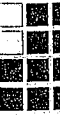


History of Testing and Assessment

CHAPTER

1



You walk into the room, prepared to take the test and know that the results will impact your future. You can feel your heart begin to pound and your stomach being to churn. "OMG, I hope I can do well, you say to yourself."

With millions of children and adults frightened by the thought of taking a test, this is not a pretty picture. But is there value in this sometimes-terrifying experience? We'll let you answer that question after you have finished reading this book. But how did test-taking start? That question will be answered in this chapter.

(Ed Neukrug)

In this chapter we will examine the history of testing and assessment. First, we will explore the differences between testing and assessment and point out how their current definitions are directly related to their history. We will then take a ride through the history of assessment, starting with ancient history and working our way to the development of modern-day assessment instruments. Along the way, we will highlight some of the people who were pioneers in the development of assessment measures and discuss some of the controversial issues that arose. As the chapter nears its conclusion, we will examine the current categories of assessment instruments, and we will finish by raising a number of ongoing concerns surrounding the use of assessment instruments.

DISTINGUISHING BETWEEN TESTING AND ASSESSMENT

Assessment
A broad array of evaluative procedures

Today, the term *assessment* includes a broad array of evaluative procedures that yield information about a person (Hunsley, 2002). Assessment procedures include the clinical interview; informal assessment techniques such as observation, rating scales, classification methods, environmental assessment, records and personal documents, and performance-based assessment; personality tests such as objective tests, projective tests, and interest inventories; and ability tests such as achievement tests and aptitude tests (see Figure 1.1).

Tests
Instruments that yield scores based on collected data—a subset of assessment

Tests are a subset of assessment techniques that yield scores based on the gathering of collective data (e.g., finding the sum of correct items on a multiple-choice exam). Assessment procedures can be formal, which means they have been well-researched and shown to be scientifically sound, valid, and reliable, or informal, which implies that such rigor has not been demonstrated, although the procedure might still yield some valuable information.

Generally, the greater the number of procedures used in assessing an individual, the greater the likelihood that they will yield a clearer snapshot of the client. Thus, using multiple assessment procedures, or a holistic approach to assessment,

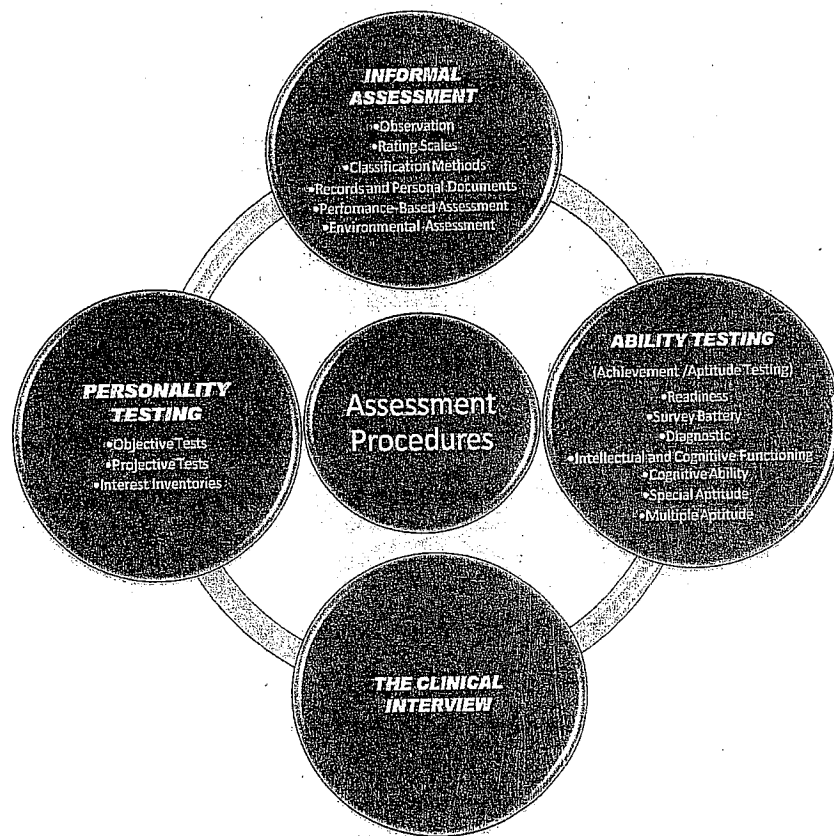


FIGURE 1.1 | Assessment Procedures

Multiple assessment procedures should always be considered

should always be considered when making important decisions about a client's life (Association for Assessment and Research in Counseling, 2012; Joint Committee on Testing Practices, 2004). In this text we will examine a broad array of formal and informal assessment procedures, all of which can be used in the decision-making process. But let's start at the beginning and see how events of the past have moved us toward our current use of assessment instruments.

THE HISTORY OF ASSESSMENT

Although the modern era of assessment began near the beginning of the twentieth century, assessment procedures can be found in ancient times. Let's examine some of the changes in assessment that have taken place over the centuries.

Ancient History

He said, "Take your son, your only son Isaac, whom you love, and go to the land of Moriah, and offer him there as a burnt offering on one of the mountains that I shall show you" (Genesis 22:5, New Revised Standard Version).

Assessment has been around for as long as humans have walked the earth. In fact, one might say that Abraham's loyalty was assessed when God asked him to kill his son Isaac. From a more down-to-earth perspective, the Chinese government is given credit for developing one of the first widely used tests when it began to assess individuals for fitness to work in government positions in approximately 2200 B.C.E. (DuBois, 1970; Higgins & Sun, 2002). With testing done under grueling conditions in hundreds of small cubicles or huts, it was not unusual that examinees would die from exhaustion (Cohen, Swerdlik, & Sturman, 2012). This kind of testing was not abolished until 1905. In the Western world, passages from Plato's (428–327 B.C.E.) writings indicate the Greeks assessed both the intellectual and physical ability of men when screening for state service (Doyle, 1974).

Precursors to Modern-Day Test Development

As experimental and controlled research spread throughout the scientific community during the 1800s, physicians and philosophers began to apply these research principles to the understanding of people, particularly in the area of cognitive functioning. For instance, working in mental asylums, the French physician Jean Esquirol (1772–1840) examined how language ability of individuals with intellectual disabilities was related to intelligence (Zusne, 1984; Drummond, 2009). Seen as having a condition called "idiocy," (Esquirol, 1838, p. 38), these individuals were viewed as having intellectual deficits as compared to a "normal" person reared in a similar environment. Esquirol's focus on language ability is often seen as the beginning of what later became known as the assessment of "verbal IQ." At around the same time, Edouard Seguin (1812–1880), also from France, suggested that the prognosis regarding intellectual deficits in children was worse if such deficits were associated with physiological problems. He suggested that physicians should "watch for a swinging walk, 'automatically busy' hands, saliva dripping from a 'meaningless mouth,' a 'lustrous and empty' look, and 'limited' or

Jean Esquirol
Used language to identify intelligence—
forerunner of
"verbal IQ"

Edouard Seguin
Developed the form board to increase motor control—
forerunner of
"performance IQ"

‘repetitive’ speech” (Zenderland, 1987, p. 54). Eventually, Seguin developed the form board to increase his patients’ motor control and sensory discrimination and to compare children and individuals with severe intellectual deficits to average children at different age groups (see Figure 1.2). Considered by some to be the forerunner of “performance IQ” measures (DuBois, 1970), versions of the form board, which is similar to the toy in which children place shapes in their respective grooves, are still used today in some performance-oriented IQ test.

Sir Francis Galton
Examined relationship
of sensory motor
responses to
intelligence

Meanwhile, intrigued by Charles Darwin’s (1809–1882) theory of evolution, scientists during the mid-1800s became engrossed in trying to understand the development of the human species (Juve, 2008; Kerr, 2008). For instance, Sir Francis Galton (1822–1911), Darwin’s cousin, became fascinated by differences among people and eventually came to believe that people inherited physical and mental characteristics (Gillham, 2001; Murdoch, 2007). He hypothesized that some inherited physical traits, such as reaction time and stronger grip strength, might be related to superior intellectual ability. His curiosity led him to examine the relationship among such characteristics, and his research spurred others to develop the statistical concept of the *correlation coefficient*, which describes the strength of the relationship among variables (DuBois, 1970; Kerr, 2008). Calculating the correlation coefficient has become an important tool in the development and refinement of tests.

Wilhelm Wundt
Developed one of the
first psychological
laboratories

Wilhelm Wundt (1832–1920), another scientist intrigued with human nature, set out to create “a new domain of science” that he called physiological psychology. Around 1875, at the University of Leipzig in Germany, Wundt developed one of the first psychological laboratories that used experimental research (Nicolas, Gyselinck, Murray, & Bandomir, 2002). Many of the experiments in Wundt’s laboratory studied reaction time of hearing, sight, and other senses in response to stimuli (Watson, 1968). A number of students who worked with Wundt helped to

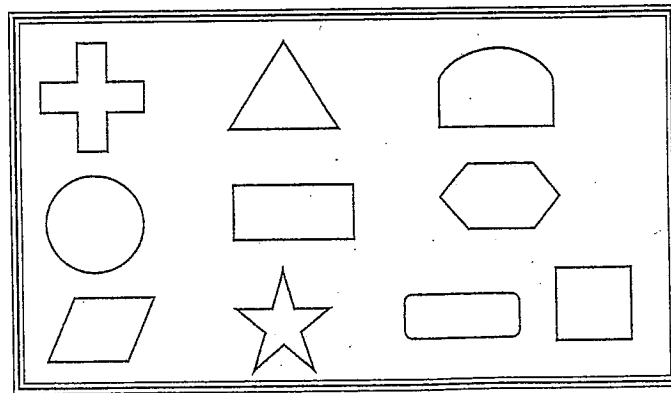


FIGURE 1.2 | Reproduction of Seguin’s Form Board

Task: Children, or individuals with intellectual disabilities, would be given ten blocks, in three piles, and asked to place them in the slots as fast as they can. They would then determine intellectual age by finding which age group the individual was most similar to.

James McKeen Cattell
Brought statistics to mental testing—
coined term mental test

G. S. Hall
Early experimental psychologist. First president of APA

foster in the new age of psychological science. For instance, **James McKeen Cattell** (1860–1944), a doctoral student under Wundt who was later greatly inspired by Galton, became one of the earliest American psychologists to use statistical concepts in understanding the person (Goodwin, 2008; Roback, 1961). Cattell's main emphasis became the assessment of what he termed *mental tests* and included examining individual differences of such things as memory span and reaction time. Another important figure, G. S. Hall (1844–1924), also worked with Wundt and eventually set up his own experimental lab at Johns Hopkins University. Hall became a mentor to other great American psychologists and was the founder and first president of the American Psychological Association in 1892 (Benjamin, 2008).

The Emergence of Ability Tests (Testing in the Cognitive Domain)

Influenced by the new scientific approach to understanding human nature, researchers at the beginning of the twentieth century began to develop instruments that could scientifically measure an individual's abilities. This era saw the emergence of ability tests, including individual intelligence tests, neuropsychological assessments, and group tests of ability.

Alfred Binet
Created first modern intelligence test

Intellectual and Cognitive Functioning: Intelligence Testing and Neuropsychological Assessment Although commonplace today, the first intelligence tests were developed by **Alfred Binet** (1857–1911) who, in 1904, was commissioned by the Ministry of Public Education in Paris, to construct a test that could be of assistance in integrating what they called “subnormal” children into the schools (Binet & Simon, 1916). Highly critical of the manner in which “mental deficiency” was diagnosed in children, Binet and his colleague Theophile Simon developed a scale that could be administered one-on-one and which would measure higher mental processes by assessing responses to a variety of different kinds of tasks (e.g., tracking a light, asking the individual to distinguish between different types of words) (Ryan, 2008). The information gained from their observations was then used to develop the first modern-day intelligence test (Watson, 1968). A relatively short time later, **Lewis Terman** (1877–1956), from Stanford University, began analyzing and methodically gathering extensive normative data on Binet and Simon's scale from hundreds of children in the Stanford, CA area (Jolly, 2008; Kerr, 2008). Based on the data, Terman made a number of revisions to the Binet and Simon scale. Originally called the Stanford Revision of the Binet and Simon scale, the test later became known as the Stanford-Binet, the name by which the revised version continues to be known as today. Terman was the first to use the term *intelligence quotient*, or “IQ,” which used a ratio of mental age to chronological age (see Box 1.1).

Lewis Terman
Enhanced Binet's work to create Stanford-Binet intelligence test

Intelligence quotient
Mental age divided by chronological age



Intelligence tests are sometimes used in neuropsychological assessment, which examines changes in brain function as the result of injury or disease process. Interest in how the brain impacts cognitive and behavioral functions, however, can be traced back to early Egypt where observations of behavioral changes following head injuries are recorded in 5,000-year-old Egyptian medical documents (Hebben & Milberg, 2009). In modern times, research conducted during World War I examined

BOX 1.1**Developing the Notion of "IQ"**

Lewis Terman wanted to develop a logical and relatively easy way of expressing an individual's intelligence. Using the data from his research, he quickly realized that he could compute a ratio score for each child by dividing a child's mental score (the age score at which the child performed) by the child's actual age. Thus, if a child was performing at the level of the average 12-year-old but was actually 9 years old, the ratio would be $12/9$ or 1.33 . Multiplying this number by 100 to eliminate the decimal point would yield an intelligence quotient ("IQ") of 133.

Note: IQ is no longer determined in this manner, and the current method of calculation will be discussed later in the text.

Use this method to determine the IQs of the children below, based on their mental age scores and their actual ages.

Child 1: mental age of 6 and chronological age of 8

Child 2: mental age of 16 and chronological age of 16

Child 3: mental age of 10 and chronological age of 9

Answers: Child 1: 75 ($6/8 \times 100$), Child 2: 100 ($16/16 \times 100$), Child 3: 111 ($10/9 \times 100$)

behaviors of soldiers who suffered from brain injuries. At that time, with the testing movement in full swing, it is not surprising that new screening and diagnostic measures to assess behavioral changes due to brain trauma were created (Lezak, Howieson, Bigler, & Tranel, 2012). During the twentieth century, as individuals became more interested in the nature of the brain and spurred on by the use of X-rays and other forms of brain imaging, the field of neuropsychology, or the study of brain function as it relates to behavior, was established. Today, when suspected changes occur in brain function due to disease, accidents, or violence, a "neuropsych" assessment, which sometimes includes an intelligence test, is often used.

Robert Yerkes
Chairman of the committee that developed the Army Alpha

Army Alpha
First modern group test—used during WWI

Group Tests of Ability (Group Testing in the Cognitive Domain) Realizing the importance of obtaining accurate information from examinees, early test developers, such as Terman and others, devised standardized directions to use in testing and stressed the importance of having trained examiners administer tests individually (Geisinger, 1994; Jolly, 2008). However, it was soon evident that individual testing, such as that conducted when doing intelligence testing, often took a particularly long time and was costly. During World War I, these practical concerns came to a head as it became critical to quickly administer tests of cognitive ability in order to place large groups of recruits in the military. At that time, Robert Yerkes, the president of the American Psychological Association, chaired a special committee to create a screening test for these new recruits. The committee, composed of many well-known psychologists, including Terman, prepared a draft of the test in just four months (Geisinger, 2000; Jones, 2007). The original test the committee developed was known as the Army Alpha (see Box 1.2 and Illustration 1.1).

Although the Army Alpha clearly had its problems, it was a large step toward the mass use of tests in decision-making and was administered to more than 1.7 million recruits in less than two years (Haney, 1981). Since there were many foreign-born recruits and large numbers of people who could not read, a second

BOX 1.2 The Army Alpha Test

The Army Alpha test was created to place recruits in the military (Jones, 2007; McKean, 1985). Based on this test, it was found that average mental age of the recruit was 13. Take the test below in the 3-minute time allotment given, then consider potential issues of bias and cultural fairness of the questions on the test.

The Army Alpha was used to determine placement in the armed forces during WWI. Below is an adaptation of the test, as printed in *Discover* magazine. Take the test and discuss your thoughts about it.

The average mental age of the recruits who took the Army Alpha test during WWI was approximately 13. Could you do better? You have three minutes to complete these sample questions, drawn verbatim from the original exam. (McKean, 1985)

The following sentences have been disarranged but can be unscrambled to make sense. Rearrange them and then answer whether each is true or false.

1. Bible earth the says inherit the the shall meek. true false
2. a battle in racket very tennis useful is true false

Answer the following questions.

3. If a train goes 200 yards in a sixth of a minute, how many feet does it go in a fifth of a second?
4. A U-boat makes 8 miles an hour under water and 15 miles on the surface. How long will it take to cross a 100-mile channel if it has to go two-fifths of the way under water?
5. The spark plug in a gas engine is found in the crank case manifold cylinder carburetor
6. The Brooklyn Nationals are called the Giants Orioles Superbas Indians
7. The product advertised as 99.44 per cent pure is
Arm & Hammer Baking Soda Crisco Ivory Soap Toledo
8. The Pierce-Arrow is made in Flint Buffalo Detroit Toledo
9. The number of Zulu legs is two four six eight

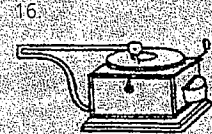
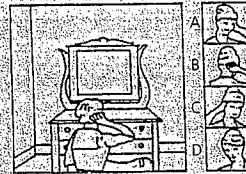
Are the following words the same or opposite in meaning?

10. vesper-matin same opposite
11. aphorism-maxim same opposite

Find the next number in the series.

12. 74, 71, 65, 56, 44, Answer
13. 3, 6, 8, 16, 18, Answer

14. Select the image that belongs in the mirror. 15. & 16. What's missing in these pictures?



Answers: 1. true, 2. false, 3. twelve feet, 4. nine hours, 5. cylinder, 6. superbas, 7. Ivory Soap, 8. Buffalo, 9. two, 10. opposite, 11. same, 12. 29, 13. 36, 14. A, 15. spoon, 16. gramophone horn

Scoring: All items except 3, 4, 10, and 11 = 1.25 points; items 3 and 4 = 1.875 points; items 10 & 11 = .625 points. Add them all up, they equal your mental age. What is wrong with this test? Examine it for problems with content, history, cross-cultural contamination, and so forth.



ILLUSTRATION 1.1 | Recruits Taking an Examination at Camp Lee, 1917

Source: U.S. Signal Corps photo number 11-SC-386 in the National Archives.

language-free version of the test, known as the Army Beta, was created. The Army Beta applied the use of form boards and mazes, and directions were given by pantomime so recruits could take the entire test without reading (Goodwin, 2008). Although crude by today's standards, the Army Alpha and Army Beta ushered in the era of group tests of ability.

In contrast with neuropsychological assessments and individual intelligence tests, which are given one-to-one, group tests of cognitive ability tend to be multiple choice and true-false measures given to groups of individuals simultaneously in an effort to assess the academic promise of each individual in the group. Probably the most well known of these has been the Scholastic Aptitude Test (now the SAT Reasoning Test, or SAT). Developed by the Educational Testing Service after World War II, the test in many ways was the brain child of James Bryant Conant, president of Harvard. Believing in a democratic, classless society, Conant thought that such tests could identify the ability of individuals and ultimately help to equalize educational opportunities (Frontline, 1999). Unfortunately, many have argued that instead of fostering equality, the SATs have been used to separate the social classes, and many in the testing movement were not as magnanimous as James Bryant Conant (see Box 1.3).

James Bryant Conant
Developed SAT to equalize educational opportunities

BOX 1.3 Eugenics and the Testing Movement

As scientists tried to make sense of the relatively new theory of evolution and as they began to examine differences among people, a number of them began attributing these differences to genetics. This information was used to bolster support for the emerging Eugenics Movement, whose members espoused a belief in selective breeding in order to improve the human race.

Individuals such as Galton, Terman, and Yerkes believed that the data retrieved from tests could help distinguish those who were naturally bright from those who, they argued, were less fortunate. Test results were used to advocate for providing incentives for the upper class to procreate and in finding methods to prevent the lower classes from having children (Gillham, 2001). Based on flimsy evidence and misguided thinking, this movement is also seen today as having racist undertones.

Believing that the Army Alpha and Army Beta measured innate ability, Terman, Yerkes, and others

used the results of these tests to support the Eugenics Movement. However, the tests were a far cry from being a measurement of intelligence, as they were saturated with cultural bias and were largely based on measuring achievement, or what had been learned, as opposed to some kind of raw, inherent intelligence. Unfortunately, their beliefs about the test and what should be done as a result of the test data may have been one of many influences that led the U.S. government to manipulate whom it would allow to immigrate to the United States (Sokal, 1987). As a result, thousands—perhaps millions—of individuals were unable to emigrate from tyrannical governments in Europe and other parts of the world (Gould, 1996).

Question to Ponder: Do you prefer partnering with someone who is bright? And if so, are you practicing your own, individualized eugenics?

Edward Thorndike
Developer of the
Stanford Achievement
Test

Frank Parsons
Leader in vocational
counseling

Paralleling the rise of group tests of cognitive ability was the administration of group tests of achievement in the schools. Traditionally, such tests had been given orally, and later in essay fashion, but the practicality of administering objective tests of academic performance to large groups of students was obvious. With the new scientific approach to testing on the rise, Edward Thorndike (1874–1949), one of the pioneers of modern-day educational and psychological assessment, and others, thought that such tests could be given in a format that was more reliable than previous tests. This move toward group testing culminated with the development of the Stanford Achievement Test in 1923 (Armstrong, 2006). Today, these tests are commonplace and are given to students en masse in school systems throughout America.

One last area where group testing became popular was in vocational counseling. With Frank Parsons (1909/1989) at its helm, vocational counseling became increasingly important at the turn of the twentieth century as large numbers of people moved to the bigger cities in search of employment. With vocational counseling seen as a process of (1) acquiring self-knowledge, (2) acquiring knowledge of the world of work, and (3) finding a suitable match through a process called “true reasoning,” thousands of individuals were anxious to discover what jobs might be a suitable match for them. And tests that could quickly measure large groups of people’s likes and dislikes, as well as their abilities, could help do that. Thus, we began to see the rise of “multiple aptitude” tests. For example, the General Aptitude Test

GATB

Developed by U.S. Employment Service to measure multiple aptitudes

Battery (GATB) was developed by the United States Employment Service to measure abilities in a number of specific areas. These areas of ability could be directly matched with job characteristics to identify appropriate occupational choices.

The Emergence of Personality Tests (Testing in the Affective Realm)

Paralleling the rise of tests in the cognitive domain, personality tests (or tests in the affective realm) began to be devised. Thus, around the turn of the twentieth century three types of personality assessment instruments were developed: interest inventories, objective personality tests, and projective personality tests. Let's take a brief look at each of these areas.

J. B. Miner

Developed one of the first group interest inventories

Interest Inventories and Vocational Assessment In 1912, Edward Thorndike, one of the first to conduct research in the field of vocational assessment, published the results of a study that examined the interests of 100 students as they progressed from elementary school through college (DuBois, 1970). As the relationship of interests to vocational choice became more obvious, in 1922, J. B. Miner developed one of the first formal interest blanks (inventories) that was used to assist large groups of high school students in selecting occupations. Miner (1922) understood that his test was only part of the total assessment process, and he explained that his inventory was "the basis for individual interviews with vocational counselors" (p. 312).

Edward Strong
Founder of the Strong Vocational Interest Blank—derivative still used today

On the heels of Miner's interest blank, in the mid-1920s Edward Strong (1884–1963) teamed up with a number of other researchers to develop what was to become the most well-known interest inventory (Cowdery, 1926; DuBois, 1970; Strong, 1926). Known as the Strong Vocational Interest Blank, the original inventory consisted of 420 items. Strong spent the rest of his life perfecting his vocational interest inventory. Having undergone numerous revisions over the years, this inventory continues to be one of the most widely used instruments in career counseling. Today, interest inventories like the Strong are often used in conjunction with multiple aptitude tests as part of the career counseling process.

Emil Kraepelin
Developed early word association test

Woodworth's Personal Data Sheet
First modern personality inventory—used during WWI

Objective Personality Assessment Although Emil Kraepelin developed a crude word association test to study schizophrenia in the 1880s, Woodworth's Personal Data Sheet is considered to be the ancestor of most modern-day personality inventories. Woodworth's instrument, which was developed to screen WWI recruits for their susceptibility to mental health problems (DuBois, 1970), had 116 items to which individuals were asked to respond by underlining "yes" or "no" to indicate whether or not the statement represented them (see Box 1.4). These rather obvious questions were then related to certain types of neuroses and pathologies. Although the test had questionable validity compared to today's instruments, it became an early model for a number of other, better-refined instruments including the Minnesota Multiphasic Personality Inventory (MMPI).

Projective Testing

Experiments such as these allow an unexpected amount of illumination to enter into the deepest recess of the character, which are opened and bared by them like the

BOX 1.4 Items from Woodworth's Personal Data Sheet

Although crude by today's standards, Woodworth's Personal Data Sheet was one of the first instruments that attempted to assess one's personality. Below are some of the original 116 items.

Answer the questions by underlining "Yes" when you mean yes, and by underlining "No" when you mean no. Try to answer every question.

- 1. Do you usually feel well and strong? YES NO
- 3. Are you often frightened in the middle of the night? YES NO

- 27. Have you ever been blind, half-blind, deaf or dumb for a time? YES NO
- 51. Have you hurt yourself by masturbation (self abuse)? YES NO
- 80. At night are you troubled by the idea that somebody is following you? YES NO
- 112. Has any of your family been a drunkard? YES NO

Source: Adapted from *A History of Psychological Testing* (pp. 160-183), by P. DuBois, 1970, Boston: Allyn and Bacon.

anatomy of an animal under the scalpel of a dissector in broad daylight (Galton, 1879, p. 431).

These words of Galton speak to the premise of projective testing: present a stimulus to an individual in an attempt to tap into the unconscious mind and discover the inner world of that person. Recognizing the importance of Galton's work, Cattell examined the kinds of associations that mentally healthy individuals made to a standard list of words (DuBois, 1970).

Carl Jung
Used word associations to identify mental illness

Herman Rorschach
Developed famous Rorschach Inkblot test

Henry Murray
Developed Thematic Apperception Test (TAT)

By 1904, Carl Jung (1875-1961) had come up with 156 stimulus words that he used in one of the earliest word association tests. Examining the responses of his clients to his list of words, Jung came up with the word *complex* to describe sets of unusual and delayed responses that individuals had to these stimulus words that seemed to point to a problematic or neurotic area in their lives (Jung, 1918/1969; Jung & Riklin, 1904; Storr, 1973). However, it was Herman Rorschach (1884-1922), a student of Jung, who developed the most well-known projective test—the Rorschach Inkblot test. Rorschach created this test by selecting ten inkblots “thrown on a piece of paper, the paper folded, and the ink spread between the two halves of the sheet” (Rorschach, 1942, p. 1). He believed the interpretation of an individual's reactions to these forms could tell volumes about the individual's unconscious life. This test was the precursor to many other kinds of projective tests, such as Henry Murray's Thematic Apperception Test, or TAT, which asks a subject to view a number of standard pictures and create a story that explains the situation.

Informal assessment procedures
User-created and situational

The Emergence of Informal Assessment Procedures

The twentieth century saw the increased use of informal assessment procedures, which are assessment instruments that are often developed by the user and designed to meet a particular testing situation. For instance, as business and industry expanded during the 1930s, the situational test became more prevalent. In these tests, businesses

took individuals who were potential hires or candidates for promotion and placed them into “contrived naturalistic situations” to assess their ability to respond to real-life situations. Meanwhile, as treatment of mental health clients improved, another informal procedure, the clinical interview, became prominent. The clinical interview became especially important as clients were increasingly being assessed for a diagnosis through the use of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM), first developed by the American Psychiatric Association in 1952 (Neukrug & Schwitzer, 2006).

During the 1960s and 1970s the use of tests in schools greatly increased and laws were passed that called for the assessment of students with disabilities. It was then that a number of informal techniques became popular, including observation, rating scales, classification techniques, and the review of records and personal documents to assess learning problems of children. Conducting an environmental assessment where many of these informal tools are used has also become popular in recent years. For instance, knowing whether a particular home (the “environment”) is a healthy place to raise a child has become an important aspect of the Child Protective Service worker’s role. Finally, in recent years, performance-based assessment as an alternative to the more traditional cognitive-based assessments (e.g., multiple choice tests) has become increasingly popular. Today, informal assessment techniques, such as those already mentioned, are used in a variety of settings in numerous ways.

Modern-Day Use of Assessment Procedures

As complex statistical analyses became possible through the use of computers, the quality of assessment instruments advanced rapidly. Today, assessment instruments can be found in every aspect of society and their uses have been vastly expanded. Although one can categorize such instruments in many ways, we have found it helpful to classify them into the following groups: (1) testing in the cognitive domain, often called “ability testing,” (2) testing in the affective domain, usually called “personality assessment,” and (3) informal assessment procedures. Figures 1.3–1.5 are graphic displays of these domains and are followed in Box 1.5 by short definitions of the various categories.

In Section III of the text, we will demonstrate how all of the assessment categories noted in Figures 1.3–1.5 and Box 1.5 are used today in a variety of ways. And the following shows how the chapters in Section III of the text correspond to the categories:

Chapter 8: Assessment of Educational Ability: Survey Battery, Diagnostic, Readiness, and Cognitive Ability Tests

Chapter 9: Intellectual and Cognitive Functioning: Intelligence Testing and Neuropsychological Assessment

Chapter 10: Career and Occupational Assessment: Interest Inventories, Multiple Aptitude, and Special Aptitude Tests

Chapter 11: Clinical Assessment: Objective and Projective Personality Tests

Chapter 12: Informal Assessment: Observation, Rating Scales, Classification Methods, Environmental Assessment, Records and Personal Documents, and Performance-Based Assessment

ch 8-9

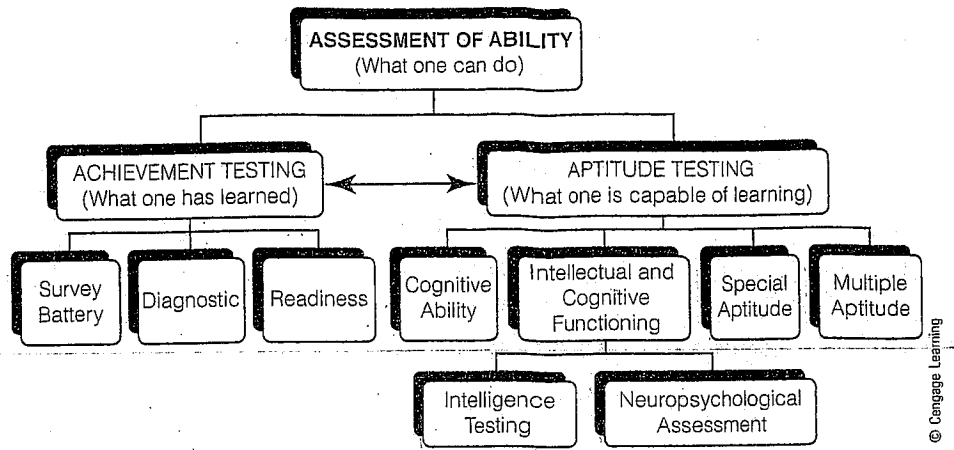


FIGURE 1.3 Assessment in the Cognitive Domain

ch 10-11

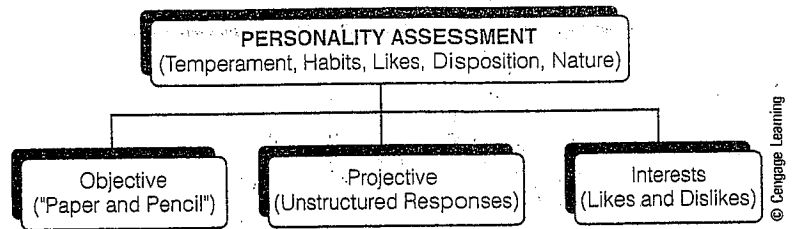


FIGURE 1.4 Assessment in the Affective Domain

ch 12

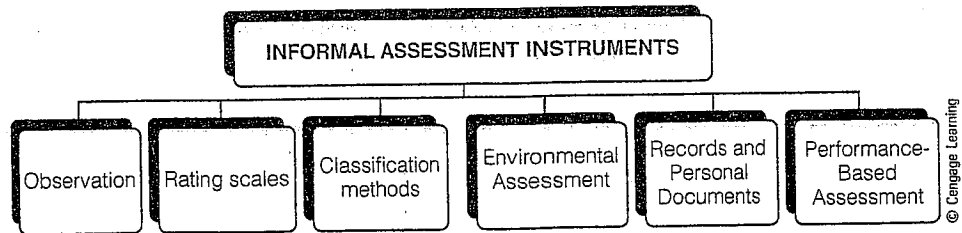


FIGURE 1.5 Informal Assessment Procedures

BOX 1.5**Brief Definitions of Assessment Categories****Assessment of Ability:**

Tests that measure what a person can do in the cognitive realm.

Achievement Testing:

Tests that measure what one has learned.

Survey Battery Tests: Tests, usually given in school settings, which measure broad content areas. Often used to assess progress in school.

Diagnostic Tests: Tests that assess problem areas of learning. Often used to assess learning disabilities.

Readiness Tests: Tests that measure one's readiness for moving ahead in school. Often used to assess readiness to enter first grade.

Aptitude Testing:

Tests that measure what one is capable of learning.

Intellectual and Cognitive Functioning: Tests that measure a broad range of cognitive functioning in the following domains: general intelligence, intellectual disabilities, giftedness, and changes in overall cognitive functioning. Includes intelligence testing that leads to an "IQ" score and neuro-psychological assessment that assesses changes in cognitive functioning over time.

Cognitive Ability Tests: Tests that measure a broad range of cognitive ability. These tests are usually based on what a student has learned in school and are useful in making predictions about the future (e.g., whether an individual might succeed in college).

Special Aptitude Tests: Tests that measure one aspect of ability. Often useful in determining the likelihood of success in a vocation (e.g., a mechanical aptitude test to determine potential success as a mechanic).

Multiple Aptitude Tests: Tests that measure many aspects of ability. Often useful in determining the likelihood of success in a number of vocations.

Personality Assessment:

Tests in the affective realm used to assess habits, temperament, likes and dislikes, character, and similar behaviors.

Objective Personality Testing: Multiple choice and true-false tests that assess various aspects of personality. Often used to increase client insight, to identify psychopathology, and to assist in treatment planning.

Projective Personality Tests: Tests that present a stimulus to which individuals can respond. Personality factors are interpreted based on the individual's response. Often used to identify psychopathology and to assist in treatment planning.

Interest Inventories: Tests that measure likes and dislikes as well as one's personality orientation toward the world of work. Generally used in career counseling.

Informal Assessment Instruments:

Often developed by the user, these tests tend to assess broad areas of ability or personality and tend to be specific to the testing situation.

Observation: Observing an individual in order to develop a deeper understanding of one or more specific behaviors (e.g., observing a student's acting-out behavior in class or assessing a client's ability to perform eye-hand coordination tasks as a means of determining potential vocational placement).

Rating Scales: Scales developed to assess any of a number of attributes of the examinee. Can be rated by the examinee or someone who knows the examinee well (e.g., rating a faculty member's teaching ability or a student's ability to make empathic responses).

Classification Methods: A tool whereby an individual identifies whether he or she has, or does not have, specific attributes or characteristics (e.g., from a list, checking adjectives that seem to be most like you).

Environmental Assessment: A naturalistic and systemic approach to assessment in which information about clients is collected from their home, work, school, or other places through observation, self-reports, and checklists.

Records and Personal Documents: Items such as diaries, personal journals, genograms, and school records that are examined

to gain a broader understanding of an individual.

Performance-Based Assessment: The evaluation of an individual using informal assessment procedures based on real-world activities that are not highly loaded for cognitive skills. These procedures are seen as an alternative to standardized testing (e.g., a portfolio).

QUESTIONS TO CONSIDER WHEN ASSESSING INDIVIDUALS

It is clear that today's assessment instruments have widespread applications. With the knowledge that many individuals have used assessment instruments for less than honorable reasons (e.g., the Eugenics Movement), it is critical that we remain vigilant about the use of such instruments. Keeping this in mind, we should continually be asking ourselves some important questions regarding the use of assessment instruments, including the following:

1. How valid is the information gained from assessment instruments, and how should that information be applied?
2. How do assessment instruments invade an individual's privacy, and does the government have, at times, the right to insist that an individual be assessed?
3. Can the use of some assessment instruments lead to labeling, and what are the implications for individuals who are "labeled"?
4. Are assessment procedures used to foster equality for all people, or do they tend to reinforce existing societal divisions based on class?

SUMMARY

We began this chapter by defining assessment and pointing out that assessment encompasses a broad range of techniques, including testing. We noted that modern-day assessment has been greatly influenced by the long history of assessment. Going back to 2200 B.C.E., we pointed out that the Chinese developed one of the first widely used tests and that hundreds of years ago the Greeks assessed intellectual and physical ability of men for state service. As we neared the modern era of testing, we pointed out that individuals such as Esquirol examined the relationship between language ability and intelligence, while others, such as Seguin, looked at the relationship between

motor control and intelligence. We noted that Darwin's theory of evolution spurred on others such as Galton, Wundt, Cattell, and Hall to examine individual differences, a focus that would be critical to the nature of assessment.

We next pointed out that some of the first ability tests were neuropsychological assessments and intelligence testing. We noted that interest in brain functioning could be traced by to early Egypt, around 2500 B.C. and moving forward in time, we noted that the late 1800s saw Alfred Binet develop the first intelligence test. Later revised by Terman, the Stanford-Binet compared an individual's chronological age to his or her

mental age. We pointed out that intelligence testing was also sometimes used as one aspect of neuropsychological assessments, which became more important with the advent of brain-imaging techniques and the use of additional methods to assess the relationship of brain function to behavior.

As the chapter continued, we noted that the early 1900s saw the development of group tests of ability, including the Army Alpha and Army Beta, achievement tests in the schools, and multiple aptitude tests. We then noted that during the 1900s, individuals such as Galton, Yerkes, and Terman were influential in the Eugenics Movement. This misguided venture attempted to use test data to show intellectual differences among cultural groups and ended up influencing government policy, including laws regarding who would be allowed to immigrate to the United States.

We pointed out that many of the first personality tests paralleled the development of ability tests. For instance, the early 1900s saw Thorndike, Miner, and Strong research the area of vocational assessment and develop some of the first interest inventories. Kraepelin developed a crude word association test, and Woodworth developed his

“personal data sheet,” which many consider the precursor of modern-day personality inventories. These early assessment instruments were soon followed by the development of one of the first projective tests, a word association test by Carl Jung who sought to uncover what kinds of reactions caused complexes. Other early projective tests included Rorschach’s Inkblot Test and Murray’s Thematic Apperception Test. As the twentieth century continued, a number of informal assessment instruments were developed, including observational techniques, rating scales, classification schemes, records and personal documents, environmental assessments, and performance-based assessments.

Near the end of the chapter, we reviewed the definitions of the various assessment categories, including those assessment techniques that are found in the following domains: ability testing (achievement and aptitude), personality testing, and informal assessment instruments. The chapter concluded by highlighting a number of important issues in assessment, including test validity, invasion of privacy, caution regarding labeling, and the importance of assuring that assessment procedures foster equality.

CHAPTER REVIEW

1. Identify some of the ancient precursors to assessment.
2. Identify some of the precursors to modern-day assessment during the 1800s.
3. Discuss how the work of Darwin, Galton, Wundt, and Cattell has influenced the development of modern-day testing.
4. Identify some of the individuals involved and describe the precursors to modern-day intelligence testing.
5. What was the Eugenics Movement, and how did it influence government policy in the United States?
6. Identify some of the early group tests of ability, the main players in their development, and their uses.
7. Identify some of the early personality tests and the main players involved in their development.
8. Describe the contributions of some of the early developers of projective testing.
9. Draw three diagrams (see Figures 1.3–1.5) that list the various kinds of achievement testing, aptitude testing, personality assessment, and informal assessment. Define each type of assessment category on your diagrams.
10. Make a list of every historical figure discussed in this chapter, and define each person’s contribution to testing and assessment.