. THE MEASUREMENT OF INTELLIGENCE AND THE AIMS OF EDUCATION

# OUTLINE

## VIII. THE MEASUREMENT OF INTELLIGENCE AND DEMOCRACY IN EDUCATION

# RIVERSIDE EDUCATIONAL MONOGRAPHS

Andress's The Teaching of Hygiene in the Grades
Atwood's The Theory and Practice of the Kindergarten
Bailey's Art Education
Betts's New Ideals in Rural Schools
Betts's The Recitation
Bloomfield's Vocational Guidance of Youth
Cabot's Volunteer Help to the Schools
Cole's Industrial Education in the Elementary School
Cooley's Language Teaching in the Grades
Cubberley's Changing Conceptions of Education
Cubberley's The Improvement of Rural Schools
Dewey's Interest and Effort in Education
Dewey's Moral Principles in Education
Dooley's The Education of the Ne'er-Do-Well
Earhart's Teaching Children to Study
Eliot's Education for Efficiency
Eliot's Concrete and Practical in Modern Education
Emerson's Education
Evans's The Teaching of High School Mathematics
Fairchild's The Teaching of Poetry in the High School
Fiske's The Meaning of Infancy
Freeman's The Teaching of Handwriting
Gates's Management of Smaller Schools
Haliburton and Smith's Teaching Poetry in the Grades
Hartwell's The Teaching of History
Haynes's Economics in the Secondary School
Hill's The Teaching of Civics
Hines's Measuring Intelligence
Horne's The Teacher as Artist
Hyde's The Teacher's Philosophy
Jenkins's Reading in the Primary Grades
Judd's The Evolution of a Democratic School System
Kendall and Stryker's History in the Elementary Grades
Kilpatrick's The Montessori System Examined
Leonard's English Composition as a Social Problem
Lewis's Democracy's High School
Losh and Weeks's Primary Number Projects
Maxwell's The Observation of Teaching
Maxwell's The Selection of Textbooks
Meredith's The Educational Bearings of Modern Psychology
Palmer's Ethical and Moral Instruction in the Schools
Palmer's Self-Cultivation in English
Palmer's The Ideal Teacher
Palmer's Trades and Professions
Perry's Status of the Teacher
Prosser's The Teacher and Old Age
Ruediger's Vitalized Teaching
Russell's Economy in Secondary Education
Smith's Establishing Industrial Schools
Snedden's The Problem of Vocational Education
Stockton's Project Work in Education
Stratton's Developing Mental Power
Suzzallo's The Teaching of Primary Arithmetic
Suzzallo's The Teaching of Spelling
Swift's Speech Defects in School Children
Terman's The Teacher's Health
Thorndike's Individuality
Tuell's The Study of Nations
Weeks's The People's School