THE MEASUREMENT *of* ADULT INTELLIGENCE

BY

DAVID WECHSLER

Chief Psychologist, Bellevue Psychiatric Hospital. Assistant Clinical Professor of Medical Psychology, New York University College of Medicine. Lecturer, School of Education, New York University.

SANS TACHE

THIRD EDITION

BALTIMORE

THE WILLIAMS & WILKINS COMPANY

1944

PREFACE TO THIRD EDITION

In preparing this third, "war" edition of the Measurement of Adult Intelligence my primary aim has been to increase the book's usefulness to those engaged in the field of applied psychometrics. I have borne in mind particularly the military applications of the Wechsler-Bellevue Scale, as it is now being used by some of the armed services of our own country as well as of Great Britain. Accordingly, most of the changes made will be found in the chapters dealing with *Clinical applications* and the *Problem of mental deterioration*. These chapters have been somewhat extended and in part rewritten. The table of 'signs' for different clinical entities has been completely revised and a new quantitative method for determining mental deterioration elaborated. Both of these presentations should be of value not only in the field of neuropsychiatric diagnosis but also in questions connected with rehabilitation.

In addition to the above, some changes have been made in the scoring of certain subtests of the Scale. The effect of these changes will be to increase the I.Q. range at the upper end of the Scale, without however altering significantly the norms for the population as a whole. In the case of one or two tests (noticeably, the Arithmetic) the change has necessitated slight alteration in the scoring of some of the subtest items. Little alteration has otherwise been made in the substance of the text. One table and two figures which were not particularly useful have been omitted and two new tables which should be of immediate and practical value added. Also added is a bibliography of most of the important studies with the Wechsler-Bellevue Scale, published up to March 1944. For editorial perusal of the text I am indebted to Miss Ventura Smith, psychologist to the Westport, Conn., public schools, who not only read the page proofs but was also kind enough to prepare a new index.

David Wechsler.

April, 1944

Made in the United States of America

First Edition, April, 1939
Second Edition, October, 1941
Reprinted, February, 1943
Third Edition, April, 1944
Reprinted, April, 1945
Reprinted, July, 1946
Reprinted, October, 1947

Composed and Printed at the
WAVERLY PRESS, INC.
for
The Williams & Wilkins Company
Baltimore, Md., U. S. A.

CONTENTS

PART I

PART II

PART III

PART I

THE NATURE AND CLASSIFICATION OF INTELLIGENCE

CHAPTER 1

THE NATURE OF INTELLIGENCE

Some years ago when interest in intelligence tests was at its height, a prominent psychologist is reported to have answered an inquiry as to what he meant by intelligence by saying that it is what intelligence tests measure. A similar attitude would not be maintained today by any considerable number of psychologists. But the continued failure of authors of intelligence tests to declare explicitly what they understand by general intelligence would almost compel one to assume that they still maintain this circular position. The lay person is entirely justified in asking, as he does, "How do you know that your test measures intelligence?", and every author of a test should be ready to answer the question, however imperfectly. Obviously, the more data the psychologist has, the easier his task will be. But he will be able to make no answer at all unless there is some provisional agreement between him and his challenger as to what they are willing to call intelligence, or at least intelligent behavior. We shall, therefore, begin by giving our own definition of intelligence and then consider its relation to the more important current theories on the subject.

Intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment. It is global[1] because it characterizes the individual's behavior as a whole; it is an aggregate because it is composed of elements or abilities which, though not entirely independent, are qualitatively differentiable. By measurement of these abilities, we ultimately evaluate intelligence. But intelligence is not identical with the mere sum of these abilities, however inclusive. There are three important reasons for this: (1) The ultimate products of intelligent behavior are not only a function of the number of abilities or their quality but also of the way in which they are combined, that is, upon their configuration. (2) Factors other than intellectual ability, for example, those of drive and incentive, enter into intelligent behavior. (3) Finally, while different orders of intelligent behavior may require varying degrees of intellectual ability, an excess of any given ability may add relatively little to the effectiveness

[1] In the adopted French sense of the term, "pertaining to or embracing the totality of a group of items or categories." Cf. Oxford Dictionary.

of the behavior as a whole. It would seem that, so far as general intelligence is concerned, intellectual ability as such merely enters as a necessary minimum. Thus, to act intelligently, one must be able to recall numerous items—i.e., have a retentive memory. But beyond a certain point this ability will not help much in coping with life situations successfully. This is true of even more important capacities, such as the ability to reason, particularly when specialized. The unusual reasoning abilities of the mathematician are more highly correlated with the thing that we ultimately measure as intelligence than sheer memory is, but possession of this ability is no surety that behavior as a whole will be very intelligent in the sense above defined. Every reader will be able to recall persons of high intellectual ability in some particular field, whom they would unhesitatingly characterize as below average in general intelligence.

Although intelligence is no mere sum of intellectual abilities, the only way we can evaluate it quantitatively is by the measurement of the various aspects of these abilities. There is no contradiction here unless we insist upon the identity of general intelligence and intellectual ability. We do not, for example, identify electricity with our modes of measuring it. Our measurements of electricity consist of quantitative records of its chemical, thermal and magnetic effects. But these effects are not identical with the "stuff" which produced them. General intelligence, like electricity, may be regarded as a kind of energy. We do not know what the ultimate nature of this energy is, but as in the case of electricity, we know it by the things it does or, better, by the things it enables us to do—such as making appropriate associations between events, drawing correct inferences from propositions, understanding the meaning of words, solving mathematical problems or building bridges. These are the effects of intelligence in the same sense as chemical dissociation, heat, and magnetic fields are the effects of electricity, but psychologists prefer the term *mental products*. We know intelligence by what it enables us to do.

Professor Thorndike was the first to develop clearly the idea that the measurement of intelligence consists essentially of some qualitative and quantitative evaluation of mental productions in terms of their number, and the excellence or speed with which they are effected. That is the only function which any measure of intelligence can possibly have. Abilities are merely these mental products sorted into different classes or types of operation. Thus, the class of operations which consists of effectually associating one fact with another and recalling either or both at an appropriate time is called learning; that of drawing inferences or educing relations between them, reasoning ability; that of merely retaining them, memory. The older psychologists were inclined to use a relatively small number of such classes based primarily on the type of mental process supposedly involved. More recently psychologists have begun to emphasize not only the processes but the content as well. They speak not only of memory but of auditory memory; not only of reasoning but of abstract, verbal or arithmetical reasoning. In a like manner some psychologists have begun to distinguish various kinds of intelligence. Thorndike, for example, has suggested subdividing intelligence into three main types: (1) abstract or verbal intelligence, involving facility in the use of symbols; (2) practical intelligence, involving facility in manipulating objects; (3) social intelligence, involving facility in dealing with human beings. The significant thing about this classification is that it emphasizes *what* a person can do as well as *how* he can do it. This distinction between function and content is fully justified by experimental evidence. The rating which an individual attains on an intelligence examination depends to a considerable degree on the type of test used. His score on a test made up largely of verbal items may differ significantly from that obtained on a test involving questions of social comprehension and still more from another test made up of items involving predominantly psychomotor reactions and the perception of spatial relationships.

Though test results show that the rating which an individual attains will frequently depend upon the type of intelligence test used, they also show a contrary tendency. When large numbers of individuals are examined with a variety of intelligence tests, those who make high scores on any one of them tend to make high scores on the remaining ones; and the same holds for those who make low and intermediate scores. This dual characteristic of human abilities—their specificity on the one hand and interdependence on the other—has long been appreciated by psychologists. But unfortunately, the reaction to this observation was not to accept it as a fact, but rather as a logical dilemma from which one had to escape. The older writers tried to escape it by accepting the scholastic formulated faculties; the modern ones, by their theory of independent unit or group traits. But more than 30 years ago Professor Carl Spearman put an end to the dilemma by showing, through rigorous mathematical proof, that all intellectual abilities could be expressed as functions of two factors, one a *general* or intellectual factor common to every ability, and another a *specific* factor, specific to any particular ability and "in every case different from that of all others". This proof first appeared as a brief article in the American Journal of Psychology (1904). It has since been subjected to a great amount of discussion, criticism and experimental investigation. We cannot enter into all this here, but can only indicate our own position by saying that Professor Spearman's generalized proof of the two-factor theory of human abilities constitutes one of the great discoveries of psychology.

As has often been the case in the history of science, the proof of the two-factor theory, in addition to being a discovery, was also an explicit formulation of an hypothesis which workers in the field had unknowingly been assuming for some time. The fact is, that from the day psychologists began to use a series of tests for measuring intelligence, they necessarily assumed the existence of a general or common factor. This becomes immediately apparent if one recalls what the actual contents of intelligence tests are. They consist of various intellectual tasks which we call tests that require the subject to do such things as to define words, reproduce facts from memory, solve problems in arithmetic and recognize likenesses and differences. The variety of tasks used, their difficulty and manner of presentation varies with the type of scale employed. But so far as measuring intelligence is concerned, these specific tasks are only a means to an end. Their object is not to test a person's memory, judgment or reasoning ability, but to measure something which it is hoped will emerge from the sum total of the subject's performance, namely, his general intelligence. One of the greatest contributions of Binet was his intuitive assumption that in the selection of tests, it made little difference what sort of tasks you used, provided that in some way it was a measure of the child's general intelligence. This explains in part the large variety of tasks employed in the original Binet scale. It also accounts for the fact why certain types of items which were found useful at one age level, were not necessarily employed at other age levels. More important than either of these details is the fact that for all practical purposes, the combining of a variety of tests into a single measure of intelligence, *ipso facto*, presupposes a certain functional unity or equivalence between them.

The functional equivalence of the test items, an assumption implicit not only in the Binet Scale but in any scale which is composed of a variety or pool of intellectual tasks, is absolutely necessary for the validation of the arithmetic employed in arriving at a final measure of intelligence. This arithmetic consists, first, of assigning some numerical value to every correct response; secondly, of adding the partial credits so obtained into a simple sum; and, thirdly, of treating equal sums as equivalent, irrespective of the nature of the test items which contribute to the total. For example, every test passed on the Stanford-Binet (between ages 3 and 10) contributes two months to the mental age (M.A.) score of the subject, irrespective as to whether the test passed calls for a repetition of a series of digits, the copying of a square, the definition of a word, or the correct reply to a common sense question. To all intents and purposes, therefore, the simple addition of these groups necessarily assumes an arithmetical equivalence of the test items so combined. If the different tests were taken to represent generically different entities, one could no more add the values assigned to them in order to obtain an M.A. total than one could add 2 dogs, 3 cats and 4 elephants, and expect the unqualified answer of 9. That, of course, does not mean that their addition is impossible. If instead of being concerned with the characteristics of the dog, the cat and the elephant, which differentiate them from one another, we restrict our interest to those which they all have in common, we can say that 2 dogs, 3 cats and 4 elephants make 9 animals. The reason we can get an answer of 9 here is because dogs, cats and elephants are in fact all animals. The addition would no longer be possible if for cats we were to substitute turnips.

The same principle is involved when we attempt to add up the number of tests correctly passed on an intelligence scale into a simple sum. The reason we can add together scores obtained from tests requiring such seemingly different abilities as those involved in solving arithmetic problems, repeating digits and defining words, is because they are in fact alike in certain ways. They are alike in that they are all measures of general intelligence. This means that all must have a common characteristic, or to use the current psychological term, a common factor. Professor Spearman has shown that such a common factor not only has to be assumed in any attempt to measure general intelligence by means of tests, but has demonstrated that its presence can always be revealed through appropriate statistical procedures. The common factor turns out to be a recurrent mathematical quantity which can be "extricated" from the tests by special correlational methods,—a quantity which he has called "g". Just what "g" is psychologically and to what extent it may be identified with general intelligence, are still matters of speculation and dispute. As will be seen shortly, the present writer is far from being in full agreement either with Professor Spearman's concept of general intelligence or even with his views regarding the best mode of measuring it, but as regards the demonstration of the existence of "g" as a common factor, there seems to be no possibility of doubt. Psychometrics, without it, loses its basic prop.

A few words as to the nature of "g". First and foremost, it is a purely mathematical quantity, "originally intended to explain correlations that exist between the most diverse sorts of cognitive performance", which recurs in all data obtainable from measures of intellectual ability. In this respect it may be said to be similar, or at least comparable, to some of the constants met with in physical and more particularly atomic measurements. But "g" is also more specific than that. It is evidently a kind of something that must be posited to explain the effects of mental work or the operations of the mind. It is, therefore, a kind of energy, or more correctly, a measure of the same. This is the second and more general interpretation of "g". Combining these two conceptions, we may say that "g" is a psychomathematical quantity which measures the mind's capacity to do intellectual work.

Everybody will agree that the capacity to do intellectual work is a necessary and important sign of general intelligence. The question is whether it is the only important or paramount factor. In this writer's opinion it is not. Professor Spearman seemingly thinks it is, although on this point he has failed to declare himself unequivocally. On the one hand, he states, "Such a factor as this ["g"] can scarcely be given the title of intelligence at all." But after having said this, he devotes several chapters[2] in an attempt to prove that the best tests of intelligence are precisely those which contain the largest amounts of "g". If this is so, then for all practical purposes, "g" and general intelligence may be said to be equivalent. This equivalence, indeed, is implied by the mathematical relationship of the "g" and "s" factors in the two-factor theory. According to this relationship an intelligence scale made up of a large number of tests especially rich in "g" would in the end be a measure of "g" exclusively.[3] In the writer's opinion, such a scale would not be a very good measure of general intelligence because it would eliminate a number of abilities essential for effective behavior.

The view that other salient factors besides "g" enter into measures of intelligence is based on several sources of evidence. The first is clinical. We know from experience that individuals attaining identical scores on intelligence tests cannot always be classified in the same way. This is perhaps most obvious in cases where test results call for practical action, as for example when they are used as a basis for deciding

[2] Spearman, C.: *The Abilities of Man.* New York, 1927.

[3] For, by pooling such tests, the "g" factor (being common) becomes cumulative, whereas the specific factors (being incidental) tend to cancel each other.

whether or not a subject should be committed to an institution for mental defectives. In such cases the test results, e.g., a Binet I.Q., cannot be used as the sole criterion. One child with an I.Q. of 75 may be definitely defective while another with an identical I.Q., or indeed one 5 or 10 points less, be far from so classifiable. Of course, the objection may be made that the classification of mental deficiency is in part a social diagnosis. But is not the capacity for social adaptation also a sign of intelligence? Should not the capacity to avoid mischief and the ability to work persistently at a task, enter into one's definition of general intelligence just as well as the ability to define words, and perceive analogies? The clinician's answer has always been "yes", and by so saying, he has implicitly assumed that there are other factors besides the intellective "g" which enter into intelligent behavior. Hitherto he has been unable to demonstrate their existence experimentally. In recent years, however, thanks to new correlational techniques, especially the method of factorial analysis as developed by Professor Thurstone in this country, a beginning has been made. Of particular importance has been the work of W. P. Alexander whose monograph on *Intelligence, Concrete and Abstract*[4] is in many ways fundamental.

Alexander set himself the problem of testing experimentally the evidence for and against the main theories now currently favored in psychological circles. The first of these is Professor Spearman's *two-factor theory* which we have already referred to. The other is the *unique traits theory,* according to which intelligence involves several abilities or factors, each *independent* of one another. More specifically, his investigation took the form of an experimental study to determine whether test results supported the view that "practical" and "verbal" intelligence were each distinct and independent capacities, or the view of Spearman according to which both were essentially the same in that they were not independent capacities but only differed with respect to their non-intellective or specific factors.

Alexander's findings were extremely interesting. They confirmed Spearman's main contention that there was one and only one common factor in all measures of intelligence and, at the same time, showed that this factor alone is not sufficient to explain the total correlational variance which existed between the tests used to measure intelligence. In addition to the common factor there are seemingly other broad factors which, while not showing the same generality, are nevertheless recurrent in a significant number of abilities which form subgroups or "com-

[4] *Brit. J. Pyschol.*, 1935.

munal clusters". The individual tests by which these abilities are measured contain a common factor of their own with respect to which they function in much the same way. Alexander has termed abilities involved in tests showing such similarity of function *functional unities.* Thus, verbal ability is one functional unity, practical ability another, and so on. But while each of these functional unities requires a separate factor to take care of its respective contribution to any global measure of intelligence, they are nevertheless "definitely related", that is, correlated with one another.[5] This means that they cannot be unitary traits in the sense implied by the unique traits theory. On the other hand, neither can they be considered as specific factors in the sense required by Spearman's two-factor theory. For, these factors, unlike the "s" factors, actually contribute a considerable amount to the correlation variance of the test composites of which they form a part.

Another important conclusion suggested by Alexander's investigation was that in order to account for the complete intercorrelation variance found among any large battery of intelligence tests one has to posit other factors in addition to the purely intellectual ones. After eliminating the general factor ("g"), and such other factors[6] as were contributed by the above described "functional unities", Alexander found that a considerable amount of his total intercorrelational variance was still unaccounted for. In addition to these factors there were apparently certain other supplementary global ones which, though not directly measurable, nevertheless contributed significant amounts to the total variance of the observed data. These latter factors he has provisionally labeled "X" and "Z". They cover such items as the subject's interest in doing the tasks set, his persistence in attacking them and his zest and desire to succeed,—items which might more familiarly be described as temperamental or personality factors, but which nevertheless must be recognized as important in all actual measures of intelligence. For this reason, one might appropriately refer to them as the non intellective-factors or, more specifically, as the *non-intellective factors in* general intelligence.[7]

[5] Thus verbal ability correlates with practical ability to the extent of .50.

[6] These were primarily the factors "v", common to tests involving verbal ability, and "f" common to tests purporting to measure practical ability.

[7] For further evidence as to the existence of these factors, see D. Wechsler: The non-intellective factors in general intelligence. *J. Abnorm. and Soc. Psychol.*, 1943, **38**: 100–104.

It thus appears that the entity or quantity which we are able to measure by intelligence tests is not a simple quantity. Certainly it is not something which can be expressed by one single factor alone, say "g", whether you define it in its most general terms as mental energy, the ability to educe relations or merely as the intellective factor. Intelligence is all this and yet something more. It is the ability to utilize this energy or to exercise this ability in contextual situations,—situations that have content and purpose as well as form and meaning. To concede as much is to admit that any practical definition of intelligence must be fundamentally a biological one in the widest sense of the term. That has been the hypothesis assumed in the construction of the intelligence scales that are to be described in the ensuing pages. We think that they measure general intelligence in the sense defined above. We shall not, however, claim that they measure all that goes to make up general intelligence, because no tests at present are capable of doing it. The only thing we can ask of an intelligence scale is that it measures sufficient portions of intelligence to enable us to use it as a fairly reliable index of the individual's global capacity.

It is important to realize that intelligence tests do not and cannot be expected to measure all of intelligence, but it is of equal importance to emphasize that they measure a great deal more than the delimited capacities to which contemporaneous theory seems desirous of restricting them. Intelligence tests measure more than mere learning ability or reasoning ability or even general intellectual ability; in addition, they inevitably measure a number of other capacities which cannot be defined as either purely cognitive or intellective,—abilities heavily loaded with factors like "X" and "Z", the non-intellective factors mentioned above. Hitherto, authors of intelligence scales when recognizing this situation, looked upon these factors as disturbing elements and tried as far as possible to eliminate them. Unfortunately, experience has shown that the more successful one is in excluding these factors, the less effective are the resulting tests as measures of general intelligence. What are needed are not tests from which the non-intellective factors have been eliminated (even if that were possible), but, on the contrary, tests in which these factors are clearly present and objectively appraisable. The performance tests described in Part II of this volume are an attempt in this direction. They are only a beginning; but the results already obtained show that tests of verbal ability, abstract reasoning and the like when used alone in a general intelligence examination give only an incomplete picture of the individual's capacity for effective adjustment and achievement.

We have concluded above that intelligence tests cannot measure all of intelligence. We must further qualify this conclusion by the statement that at different ages they measure different portions of it. In general it might be stated that as the individual grows older our tests become less and less an effective measure of his global intelligence. The failure to realize this has led to a great deal of confusion in interpreting test results, particularly when used for the evaluation of adult intelligence. The difficulty is most acutely apparent when intelligence levels are given in the now almost universal notation of mental age. Most psychologists are aware of the fact that when an adult of 30 scores a mental age of 12, and a child of 12 scores a mental age of 12, their intelligence is not identical, yet there does not seem to be any general understanding as to why they are not identical. If what we have said above is correct, the answer is not far to find. The basic reason a mental age of 12 at 12 does not mean the same thing as a mental age of 12 at 30, is that the measured abilities and hence the M.A. scores represent different portions of the subjects' respective total intelligence. At age 12 the tests are capable of tapping far more of the individual's capacities than at 30.[8] How much more we cannot say without further investigation. But even if the differences were small it would challenge our entire present method of defining at least adult intelligence in terms of M.A. levels. The same observation may be made as regards intelligence quotients with even greater force, because as we shall have occasion to see, the I.Q. concept as a mental age score divided by a chronological age score, presupposes a constancy of relationship between the two, which in point of fact does not exist.

Sufficient has been said to show that the definition of general intelligence, far from being a mere theoretical question, is one which enters immediately into any practical attempt at measurement. It is particularly important when we come to measuring adult intelligence, because we are at once confronted with a wider range of criteria against which our definition may be checked. But before this checking can be done, the tests themselves must of course have been available. The scales presented in the following pages are an attempt to supply this. The extent of the present need for them will form the content of our next chapter.

It should be added that the same might be said of age 6 as compared to age 12.

CHAPTER 2

NEED FOR AN ADULT INTELLIGENCE TEST

Although the earliest investigations in the field of psychometrics were made largely with adult subjects, the great bulk of the test data which now forms the basis of intelligence scales, has been derived from the examination of school children. The reasons for this are several. Perhaps the most important is the relative ease with which one may obtain young subjects; children are nearly always available through the schools. Another is the fact that it is generally much easier to devise children's tests than adults' tests, both from the point of view of the definition of ability measured, as well as the likelihood of interest and appeal. Finally, there is the fact that the results obtained by the testing of children lend themselves more readily to concrete applications.

Demonstration of the value of intelligence tests for children has come from two important fields. The first of these is the school. From the very beginning it was apparent that intelligence tests, particularly of the Binet type, correlated very highly with scholastic achievement. A well trained psychologist after an hour of standardized testing could often more accurately differentiate between the superior, the average, and the dull than the teacher who may have had the children under daily observation for a period of several months. Here, then, was an instrument that could be of great value in the matter of proper grading of school children, particularly at the outset of their training. Naturally psychologists and educators who were at all interested in tests, devoted a large part of their energies to the development of these important possibilities.

The other field in which intelligence tests early demonstrated their usefulness was in the diagnosis of mental deficiency. Tests supplied for the first time a quantitative measure of the degree of mental defect. And here the tests seemed not only applicable to the diagnosis of retardation among children, but for the classification of defective adults as well. The fact that the tests were originally standardized on children, was looked upon by those who applied them as no great limitation upon their usefulness. Indeed, because the tests made possible a comparison of adult functioning with that of children, they offered the clinician a new basis for classifying degrees of deficiency. We refer, of course, to the definition of degrees of deficiency in terms of mental age levels.