

I.

HISTORICAL BACKGROUND

Uses of Psychological Tests

Psychological tests have been devised and are used primarily for the determination and analysis of individual differences in general intelligence, specific aptitudes, educational achievement, vocational fitness, and nonintellectual personality traits. Tests also have long been used for a variety of psychological, educational, cultural, sociological, and employment studies of groups rather than for the study of a particular individual. Among these studies of groups, the following have been most common and include the most important fields of investigation: the nature and course of mental development; intellectual and nonintellectual personality differences associated with age, sex, and racial membership; differences that might be attributed to hereditary or to environmental factors; differences among persons at different occupational levels and among their children; intellectual and other personality traits of atypical groups such as the mentally gifted, the mentally retarded, the neurotic, and the psychotic.

Psychological tests, especially those of general intelligence and of specific aptitudes, have had very extensive use in educational classification, selection, and planning, from the first grade (and sometimes earlier) through the university. Prior to World War II, schools and colleges were the largest users of psychological tests. During and after World War II, however, so many types of tests were administered to so many men and

women in all branches of the military services that the armed forces, along with educational institutions, must now be regarded as the major users of psychological devices.

When tests are used for the determination and analysis of an individual's intellectual abilities or nonintellectual traits, the purpose might be to provide educational and vocational guidance; to place an individual in a special class for superior pupils or in one for the mentally retarded; to discern weaknesses in order to provide remedial instruction; or to discover causes, intellectual or otherwise, which might account for behavior problems in school.

In clinics, psychological tests are used primarily for individual diagnosis of factors associated with personal problems of learning, behavior, attitudes, or specific interpersonal relations.

In business and industry, tests are helpful in selecting and classifying personnel for placement in jobs that range from the simpler semiskilled to the highly skilled, from the selection of filing clerks and salespersons to top management. For any of these positions, however, test results are only one source of information—though an important one.

The foregoing discussion emphasizes the fact that psychological tests and testing play a significant role in a wide variety of situations and can significantly affect the lives of many persons. But even though they are significant educational, vocational, and diagnostic assets today, psychological tests did not begin to assume appreciable significance until about 1910-15.

The Nineteenth Century

Although the fact that persons differ in intellectual and other psychological characteristics had been apparent to observers for many centuries, it was only about a hundred years ago that these differences were first studied scientifically and subjected to measurement and objective evaluation.

Francis Galton (1822-1911) was the first scientist to undertake systematic and statistical investigations of individual differences. He was preceded, before the middle of the nineteenth century, by other men who are important in the history of psychology; but these men, who belonged to one of two groups, were not concerned with devising means of measuring individual differences. Some were nonexperimental, speculative psychologists who were concerned largely with problems of the dualism of mind and matter, the nature of ideas, intellectual "faculties," and classical associationism. Others, though experimentally oriented,

directed their attention to general problems and theories rather than to variations and differences in human abilities.

Among these was Ernst Heinrich Weber (1795–1878), educated as an anatomist and physiologist, who experimented on weight discrimination, vision, hearing, and the “two-point threshold” of the skin. He is best remembered for his quantitative experimental approach to psychological problems and for what we know as Weber’s law.¹ Gustav Theodor Fechner (1801–87), who started his career in physics and chemistry, was basically concerned with the application of the exact methods of the natural sciences to the study of man’s “inner world,” that is, the relations of mental processes to physical phenomena. Johannes Müller (1801–58), a professor of physiology, was especially interested in the physiology of the senses and in reflex action. In his significant experiments in space perception, he attempted to reconcile the opposed theories of “nativism” versus “empiricism.” William Hamilton (1788–1856) and James Mill (1773–1836) were concerned with reformulating more completely and rigorously the classical association theory.

One of the most significant writers in psychology at mid-nineteenth century was Alexander Bain (1818–1903), who was Professor of Logic, Mental Philosophy, and English Literature in Aberdeen University. His two most distinguished works were *The Senses and the Intellect* (1855) and *The Emotions and the Will* (1859). Bain’s approach was principally through physiology; he utilized, organized, and interpreted findings of the German experimentalists in a systematic restatement of associationism. Perhaps Bain’s most important contribution was his pioneering effort to contain the entire range of human experience within a system of psychology.

Although Wundt’s principal work was done somewhat later than Galton’s, he is significant not only for his actual contributions but also as an example of nineteenth-century neglect of differential psychology. Wundt (1832–1920), who established the first laboratory of psychology, in 1879, at Leipzig University, employed physiological methods and introspection in his and his students’ research. He held that “. . . a genuinely psychological experiment involved an objectively knowable and preferably a measurable stimulus, applied under [specific] conditions, resulting in a response objectively known and measured. But there were certain intervening steps which [could be known only] through introspection, sometimes supplemented by instrumentation” (31, p. 161). Thus, Wundt’s method emphasized the necessity of knowing and stating con-

¹ This law states that the least added difference of a stimulus that can be noticed is a constant proportional part of the original stimulus.

ssciously experienced events as they are related to objective and measurable stimuli and responses. For Wundt, introspection became the most important method of the experimental psychologist. These methods he applied to experimental study of vision, hearing, reaction time, psychophysical problems,² and to the analysis of word associations. It is interesting to note that one of Wundt's students from the United States, James McKeen Cattell (1860-1944), was impressed by the range of individual differences he found in his experiments. Although he was discouraged by Wundt from pursuing the subject, he persisted in doing so for many years.

These several examples will suffice to indicate the major interests of nineteenth-century psychologists, from which the pioneers in psychological testing had to break away. Yet the work of these early psychologists did significantly influence the types of testing first used in experimental work on individual differences.

Interest in the Mentally Deficient

In France, during the first half of the nineteenth century, interest in more accurate differentiation among individuals with regard to mental abilities was stimulated by a number of men, of whom two of the outstanding will be mentioned: Jean Esquirol (1772-1840) and Édouard Seguin (1812-80). They were concerned with mental deficiency and mental disease (14, 37).

Esquirol, in the first place, made explicit the distinction between mental deficiency and mental illness. These abnormal conditions were at that time generally undifferentiated and confused.³ He also distinguished among the several levels of mental deficiency. Esquirol spoke of the "weak-minded" and of several grades, or levels, of "idiocy"; the former term he applied to what, for many years now, has been called "morosity" (and probably also includes borderline cases), while the latter term refers to the current terms "imbecility" and "idiocy." These groups, however, were not precisely defined or delineated, although Esquirol did attempt unsuccessfully to distinguish and classify mentally deficient individuals on the basis of physical measurements, especially size and formation of the

² Psychophysics is the study of the relation between the physical attributes of the stimulus and the quantitative attributes of sensation.

³ Esquirol's distinction, essentially the one that has been current since then, is now widely understood: namely, that mental deficiency is a condition of seriously subnormal mental development due to congenital causes or to accidental causes occurring during early childhood, whereas mental illness (psychosis) is a severe disorder which may be marked by progressive impairment of mental functions and behavior, and by personality disintegration.

skull. It remained for Binet and his collaborator, Simon, to devise the first standard scale of intelligence and behavioral criteria that would differentiate the three levels of mental deficiency: *moron*, *imbecile*, and *idiot*.

Esquirol did, however, correctly discern the fact that development and use of language is one of the most useful and valid psychological criteria for differentiating levels of mental deficiency. This observation is of historical interest because for many years now the development, use, organization, and interpretation of verbal materials have been regarded by numerous psychologists as one of the major aspects—in some instances, *the* major aspect—of mental ability. Especially noteworthy among these psychologists is the late Lewis M. Terman, about whose work much more will be said in subsequent chapters.

Seguin is noteworthy for his pioneering work and methods in the training of mental defectives. He was placed in charge of a school for this purpose in 1842, after having had his own small school for the training of mental defectives for five years. Seguin believed that with appropriate help these individuals could improve in behavior, in utilization of their limited mental capacity, in their economic adequacy, and in their personalities generally. In 1846, his book on the treatment of mental defectives appeared (37). Like Esquirol, he attempted to find a basis for distinguishing between idiocy and imbecility, and between these and "backwardness." In 1848, Seguin migrated to the United States where, as in France, he stimulated interest in the study and training of mental defectives. His methods emphasized the development of greater sensory sensitivity and discrimination and of improved motor control and utilization.

Both Esquirol and Seguin are of significance to us because of their efforts to establish psychological criteria upon which to base differentiations among levels of mental deficiency; and, as will be seen later, it was this problem which provided the strongest original motive force to the testing movement after 1900. Seguin, furthermore, is noteworthy for his Form Board, which carries his name and is part of several performance test batteries currently in use.

Francis Galton's Contributions

It is clear from the foregoing brief account that until the last quarter of the nineteenth century there was scant recognition of individual differences as a subject worthy of study and research by psychologists. This indifference, no doubt, retarded the development of psychological tests that would be necessary for their measurement. Galton, though interested in and influenced by the psychological work of his predecessors

and contemporaries, was even more strongly influenced by the development of the biological sciences then ascendant among British scientists. Consequently, his efforts were devoted largely to investigations of individual differences more from biological interests than from psychological. In the introduction to his *Inquiries into Human Faculty* (1883), he states (18):

My general object has been to take note of the varied hereditary faculties of different men, and of the great differences in different families and races, to learn how far history may have shown the practicability of supplanting inefficient human stock by better strains, and to consider whether it might not be our duty to do so by such efforts as may be reasonable, thus exerting ourselves to further the ends of evolution more rapidly and with less distress than if events were left to their own course.

This quotation is evidence of Galton's sustained interest in developing a science of genetics and eugenics. It also indicates a problem with which psychologists have since been concerned—the roles of heredity and environment (or, as Galton named them, “nature and nurture”) in the development of man's intelligence. For the study of this problem, objective psychological tests have been indispensable.

Prior to the appearance of the volume mentioned above, Galton had published the results of his earlier studies in *Hereditary Genius* (1869), and *English Men of Science: Their Nature and Nurture* (1874). His *Inquiries into Human Faculty* was followed by *Natural Inheritance* (1889) and *Noteworthy Families* (1906), the last with Schuster. In addition to these larger works published during this period of about forty years, Galton produced numerous articles on the general subjects indicated by the titles of his books. At the same time, his statistical techniques for the analysis of data provided the basis for the elaborated, extended, and refined statistical methods used by such men as Karl Pearson, British biometrician, and Charles Spearman, British psychologist, who was one of the earliest and most noteworthy men to engage in the analysis of human abilities (38).

Galton not only stimulated investigations of individual differences; he also strongly influenced the direction of the experimental efforts to measure intelligence by means of tests of imagery and sensory discrimination. He devised a test for the measurement of the delicacy of weight discrimination; he invented what is now known as the Galton whistle for measuring sensitivity to high tones. In addition, he suggested devices for testing visual and auditory discriminations, reaction time, and muscular strength.

Galton assumed, apparently, that the simpler and measurable sensory capacities should be significantly correlated with intelligence. That this

was his hypothesis is shown by the fact that as subjects for study he selected persons of extreme differences in mental ability in order to learn whether their differences in sensory discrimination corresponded with the known differences in their mental abilities. Although it has long since been learned that sensory and sensory-motor tests have very little value for the study of the higher and more complex processes called intelligence, Galton's work, nevertheless, did strongly affect the course taken by test experimenters until about 1900, when the influence of Alfred Binet, the French psychologist, was felt.

Binet's Contributions

It is impossible in a short space to present a full review and evaluation of the character, range, and importance of Alfred Binet's contributions to individual psychology. An attempt will be made, however, to indicate his supreme importance in the field of mental measurements and individual differences.

Young (48) has quite properly said that "the contribution of Alfred Binet stands supreme for its general originality and the fact that he synthesized the growing movement into his now well-known scale." Binet and his collaborators objected to the types of psychological testing which followed Galton's work, on the ground that they were too simple in nature and would contribute little to the understanding of differences in the complex and higher mental processes; for it is in these higher processes that individual differences are most marked, and it is these which distinguish individuals most significantly and characteristically in daily activity; whereas it is in the simpler sensory and motor processes that persons differ least. Binet was quite ready to admit that the simpler processes lent themselves to more precise measurement and, therefore, yielded more nearly constant results. Yet his interests were strongest in individuals rather than in the study of sensations or ideas. Thus he was ready to sacrifice the greater quantitative precision of sensory-motor tests in order to obtain a more nearly accurate study of the integrated mentality of the individual. He argued that in the measurement of the higher functions, the greatest precision, though desirable, was not as essential as in measuring the simpler functions, because of the very fact that individual differences are more marked in the former. Binet made it clear, however, that his proposed scale would not measure in a physical sense, in the same way, for example, that a line is measured. It would, however, yield "a classification, a hierarchy among diverse intelligences; and for the necessities of practice this classification is equivalent to a measure" (2, p. 40). He and his collaborators were interested, consequently, in

establishing the extent and nature of variations of the mental processes from one individual to another, and in the determination of the inter-relations of the various processes *within* the individual. Binet and Henri (a collaborator) proposed, therefore, to study the following functions: memory, the nature of mental images, imagination, attention, comprehension, suggestibility, esthetic feeling or appreciation, moral sentiments, muscular strength and strength of will, motor skill, and visual judgment. These are, they believed, "faculties" which differ much from one individual to another and are such that knowledge of their state for an individual gives us a general idea of this person and permits us to distinguish him from other individuals within the same milieu. Here we have the beginnings of the tests which a few years later proved so useful in the construction of Binet's scales.

The range and number of publications coming from Binet and his collaborators were remarkable (45). They—especially Binet—interested themselves in and investigated an unusual variety of problems relevant to individual psychology including such matters as handwriting, head measurements, physical growth, physiognomy, and palmistry. Yet his abiding interest was in the problems of measuring intelligence and in differentiating between the mental level of one person and another.

The Binet-Simon Scale. In 1904 a practical situation arose in which Binet had an opportunity to apply his principles with regard to the differentiating of individuals and to make a great contribution to the study of individual differences in mental ability. The French Minister of Public Instruction appointed a commission to recommend means of educating subnormal children in the schools of Paris because these children were unable to profit from regular instruction. The plan, therefore, was to eliminate subnormal children from ordinary schools and to give them instruction in a special school. Admission was to be determined by a medical and a psychological examination. Obviously, the first device needed was an objective means of selecting those of subnormal mentality. Subjective opinions were worse than useless; for not only was there disagreement among different "experts," but serious injustices might result in some cases. It was to meet this problem that the first intelligence scale was constructed. This first one is known as the 1905 Binet-Simon scale (4, 5). In it we find the fundamental concept underlying all tests that measure the mentality of children. This principle is that we may identify differences in mental development—in degree of brightness or dullness—with differences in the levels of development as represented by the average capacities of children of various ages.

In the construction of their first scale, Binet and Simon limited themselves to the definite and practical problem of creating a device with

which school children's intellectual abilities might be measured, and with which the normal might be distinguished from the subnormal. They devoted their efforts to the evaluation and the quantitative determination of "general intelligence," that is, intellectual level, and to comparisons with normal children. They recognized that the determination of *special aptitudes* was a matter for later investigation. In fact, that very problem has been studied rather intensively within more recent years.⁴

Binet's first scale (1905), which he himself tested in Paris, was also tried out by other psychologists in Europe. As a result of these trials and consequent suggestions and criticisms, a second and considerably revised scale was constructed and appeared in 1908. Again, other psychologists collaborated by using this new scale in their own countries: Decroly and Degand in Belgium, Goddard in the United States, Bobertag in Germany, and Ferrari in Italy. Binet took account of their findings and criticisms. As a result of these and his own investigations, he published another revised scale in 1911. This was Binet's final contribution to the field of mental testing, for he died the same year.⁵

Binet, the synthesizer and the originator, provided the original major impetus to the study of individual differences by means of standardized tests. Since 1911, revisions and adaptations of his scale have been made in a number of countries. Most later developments have been expansions, modifications, and improved standardizations of the 1911 scale. Understandably, the principal interest for some years following Binet was in the identification and classification of mentally defective individuals.

Developments in the United States

Early Experiments. One of the most important of the early American psychologists in the study of individual differences was James McKeen Cattell (1860-1944), a man much younger than Galton, but still his contemporary. "It was Cattell," says Professor R. L. Thorndike, "[who] . . . was perhaps the first rebel from within the ranks of psychologists . . . to set his face against the narrowness of the Wundtian School where . . . individual diversities were hidden in averages, or even discarded as erroneous. . . . Cattell was bold enough to declare, in reference to reaction times, that . . . 'The individual difference is a matter of special interest.' Wundt opposed any study of individual differences in themselves" (48, p. 32).

The term "mental tests" was first employed by Cattell in a publication

⁴ Binet and Simon excluded from consideration those persons who had suffered mental disorganization; that is, the demented.

⁵ Binet's scales are examined in some detail in Chapter 8.

of 1890 in which he described tests then being used in his laboratory in the University of Pennsylvania (11). Cattell's tests were of memory, imagery, keenness of eyesight and of hearing, afterimages, color vision, color preferences, perception of pitch and of weight, perception of time intervals, sensitivity to pain, rate of perception and of movement, accuracy of hand movement, and reaction time.

The last of these was the most important of his early contributions to differential psychology; for much of the subsequent interest in reaction-time experiments is attributable to Cattell's work (20). One of the most direct methods with which certain of the simpler mental processes, such as discrimination and choice, can be studied is the precise measurement of the time an individual requires to respond to a given stimulus or to perform a specified act, usually a very simple one. Although many experiments on reaction time followed those of Cattell, and although these have added considerable information about speed of response to some types of stimuli, they have not made significant contributions to our understanding of higher, complex mental processes; for reaction time, it has been found, has little or no value in estimating intellectual abilities.

The time factor in itself is a relatively minor aspect of most mental tests, except in those devised specifically to measure speed of performance, usually in a restricted type of activity for a specified purpose. A good example is the rate at which one can discern likenesses and differences between two sets of digits or letters of the alphabet—a form of clerical test. However, Cattell justified his tests of sensory discrimination, motor activity, and simple reactions on the ground that his purpose at that time was principally anthropometric; therefore, measurement of the senses properly belonged within the scope of his research (12). He and his collaborators realized that the more complex mental processes should be measured; but they were also aware of the fact that much research and analysis had yet to be done before adequate mental tests could be devised for the measurement of these processes.

Other investigators in this country and abroad were experimenting with psychological tests, following very much the same paths at those of Galton and Cattell. Jastrow tried out tests of touch and cutaneous sensitivity, and tests of vision, memory, and reaction time (25). Gilbert used measures of height, weight, and lung capacity; also tests of sensation, rapidity of tapping, reaction time, memory, and suggestibility. Against these he compared teachers' ratings of their pupils' mental abilities (22).

The importance of studying individual differences by objective scientific methods and through comprehensive research was emphasized as early as 1895 when the American Psychological Association appointed a committee of which Cattell was a member ". . . to consider the feasibility

of cooperation among the various psychological laboratories in the collection of mental and physical statistics" (12). Also, in 1896 the American Association for the Advancement of Science appointed a committee ". . . to organize an ethnographic survey of the white races in the United States" (12). Cattell, who was also a member of this committee, stressed the importance of including psychological tests in the survey and of cooperating with the committee of the Psychological Association.

The development of testing had assumed importance to educators at an early date. In 1899, President Harper of the University of Chicago ". . . recommended that a special study be made of the college student's character, intellectual capacity, and tastes by the questionnaire method" (10). Further, in 1909, a committee of the National Education Association presented a report regarding psychological tests for mentally deficient children (8). From the report, it appears that the tests were looked upon as applicable chiefly to the subnormal and to other exceptional children.

Revisions of the Binet. The great development in testing and studying individual differences in the United States occurred after Binet's work was made known. Goddard was the first to revise the Binet scales for use in this country. In 1911, he published his standardization of Binet's 1908 revision, with which he had been acquainted since 1909 (23). At that time, he was director of the laboratory of psychology at the Vineland (New Jersey) Training School for Feeble-minded Children. Thus, as in France under the guidance of Binet, the scale in this country was first used almost entirely for the study and selection of mentally deficient individuals. The Binet was made a part of the routine procedure at Vineland, and it was rapidly adopted for use by psychologists in other institutions.

Goddard and Kuhlmann, who in 1911 published a revision of Binet's 1908 scale, made the test known and were largely responsible for its early spread among clinical psychologists (30). Lewis M. Terman, who had already interested himself in psychological differences among individuals, brought the scale before the schools of the country. In 1912, he published a tentative revision of the Binet; in 1915, he completed this revision with collaborators. In 1916, he published *The Measurement of Intelligence*, which presented the scale in its revised form, its standardization and directions for administering and scoring, as well as brief explanations of the psychological justification for each part (41).

In 1937, a revised and much improved edition in two forms was published in collaboration with Maude E. Merrill (42). Inevitably, of course, another revision of the scale had to be prepared. This last edition appeared in 1960, again under the coauthorship of Terman and Merrill, although Terman had died in December, 1956 (43). The 1916 and 1937

editions have been widely used in clinics, schools, and other agencies; and the 1960 edition, it is reasonable to assume, will also enjoy widespread currency.

Group Tests. Shortly after 1916, the most significant occurrence in psychological testing was the development of group tests. The Binet and its several revisions are administered to each person singly, the length of time required varying with the age, brightness, and responsiveness of the individual being tested. As a result, it is costly in time and money to test large numbers of persons one by one, and in some instances it is impossible to do so. Therefore, if many people are to be tested at once, as is the case in the schools and the armed forces, a group test will have distinct advantages if it yields sufficiently accurate and dependable results.

Psychologists had already begun to study, by group tests, some of the mental processes required in school work. So it was not a very long or entirely new step to try devising a single scale in which a variety of items, testing several mental processes, would be combined. This tendency in group testing received its greatest impetus in 1917 with the entrance of the United States into World War I. At that time the government agreed with the views of a group of psychologists that it would be desirable to examine the newly drafted men to determine their general mental capacity and vocational fitness by means of the best available psychological methods. The need was a pressing one, and a group-testing method was imperative. This army problem enlisted the interests and cooperation of many psychologists, some of whom had already made contributions to the field of measurements, and some of whom were already experimenting with group methods. Pooling their efforts and resources, they emerged with the well-known army tests, Alpha and Beta, the former being verbal in content and the latter nonverbal.

With their army data, these psychologists opened up numerous fields in which group tests might be used, and at the same time gave rise to a number of controversial questions. Among these were the relative influence of heredity and environment, racial and national differences, occupational and regional differences, and the age at which maximum mental capacity is reached.

In the ensuing years, a large volume of research on these and other problems was published. For present purposes, it is sufficient to note that this use of tests in the army and the results achieved demonstrated the possible values of group scales and supplied the impetus for their use in other areas, especially in the schools (7, 32).

There are today a large number of group tests designed for use at educational levels from kindergarten to university. Of these, some are highly

reliable (see Chapter 4) and have reasonably good validity (see Chapter 5), whereas others do not withstand scrutiny and evaluation.

Performance Tests. Not long after the introduction of the Stanford-Binet scale, its emphasis upon language was criticized by some psychologists and educators. It was maintained that this scale, valuable though it is, needed to be supplemented by tests which do not require ability to deal with words, numbers, and abstract concepts. Accordingly, "performance tests" were developed to meet this criticism and to provide means of testing individuals with language handicaps, as well as the deaf, the blind, and others for whom an adequate rating could not be obtained with tests that depended largely on language, numbers, and abstractions.

A performance test provides a perceptual situation in which the subject manipulates items such as form boards, blocks, pictures, and disassembled objects instead of reasoning with symbols. Some psychologists apply the term also to "pencil-and-paper" tests that utilize nonverbal materials such as printed geometric forms, pictorial representations, printed cubes, substituting digits for symbols, and the like. It seems preferable, however, to designate these simply as "nonverbal" tests because they do not involve actual manipulation of objects as do performance tests. Both types of test materials, performance and nonverbal pencil-and-paper, are now used extensively. Some scales, such as the Arthur and the Pintner-Paterson, are built entirely of performance materials; other scales combine one or both with verbal materials.

Aptitude Tests. Another type of instrument, the development of which received impetus in World War I, is the aptitude test. Each of these, unlike tests of general ability, is intended to measure an individual's ability to perform a task of a limited or specific kind, for example, clerical, mechanical, or musical aptitudes. Interest in and development of aptitude testing may be ascribed to several causes: the army's need, during World War I, to select men for tasks requiring specific skills; the desire, in vocational guidance and personnel assignment, to find the right person for a specific job; the opposition of some educators and psychologists to what they called the "super-faculty" of general intelligence; and the belief of some of them that only specific aptitudes, such as mechanical and clerical, could be satisfactorily measured.⁶ As a matter

⁶ Some psychologists prefer to avoid the use of the term "intelligence," and to speak, instead, of "general aptitude," "general ability," "scholastic aptitude," and the like. We shall continue to use the term "intelligence" because: (1) it has a long and respectable history in psychology; (2) many of the most important tests with which we shall be concerned are called tests of intelligence; (3) we shall have to deal with what psychologists have long called theories and definitions of intelligence; and (4) because there seems to be no merit in substituting the term "general ability" or "general aptitude" for "general intelligence." Furthermore, even those who would reject the term "intelligence" must and do use the concept of "intelligence quotient."

of fact, tests of general intelligence and those of specific aptitudes do not and need not stand opposed; they are supplemental.

Aptitude tests have been developed to predict educability and performance in music and drawing, in mechanical and clerical occupations, in engineering, in medicine and law, and in other areas as well. Others in this category are intended to evaluate aptitudes for the study of specific types of subject matter, such as science, foreign languages, and mathematics.

Occupational Interest Inventories. To supplement tests of aptitude and those of intelligence, several self-answering occupational preference questionnaires, or inventories, have been devised to provide information regarding an individual's interests in a variety of activities; for these, it has been found to have some relevance to and predictive value for certain broad vocational areas or for certain specific occupations.

Tests of Educational Achievement. Closely associated with the testing of aptitudes is the measurement of educational achievement and the construction of objective measures for that purpose. These are not designed primarily for prediction; instead, they are intended to measure the individual's actual learning in educational subject matter after a period of instruction. They have proved to be highly valuable in the determination of individual difficulties in learning, in the discovery of strong scholastic interests, in the discovery of special abilities or disabilities, and, in combination with other factors, in plotting the educational career of the individual child.

Educational achievement tests have other values as well: they provide objective measures of progress, as opposed to teachers' ratings that may be too subjective; they permit intergroup comparisons based on a reasonably objective determination; and they facilitate experimental evaluation of varied teaching methods.

Test Batteries. During World War II many test "batteries"⁷ were constructed. Those that made use of specific aptitudes and subject-matter knowledge—especially the former—were most important. Batteries were devised for the selection and training of personnel in a great variety of assignments in the several branches of the armed forces: radio and radar operators, pilots, navigators, gunners, flight engineers, and other specialties (1). The development of these batteries in the armed forces stimulated research on and use of similar tests for the selection and training of personnel in civilian occupations.

Multifactor Tests. These, also called "differential aptitude tests," are relatively recent developments in psychological measurement and evaluation. Interest in them has increased markedly since about 1945,

⁷ A "battery" of tests is a group of tests used in combination for a specified purpose.

although research on the subject began as early as the 1920's, when T. L. Kelley (28) and, later, L. L. Thurstone (44), published their work on *factorial analysis* of human abilities.⁸ Factorial analysis provided the statistical tools for the development of multifactor tests, which isolate and measure relatively "pure" mental operations (factors) or "constellations" of closely related factors, rather than general intelligence or general ability. In other words *multifactor tests isolate the elements that constitute mental operations*. The psychological principle upon which these instruments are based is the theory that the factors, or elements, are relatively independent of one another; hence, it was concluded that they should be measured independently.

Multifactor scales were expected to be especially valuable in educational and vocational counseling because they consist of separate tests of numerical operations, space relations, form perception, name perception, verbal reasoning, rote memory associations, and others restricted in complexity and range of mental operations. Each factor, or test, is thought to have special educational and vocational relevance and predictive value in itself; and a combination of factors is thought to have predictive value for specific areas of learning or occupations. The use of multifactor scales, therefore, would yield a "profile" ⁹ of scores for each of the several factors or "constellation" of factors, rather than a general, over-all rating for the entire scale, such as those derived from the Stanford-Binet, the Wechsler, and numerous group tests. All of these will be described and evaluated in subsequent chapters.

Personality Tests

Efforts to evaluate and test nonintellectual traits of personality were apparent in the nineteenth century beginning with Galton in 1879 (17) and followed by Pearson (35), who devised questionnaires and rating scales. During the last decade of that century and the first of the twentieth, word-association tests were tried out by Jung of Switzerland (26, 27) and Kent and Rosanoff in the United States (29) in an effort to expose some of the "deeper" personality traits and, if possible, to assist in differentiating among the various mental disorders. Although word-association tests are still used today in psychological clinics and elsewhere in diagnosing personality traits, they are much less frequently employed than inventories and projective techniques.

⁸ Spearman preceded both Kelley and Thurstone in making statistical analyses of human abilities; but he is not associated with the multifactor test movement.

⁹ A psychological "profile" is a chart representing an individual's score or relative position in each of several types of performance, with separate scores made comparable by statistical treatment.

With widespread use of individual tests of intelligence in schools, clinics, and hospitals, it became increasingly clear that in some cases an individual's performance on a test, his successes and failures, and the content and quality of his responses, were not only evidence of intellectual functioning, but were also affected, in greater or lesser degree, by non-intellectual traits of personality. The recognition of this fact, in addition to the growing interest in the scientific and clinical study of personality per se, provided the stimulus for the development of the several varieties of personality tests. Personnel problems during World War I provided impetus for their growth as well.

Today the tests are used extensively for the analysis of desirable and undesirable traits in a wide range of civilian and military occupations. In addition, psychologists employ personality tests in studies of differences between subgroups within the same general society and of differences between various cultural, national, and racial groups.

The large current crop of personality tests now available varies in quality from those that are poorly conceived, inadequately validated, and therefore useless, to those having considerable value in the hands of competent psychologists.¹⁰

Rating Scales. The earliest device employed, the rating scale, is a means of obtaining the judgments of a number of respondents with reference to a limited number of traits of a given individual. They were tried out and used during World War I, well before they were formalized and scaled both by statistical methods and by psychological analysis of personality and behavior traits relevant to specified situations.

Self-Rating Inventories. The first self-report, questionnaire type of personality inventory is the Personal Data Sheet, devised by R. S. Woodworth for use in World War I and published in 1919. Employed with moderate success, its purpose was to identify men who would prove to be poor prospects for military service because of undesirable personality and behavioral characteristics. This questionnaire consists of a list of items in the form of questions about himself, to be answered by the individual. The aim of the questionnaire is to detect personality and behavioral symptoms that are regarded as indicative of maladjustment. The questions on the Data Sheet took the place of an individual interview. Men whose responses indicated a sufficient number of undesirable symptoms were later interviewed individually. The types of questions asked and the aspects of personality sampled were forerunners of many of those included, with very little modification, in subsequent inventories.

¹⁰ Cf. O. K. Buros (9). In this volume, 145 personality tests of all types are critically reviewed.

Since the appearance of the Woodworth Personal Data Sheet, dozens of personality inventories, representing several different types, have been published. In general, the emphasis of the items—questions or statements—in these instruments is on what the individual respondent actually does in various kinds of situations and on how he feels about what he does in these situations. Relatively few of these inventories, however, have survived scientific analysis and practical use. Until the early 1930's these were, however, the principal instruments used to evaluate personality traits in a systematic and scientific, or quasi-scientific, manner (40).

Projective Tests. In the early 1930's a newer type of instrument became prominent in American psychology: the projective test of personality. This instrument is much more subtle than the self-rating inventory; it presents more or less equivocal, undefined ("unstructured") stimulus situations, usually in the form of pictures, inkblots, or incomplete sentences. Thus, the person being tested has a greater opportunity to impose upon the test his own private and particular personality traits than would be exposed by means of the questionnaire type of inventory.

The best known of the projectives is the Rorschach Inkblot Test, first published in Switzerland in 1921, although not introduced into the United States until the early 1930's. Rorschach, a Swiss psychiatrist, began his experimentation with inkblots as a means of stimulating and testing imagination. In the course of his work (1911-21), he perceived the possibilities inherent in the inkblot test as a device for differentiating among various kinds and traits of personalities. Although Rorschach's work on inkblots was the most extensive of any up to that time, he was not the first investigator to discern the possibilities of inkblots in psychological experimentation. As a matter of fact, these had been used for some years in psychological laboratories to study fertility of imagination and of invention¹¹ (46, part II). Since the introduction of what has come to be known as "the Rorschach," it has been extensively used in private psychological practice, in clinics, and in hospitals for diagnostic purposes; in business and industry for some types of personnel selection; in researches in cultural anthropology; and in researches on personality theory. Interest in and use of the Rorschach can be inferred from the huge number of professional publications on the subject, which did not begin to appear in appreciable numbers until about 1935.¹²

Another projective instrument of major importance is the Thematic Apperception Test, introduced by H. A. Murray and C. D. Morgan in

¹¹ Among those who early suggested the use of inkblots were Binet and Henri, in 1895.

¹² In O. K. Buros (9), 2297 publications are listed.

1935. This test consists of thirty rather ambiguous pictures, each on a separate card, and one blank card. The person being examined is asked to make up a story of his own for each picture. The psychological principle involved is that in his stories the examinee will, probably unwittingly, give expression to his needs, values, attitudes, and feelings about persons, situations, and the world around him, as well as to the pressures he is experiencing from sources outside of himself. This instrument, too, has been and is being widely used in a variety of psychological settings. While the number of publications on the TAT, as it is professionally known, is not so great as that on the Rorschach, it has, nevertheless, been the subject of many studies and researches.¹³

Since the appearance of the Rorschach and the TAT, a variety of other projective devices and techniques have been made available. Some of these are special adaptations of the two foregoing tests; others offer rather different approaches for the same general purpose, that is, to elicit responses which will reveal aspects and traits of personality that inventories and rating scales are incapable of eliciting. Since 1945 and to the present time, projective tests have occupied a position of primary importance in practical applications and in research.

The types of techniques for obtaining evaluations of aspects of personality thus far mentioned do not exhaust the list. Among other and more tenuous kinds of procedures used are storytelling and story completion, drawing and painting, and "situational tests," in which an individual's behavior is observed and rated in a setting that simulates reality (34). Contrived play activities, usually of one child who is being observed, are used for two purposes: to permit the child to project some of his inner traits and to serve as a form of psychotherapy. Sociometric methods, whereby an individual's social currency or acceptability is obtained from ratings made by his peers, is an adaptation and extension of the older rating scale (33).

Although all of these procedures are used in their appropriate settings, they are much less commonly employed in personality evaluations than are self-rating inventories, the Rorschach, and the TAT, because, being tenuous, they are not susceptible to standardization and objectification. To be sure, personality inventories and the more widely used projective tests present their own problems in standardization. However, progress has been and continues to be made with these and their development has proceeded far enough to provide sufficient common ground and research information, so that in the hands of qualified psychologists they are of value.

¹³In O. K. Buros, *op. cit.*, 610 publications are reported.

THE PRESENT SITUATION

The Present Situation

Psychological tests of intelligence, whether based upon the theory of "general ability" or upon one of relatively independent factors (or aptitudes), and tests of specific aptitudes and skills are now at a reasonably advanced stage of development. This is so because they have been in the process of evolution and improvement for many years, a tremendous amount of research has been devoted to them by numerous psychologists, and they have been used in a variety of practical situations where their validity could be evaluated. Another reason is the fact that determination of the mental functions, or operations to be tested, though not simple, has not been as difficult as the determination by testing of nonintellective traits of personality.

Because "personality" is so all-inclusive a concept, and because its manifestations are often complex and covert, development and use of self-rating inventories and projective tests are as yet not on so secure a foundation as tests of mental abilities, of specific aptitudes and skills, and of educational achievement. In subsequent chapters, we shall discuss the principles upon which all of these types of tests are based, as well as their values and their limitations.

The great variety of psychological tests in existence has already been mentioned. The numerous uses to which they are put and the important part they may have in the determination of an individual's educational, vocational, or general welfare have been indicated. It is essential, therefore, that anyone who employs these tests in a professional capacity should understand the basic psychological and statistical principles upon which they rest. It is necessary that everyone—teachers, psychiatrists, guidance counselors, personnel administrators—who interprets the results of test findings should be familiar with their essential theory as well as with the meanings of the technical terms.

Since the end of World War I, the use of psychological tests has continuously increased, because they are needed and because they have improved steadily. Education in the United States has become more nearly universal; individuals of inferior and those of seriously deficient mental abilities are being retained in public schools much longer than was the case in earlier years. Thus, the range of intelligence found in schools extends from the very low to the highest levels, making it essential that each individual's educational potential and promise be known as accurately as available psychological means permit. The general increase in years of schooling, not to speak of the tremendous growth in numbers of students, has extended to college and university, so that the importance

of knowledge of individual variations in mental ability at higher educational levels has also grown.

Educational and vocational guidance, at all levels, have consequently assumed increasing significance. With the availability of standardized tests, even with their defects, guidance has been placed upon a more objective basis, instead of remaining a matter of subjective, perhaps even casual, advice.

For many years schools and, now more recently, colleges and universities have been concerned with the learning difficulties of individuals. Are these difficulties due to inferior general intelligence? Or are they due to specific disabilities, as in reading or spelling? Or to defective perception of spatial relations? Or perhaps to defects of the visual-motor function? Is an individual's lack of aptitude in shopwork attributable to inferior manual dexterity? Is the individual's learning impaired or retarded by poor ability for recall of rote or meaningful materials although his level of general ability might otherwise be adequate for learning? Answers to these and other important educational problems have been provided or at least facilitated by the use of psychological tests.¹⁴

The types and numbers of occupations have multiplied, and specializations within the types themselves have increased. It is unnecessary to detail the vocational changes and developments that have taken place with technological and scientific developments, but it does seem necessary to point out that for purposes of psychological testing and vocational guidance, occupations designated by the same name are not necessarily identical in regard to skills, knowledge, specialized functions, and interests involved. For example, there are various factors that combine in different ways to create not a single, unitary aptitude called "mechanical"; but, rather, there are several different aspects of mechanical aptitude, although all have something in common. "Engineering aptitude" is not a single, unitary function either. There are differences in requirements for learning and achieving proficiency in civil, mechanical, electrical, and chemical engineering, although, of course, their requirements are not mutually exclusive. Nor is "clerical aptitude" a single, unitary function. The fact that each of these general areas of training and employment is complex and divisible gives increased significance to psychological testing and insightful guidance.

Tests of personality are being used in some business and industrial organizations in the selection of management personnel, whereas in certain professions, tests are utilized in selecting individuals to be educated for

¹⁴ We are assuming throughout, of course, that tests are administered and interpreted by qualified professional persons. More will be said on this matter in subsequent chapters.

practice in them. These professions include medicine, in which there have been researches on desirable personality traits of medical students. Some engineering schools would like to identify those nonintellectual traits that distinguish the successful from the unsuccessful students of the profession. Psychologists are desirous of determining personality characteristics of the more promising students of clinical psychology. Some religious denominations require that candidates for admission to their theological schools take tests of nonintellectual personality traits as well as of mental ability.

Finally, there is the whole area of "mental health," to which so much attention has been given since the termination of World War II. Schools and colleges are concerned over individuals who present more than ordinary degrees of personality difficulties or of problems of behavior. Numerous bureaus of child guidance have been established within school systems; there are mental health clinics in many sections of the country; federal hospitals (for example, of the Veterans' Administration) have psychological divisions, as do many state and some private hospitals. In all of these settings, psychological testing of all types, especially involving nonintellectual personality traits, is one of the established practices. And it is not uncommon for private welfare agencies to have on their staffs psychologists whose work consists of psychological diagnosis by means of tests, or of the practice of psychotherapy, which is often based upon or facilitated by diagnostic testing, or of both. Also, many psychologists in private practice make diagnostic testing a significant or a major part of their work.

This brief account of the current role and extent of psychological testing should be sufficient to emphasize the development of this branch of psychology since its relatively modest beginnings, shortly after the turn of the twentieth century, when the principal purpose of testing was the identification and special schooling of mentally deficient children.

References

1. *Aviation Psychology Program in the Army Air Forces*. Reports numbers 1-19. Washington, D.C.: U.S. Government Printing Office, 1947-48.
2. Binet, A. *The Development of Intelligence in Children* (translated by E. S. Kite). Vineland, N.J.: Training School, 1916.
3. Binet, A., and V. Henri. La psychologie individuelle. *L'Année Psychologique*, 1896, 2, 411-465.
4. Binet, A., and T. Simon. Sur la nécessité d'établir un diagnostic scientifique des états inférieurs de l'intelligence. *L'Année Psychologique*, 1905, 11, 163-190.

5. Binet, A., and T. Simon. Methodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Année Psychologique*, 1903, 11, 191-244.
6. Boring, E. G. *History of Experimental Psychology* (2nd ed.). New York: Appleton-Century-Crofts, Inc., 1950.
7. Brigham, C. C. *A Study of American Intelligence*. Princeton: Princeton University Press, 1923.
8. Bruner, F. G., E. Barnes and W. F. Dearborn. Report of the committee on books and tests pertaining to the study of exceptional and mentally deficient children. *Proceedings of the National Education Association*, 1909, 7, 901-914.
9. Buros, O. K. (ed.). *Fifth Mental Measurements Yearbook*. Highland Park, N.J.: The Gryphon Press, 1959.
10. Carothers, F. E. Psychological examinations of college students. *Archives of Psychology*, 1921, no. 46.
11. Cattell, J. McK. Mental tests and measurements. *Mind*, 1890, 15, 373-381.
12. Cattell, J. McK., and L. Farrand. Physical and mental measurements of students of Columbia University. *Psychological Review*, 1896, 3, 618-648.
13. Ebbinghaus, H. Über eine neue Methode zur Prüfung Geistiger Fähigkeiten und ihre Anwendung bei Schulkindern. *Zeitschrift für Psychologie*, 1897, 13, 401-459.
14. Esquirol, J. E. D. *Des maladies mentales considérées sous les rapports médical, hygienique, et médico-legal*. Paris: J. B. Bailliere, 1838. 2 vols.
15. Galton, F. *Hereditary Genius*. London: Macmillan & Co., 1869.
16. Galton, F. *English Men of Science: Their Nature and Nurture*. London: Macmillan & Co., 1874.
17. Galton, F. Psychometric experiments. *Brain*, 1879, 2, 149-162.
18. Galton, F. *Inquiries into Human Faculty and Its Development*. London: Macmillan & Co., 1883.
19. Galton, F. *Natural Inheritance*. London: Macmillan & Co., 1889.
20. Garrett, H. E. *Great Experiments in Psychology*. New York: Appleton-Century-Crofts, Inc., 1951.
21. Gilbert, J. A. Researches on the mental and physical development of children. *Studies from the Yale Psychological Laboratory*, 1894, 2, 40-100.
22. Gilbert, J. A. Researches upon school children and college students. *University of Iowa Studies in Psychology*, 1897, 1, 1-39.
23. Goddard, H. H. A revision of the Binet scale. Vineland, N.J.: *The Training School Bulletin*, 1911, 8, 56-62.
24. Huey, E. B. The Binet Scale for measuring intelligence and retardation. *Journal of Educational Psychology*, 1910, 1, 435-444.
25. Jastrow, J. Some anthropometric and psychologic tests on college students: A preliminary survey. *American Journal of Psychology*, 1891-1892, 4, 420-428.
26. Jung, C. G. The association method. *American Journal of Psychology*, 1910 21, 219-269.
27. Jung, C. G. *Studies in Word Association* (translated by M. D. Eder). London: William Heinemann, Ltd., 1918.

28. Kelley, T. L. *Crossroads in the Mind of Man*. Stanford, Calif.: Stanford University Press, 1928.
29. Kent, G. H., and A. J. Rosanoff. A study of association in insanity. *American Journal of Insanity*, 1910, 67, 37-96, 317-390.
30. Kohs, S. C. Annotated bibliography of the Binet-Simon scale. *Journal of Educational Psychology*, 1914, 5, 215-224, 279-290, 335-346.
31. Murphy, G. *An Historical Introduction to Modern Psychology*. New York: Harcourt, Brace & World, Inc., 1929.
32. National Academy of Sciences. *Memoirs*. Washington, D.C.: U.S. Government Printing Office, 1921, 15.
33. Northway, M. L. A method for depicting social relationships by sociometric testing. *Sociometry*, 1940, 3, 144-150.
34. Office of the Strategic Services Staff. *Assessment of Men*. New York: Holt, Rinehart and Winston, Inc., 1948.
35. Pearson, K. On the laws of inheritance in man. *Biometrika*, 1904, 3, 131-190.
36. Peterson, J. *Early Conceptions and Tests of Intelligence*. New York: Harcourt, Brace & World, Inc., 1926.
37. Seguin, E. *Idiocy: Its Treatment by the Physiological Method*. New York: Bureau of Publications, Teachers College, Columbia University, 1907.
38. Spearman, C. General intelligence objectively determined and measured. *American Journal of Psychology*, 1904, 15, 201-292.
39. Stern, W. *Über Psychologie der individuellen Differenzen*. Leipzig: Barth, 1900.
40. Symonds, P. M. *Diagnosing Personality and Conduct*. New York: Appleton-Century-Crofts, Inc., 1931.
41. Terman, L. M. *The Measurement of Intelligence*. Boston: Houghton Mifflin Company, 1916.
42. Terman, L. M., and M. E. Merrill. *Measuring Intelligence*. Boston: Houghton Mifflin Company, 1937.
43. Terman, L. M., and M. E. Merrill. *Stanford-Binet Intelligence Scale*. Boston: Houghton Mifflin Company, 1960.
44. Thurstone, L. L. *The Vectors of the Mind*. Chicago: University of Chicago Press, 1935.
45. Varon, E. J. The development of Alfred Binet's psychology. *Psychological Monographs*, 1935, 46, no. 3.
46. Whipple, G. M. *Manual of Mental and Physical Tests*. Baltimore: Warwick and York Incorporated; 1914, part I, Simpler Processes; 1915, part II, Complex Processes.
47. Yoakum, C. S., and R. M. Yerkes. *Army Mental Tests*. New York: Holt Rinehart and Winston, Inc., 1920.
48. Young, K. The history of mental testing. *Pedagogical Seminary*, 1924, 31-48.