

## *Historical, Cultural, and Legal/Ethical Considerations*

**W**e continue our broad overview of the field of psychological testing and assessment with a look backward, the better to appreciate the historical context of the enterprise. We also present “food for thought” regarding cultural and legal/ethical matters. Consider this “food” only as an appetizer; material on historical, cultural, and legal/ethical considerations is interwoven where appropriate throughout this book.

### **A Historical Perspective**

#### *Antiquity to the Nineteenth Century*

It is believed that tests and testing programs first came into being in China as early as 2200 B.C.E. (DuBois, 1966, 1970). Testing was instituted as a means of selecting who, of many applicants, would obtain government jobs. In a culture where one’s position in society had a long tradition of being determined solely by the family into which one was born, the fact that one could improve one’s lot in life by scoring high on an examination was a significant step forward. In reality, passing the examinations required knowledge that usually came either from long hours of study or work with a tutor. Given those facts of life, it was likely that only the land-owning gentry could afford to have their children spend the time necessary to prepare for the tests. Still, tales emerged of some people who were able to vastly improve their lot in life by passing the state-sponsored examinations.

What were the job applicants tested on? As might be expected, the content of the examination changed over time and with the cultural expectations of the day—as well as with the values of the ruling dynasty. In general, proficiency in endeavors such as music, archery, horsemanship, writing, and arithmetic were examined. Also important were subjects such as agriculture, geography, revenue, civil law, and military strategy. Knowledge and skill with respect to the rites and ceremonies of public and social life were also evaluated. During the Song dynasty, emphasis was placed on knowledge of classical literature. Testtakers who demonstrated their command of the classics were perceived as having acquired the wisdom of the past; they were therefore entitled to a government position. During some dynasties, testing was virtually suspended and government positions were given to family members or friends, or simply sold.

**Figure 2-1**  
**Releasing the Roll**

For a period of about three thousand years, forms of proficiency testing existed in China. Some time after taking a test, men—the tests were open only to men with the exception of a brief period in the 1800s—gathered to see who passed when the results were posted on a wall (sound familiar?). This posting was referred to as “releasing the roll.”



In dynasties where state-sponsored examinations, referred to as *imperial examinations*, for official positions were in force, the consequential privileges for succeeding varied. During some periods, in addition to a government job, those who passed the examination were entitled to wear special garb; this entitled them to be accorded special courtesies by anyone they happened to meet. In some dynasties, passing the examinations could result in exemption from taxes. Passing the examination could also exempt one from government-sponsored interrogation by torture if the individual was suspected of committing a crime. Clearly, it paid to do well on these difficult examinations.

**JUST THINK . . .**

What parallels can you draw between doing well on difficult examinations in ancient China and doing well on difficult examinations today?

Also intriguing from a historical perspective are ancient Greco-Roman writings indicative of attempts to categorize people in terms of personality types. Such categorizations

typically included reference to an overabundance or deficiency in some bodily fluid (such as blood or phlegm) as a factor believed to influence personality. During the Middle Ages, a question of critical importance was “Who is in league with the Devil?” and various measurement procedures were devised to address this question. It would not be until the Renaissance that measurement in the modern sense began to emerge. By the eighteenth century, Christian von Wolff (1732, 1734) had anticipated psychology as a science and psychological measurement as a specialty within that science.

In 1859, a book was published entitled *On the Origin of Species by Means of Natural Selection* by Charles Darwin (1809–1882). In this important, far-reaching work, Darwin

argued that chance variation in species would be selected or rejected by nature according to adaptivity and survival value. He further argued that humans had descended from the ape as a result of such chance genetic variations. This revolutionary notion aroused interest, admiration, and a good deal of enmity. The enmity arose primarily from members of the religious community who interpreted Darwin's ideas as an affront to the biblical account of creation in Genesis. Still, the notion of an evolutionary link between human beings and animals conferred a new scientific respectability on experimentation with animals. It also raised questions about how animals and humans compare with respect to states of consciousness—questions that would beg for answers in laboratories of future behavioral scientists.<sup>1</sup>

History records that it was Darwin who spurred scientific interest in individual differences. Darwin (1859) wrote:

The many slight differences which appear in the offspring from the same parents . . . may be called individual differences. . . . These individual differences are of the highest importance . . . [for they] afford materials for natural selection to act on. (p. 125)

Indeed, Darwin's writing on individual differences kindled interest in research on heredity in his half cousin, Francis Galton. In the course of his efforts to explore and quantify individual differences between people, Galton became an extremely influential contributor to the field of measurement (Forrest, 1974). Galton (1869) aspired to classify people "according to their natural gifts" (p. 1) and to ascertain their "deviation from an average" (p. 11). Along the way, Galton would be credited with devising or contributing to the development of many contemporary tools of psychological assessment including questionnaires, rating scales, and self-report inventories.

Galton's initial work on heredity was done with sweet peas, in part because there tended to be fewer variations among the peas in a single pod. In this work, Galton pioneered the use of a statistical concept central to psychological experimentation and testing: the coefficient of correlation. Although Karl Pearson (1857–1936) developed the product-moment correlation technique, its roots can be traced directly to the work of Galton (Magnello & Spies, 1984). From heredity in peas, Galton's interest turned to heredity in humans and various ways of measuring aspects of people and their abilities.

At an exhibition in London in 1884, Galton displayed his Anthropometric Laboratory where, for three or four pence—depending on whether you were already registered or not—you could be measured on variables such as height (standing), height (sitting), arm span, weight, breathing capacity, strength of pull, strength of squeeze, swiftness of blow, keenness of sight, memory of form, discrimination of color, and steadiness of hand. Through his own efforts and his urging of educational institutions to keep anthropometric records on their students, Galton excited widespread interest in the measurement of psychology-related variables.

Assessment was also an important activity at the first experimental psychology laboratory, founded at the University of Leipzig in Germany by Wilhelm Max Wundt (1832–1920), a medical doctor whose title at the university was professor of philosophy. Wundt and his students tried to formulate a general description of human abilities with respect to variables such as reaction time, perception, and attention span. In contrast to Galton, Wundt focused on questions relating to how people were similar, not

### JUST THINK . . .

A critical "diagnostic" question during the Middle Ages was "Who is in league with the Devil?" What would you say the most critical diagnostic question is today?

---

1. The influence of Darwin's thinking is also apparent in the theory of personality formulated by Sigmund Freud. From a Darwinian perspective, the strongest people with the most efficient sex drives would have been most responsible for contributing to the human gene pool. In this context, Freud's notion of the primary importance of instinctual sexual and aggressive urges can be better understood.

different. In fact, individual differences were viewed by Wundt as a frustrating source of error in experimentation. Wundt attempted to control all extraneous variables in an effort to reduce error to a minimum. As we will see, such attempts are fairly routine in contemporary assessment. The objective is to ensure that any observed differences in performance are indeed due to differences between the people being measured and not to any extraneous variables. Manuals for the administration of many tests provide explicit instructions designed to hold constant or “standardize” the conditions under which the test is administered. This is so that any differences in scores on the test are due to differences in the testtakers rather than to differences in the conditions under which

### JUST THINK . . .

Which orientation in assessment research appeals to you more, the Galtonian orientation (researching how individuals differ) or the Wundtian one (researching how individuals are the same)? Why?

the test is administered. In Chapter 4 we will go in to more detail about the meaning of terms such as *standardized* and *standardization* as applied to tests.

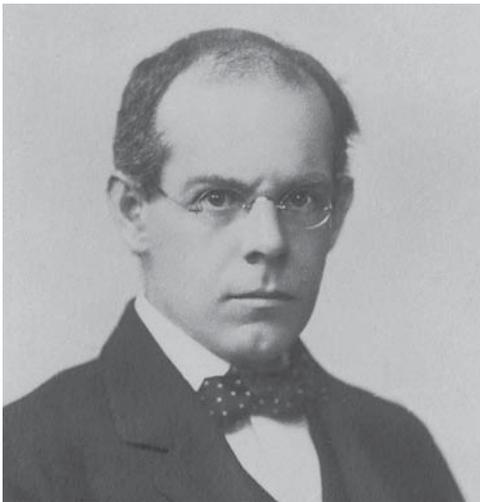
In spite of the prevailing research focus on people’s similarities, one of Wundt’s students at Leipzig, an American named James McKeen Cattell (Figure 2–2), completed a doctoral dissertation that dealt with individual differences—specifically, individual differences in reaction time. After receiving his doctoral degree from Leipzig, Cattell returned to the United States, teaching at Bryn Mawr and then at the University of Pennsylvania before leaving for Europe to teach at Cambridge. At Cambridge, Cattell came in contact with Galton, whom he later described as “the greatest man I have known” (Roback, 1961, p. 96).

Inspired by his interaction with Galton, Cattell returned to the University of Pennsylvania in 1888 and coined the term *mental test* in an 1890 publication. Boring (1950, p. 283) noted that “Cattell more than any other person was in this fashion responsible for getting mental testing underway in America, and it is plain that his motivation was similar to Galton’s and that he was influenced, or at least reinforced, by Galton.” Cattell went on to become professor and chair of the psychology department at Columbia University. Over the next 26 years, he not only trained many psychologists but also founded a number of publications (such as *Psychological Review*, *Science*, and *American Men of Science*). In 1921, Cattell was instrumental in founding the Psychological Corporation, which named 20 of the country’s leading psychologists as its directors. The goal of the corporation was the “advancement of psychology and the promotion of the useful applications of psychology.”<sup>2</sup>

Other students of Wundt at Leipzig included Charles Spearman, Victor Henri, Emil Kraepelin, E. B. Titchener, G. Stanley Hall, and Lightner Witmer. Spearman is credited with originating the concept of test reliability as well as building the mathematical framework for the statistical technique of factor analysis. Victor Henri is the Frenchman who would collaborate with Alfred Binet on papers suggesting how mental tests could be used to measure higher mental processes (for example, Binet & Henri, 1895a, 1895b, 1895c). Psychiatrist Emil Kraepelin was an early experimenter with the word association technique as a formal test (Kraepelin, 1892, 1895). Lightner Witmer received his Ph.D. from Leipzig and went on to succeed Cattell as director of the psychology laboratory at the University of Pennsylvania. Witmer has been cited as the “little-known founder of clinical psychology” (McReynolds, 1987), owing at least in part to his being challenged to treat a “chronic bad speller” in March of 1896 (Brotemarkle, 1947). Later that year, Witmer founded the first psychological clinic in the United States at the University of Pennsylvania. In 1907, Witmer founded the journal *Psychological Clinic*. The first article in that journal was entitled “Clinical Psychology” (Witmer, 1907).

---

2. Today, many of the products and services of what was once known as the Psychological Corporation have been absorbed under the “PsychCorp” brand of a corporate parent, Pearson Assessment, Inc.



**Figure 2–2**  
**The Cattells, James McKeen and Psyche**

*The psychologist who coined the term mental test, James McKeen Cattell (1860–1944), has often been mistakenly credited (along with another psychologist, Raymond B. Cattell, no relation) with the authorship of a measure of infant intelligence called the Cattell Infant Intelligence Scale (CIIS). Actually, it was Psyche (1893–1989), the third of seven children of Cattell and his wife, Josephine Owen, who created the CIIS. From 1919 through 1921, Psyche assisted her famous father in statistical analyses for the third edition of *American Men of Science*. In 1927, she earned a doctor of education degree at Harvard. In 1931, she adopted a son, becoming one of the first unmarried women to do so (Sokal, 1991). Later in the decade she adopted a daughter. Her book *The Measurement of Intelligence in Infants and Young Children* was published in 1940, and it was in that book that the CIIS was introduced. Later in her career, she would write a popular book, *Raising Children with Love and Limits*, which refuted the permissiveness advocated by child-rearing authorities such as Benjamin Spock.*

### *The Twentieth Century*

Much of the nineteenth-century testing that could be described as psychological in nature involved the measurement of sensory abilities, reaction time, and the like. Generally, the public was fascinated by such testing. However, there was no widespread belief that testing for variables such as reaction time had any applied value. But all of that would change in the early 1900s with the birth of the first formal tests of intelligence, tests that could really be useful for reasons readily understandable to anyone who had school-age children. As we will see, public receptivity to psychological tests would shift from mild curiosity to outright enthusiasm as more and more instruments that purportedly quantified mental ability were introduced. Soon there would be tests to measure sundry mental characteristics such as personality, interests, attitudes, values, and widely varied mental abilities. It all began with a single test designed for use with young Paris pupils.

**The measurement of intelligence** As early as 1895, Alfred Binet (1857–1911) and his colleague Victor Henri published several articles in which they argued for the measurement of abilities such as memory and social comprehension. Ten years later, Binet and collaborator Theodore Simon published a 30-item “measuring scale of intelligence” designed to help identify mentally retarded Paris schoolchildren (Binet & Simon, 1905). The Binet test would go through many revisions and translations—and, in the process, launch both the intelligence testing movement and the clinical testing movement.

### JUST THINK . . .

In the early 1900s, the Binet test was being used worldwide for various purposes far beyond identifying exceptional Paris schoolchildren. What do you think were some of the other uses of the test? How appropriate do you think it was to use this test for these other purposes?

Before long, psychological tests were being used with regularity in such diverse settings as schools, hospitals, clinics, courts, reformatories, and prisons (Pintner, 1931).

In 1939, David Wechsler, a clinical psychologist at Bellevue Hospital in New York City, introduced a test designed to measure adult intelligence. For Wechsler, intelligence was “the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment” (1939, p. 3). Originally christened the Wechsler-Bellevue Intelligence Scale, the test

was subsequently revised and renamed the Wechsler Adult Intelligence Scale (WAIS). The WAIS has been revised several times since then, and versions of Wechsler’s test have been published that extend the age range of testtakers from young children through senior adulthood. These tests will be discussed in greater detail in the chapters that deal with the assessment of intelligence.

### JUST THINK . . .

Should the definition of *intelligence* change as one moves from infancy through childhood, adolescence, adulthood, and senior adulthood?

A natural outgrowth of the individually administered intelligence test devised by Binet was the *group* intelligence test. Group intelligence tests came into being in the United States in response to the military’s need for an efficient method of screening the intellectual ability of World War I recruits. This same need again became urgent as the

United States prepared for entry into the Second World War. Psychologists would again be called upon by the government service to develop group tests, administer them to recruits, and interpret the test data.

### JUST THINK . . .

What do you think are the advantages of group intelligence testing? What do you think are the disadvantages of group intelligence testing?

After the war, psychologists returning from military service brought back a wealth of applied testing skills that would be useful in civilian as well as governmental applications. Psychological tests were increasingly used in diverse settings, including large corporations and private organizations. New tests were being developed at a brisk pace to measure various abilities and interests as well as personality.

**The measurement of personality** Public receptivity to tests of intellectual ability spurred the development of many other types of tests (Garrett & Schneck, 1933; Pintner, 1931). Only eight years after the publication of Binet’s scale, the field of psychology was being criticized for being too test oriented (Sylvester, 1913). By the late 1930s, approximately four thousand different psychological tests were in print (Buros, 1938), and “clinical psychology” was synonymous with “mental testing” (Institute for Juvenile Research, 1937; Tulchin, 1939).

World War I had brought with it not only the need to screen the intellectual functioning of recruits but also the need to screen for recruits’ general adjustment. A government Committee on Emotional Fitness chaired by psychologist Robert S. Woodworth was assigned the task of developing a measure of adjustment and emotional stability that could be administered quickly and efficiently to groups of recruits. The committee developed several experimental versions of what were, in essence, paper-and-pencil psychiatric interviews. To disguise the true purpose of one such test, the questionnaire was labeled as a “Personal Data Sheet.” Draftees and volunteers were asked to indicate *yes* or *no* to a series of questions that probed for the existence of various kinds of psychopathology. For example, one of the test questions was “Are you troubled with the idea that people are watching you on the street?”

The Personal Data Sheet developed by Woodworth and his colleagues never went beyond the experimental stages, for the treaty of peace rendered the development of

this and other tests less urgent. After the war, Woodworth developed a personality test for civilian use that was based on the Personal Data Sheet. He called it the Woodworth Psychoneurotic Inventory. This instrument was the first widely used **self-report test** of personality—a method of assessment that would soon be employed in a long line of succeeding personality tests.

Personality tests that employ self-report methodologies have both advantages and disadvantages. On the face of it, respondents are arguably the best-qualified people to provide answers about themselves. However, there are also compelling arguments *against* respondents supplying such information. For example, respondents may have poor insight into themselves. One might honestly believe something about oneself that in reality is not true. Regardless of the quality of insight, some respondents are unwilling to reveal anything about themselves that is very personal or that could put them in a negative light. Given these shortcomings of the self-report method of personality assessment, a need existed for alternative types of personality tests.

Filling the need for measures of personality that did not rely on self-report were various methods. One such method or approach to personality assessment came to be described as *projective* in nature. As we will see later in this book, a **projective test** is one in which an individual is assumed to “project” onto some ambiguous stimulus his or her own unique needs, fears, hopes, and motivation. The ambiguous stimulus might be an inkblot, a drawing, a photograph, or something else. Perhaps the best known of all projective tests is the Rorschach, a series of inkblots developed by the Swiss psychiatrist Hermann Rorschach. The use of pictures as projective stimuli was popularized in the late 1930s by Henry A. Murray, Christiana D. Morgan, and their colleagues at the Harvard Psychological Clinic. When pictures or photos are used as projective stimuli, respondents are typically asked to tell a story about the picture they are shown. The stories told are then analyzed in terms of what needs and motivations the respondents may be projecting onto the ambiguous pictures. Projective and many other types of instruments used in personality assessment will be discussed in a subsequent chapter devoted to that subject.

**The academic and applied traditions** Like the development of its parent field of psychology, the development of psychological measurement can be traced along two distinct threads: the academic and the applied. In the tradition of Galton, Wundt, and other scholars, psychological testing and assessment are practiced today in university psychology laboratories as a means of furthering knowledge about human and animal behavior. There is also a very strong applied tradition, one that dates back in modern times to the work of people like Binet and in ancient times to China and the administration of competitive civil service examinations. Which child should be placed in which class? Which person is best suited for the job? Society requires answers to questions such as these, and tests and measures used in a competent manner can help provide answers.

Today, perhaps more than ever before, there is a great appreciation for the role of culture in the human experience. So, whether in academic or applied settings, assessment professionals recognize the need for cultural sensitivity in the development and use of psychological tests. Let’s briefly look at some of the major issues that such sensitivity entails. These and related issues are elaborated and expanded on throughout this book as they relate to specific aspects of testing and assessment.

#### JUST THINK . . .

Who is best qualified to provide information about one’s personality? Describe the ideal candidate for personality by means of self-report.

#### JUST THINK . . .

What potential problems do you think might attend the use of projective methods to assess personality?

## Culture and Assessment

**Culture** may be defined as “the socially transmitted behavior patterns, beliefs, and products of work of a particular population, community, or group of people” (Cohen, 1994, p. 5). As taught to us by parents, peers, and societal institutions such as schools, culture prescribes many behaviors and ways of thinking. Spoken language, attitudes toward elders, and techniques of child rearing are but a few critical manifestations of culture. Culture teaches specific rituals to be performed at birth, marriage, death, and other momentous occasions. Culture imparts much about what is to be valued or prized as well as what is to be rejected or despised. Culture teaches a point of view about what it means to be born of one or another gender, race, or ethnic background. Culture

### JUST THINK . . .

Can you think of one or two ways in which you are a product of your culture? How about one or two ways this fact might come through on a psychological test?

teaches us something about what we can expect from other people and what we can expect from ourselves. Indeed, the influence of culture on an individual’s thoughts and behavior may be a great deal stronger than most of us would acknowledge at first blush.

Professionals involved in the assessment enterprise have shown increasing sensitivity to the role of culture in many different aspects of measurement. This sensitivity is manifested in greater consideration of cultural issues with respect to every aspect of test development and use, including decision making on the basis of test data. Unfortunately, it was not always that way.

### *Evolving Interest in Culture-Related Issues*

Soon after Alfred Binet introduced intelligence testing in France, the U.S. Public Health Service began using such tests to measure the intelligence of people seeking to immigrate to the United States (Figure 2–3). Henry H. Goddard, who had been highly instrumental in getting Binet’s test adopted for use in various settings in the United States, was the chief researcher assigned to the project. Early on, Goddard raised questions about how meaningful such tests are when used with people from various cultural and language backgrounds. Goddard (1913) used interpreters in test administration, employed a bilingual psychologist, and administered mental tests to selected immigrants who appeared mentally retarded to trained observers. Although seemingly sensitive to cultural issues in assessment, Goddard’s legacy with regard to such sensitivity is, at best, controversial. Goddard found most immigrants from various nationalities to be mentally deficient when tested. In one widely quoted report, 35 Jews, 22 Hungarians, 50 Italians, and 45 Russians were selected for testing among the masses of immigrants being processed for entry into the United States at Ellis Island. Reporting on his findings in a paper entitled “Mental Tests and the Immigrant,” Goddard (1917) concluded that, in this sample, 83% of the Jews, 80% of the Hungarians, 79% of the Italians, and 87% of the Russians were feebleminded. Although Goddard had written extensively on the genetic nature of mental deficiency, it is to his credit that he did not summarily conclude that these test findings were the result of hereditary. Rather, Goddard (1917) wondered aloud whether the findings were due to “hereditary defect” or “apparent defect due to deprivation” (p. 243). In reality, the findings were largely the result of using a translated Binet test that overestimated mental deficiency in native English-speaking populations, let alone immigrant populations (Terman, 1916).

Goddard’s research, although leaving much to be desired methodologically, fueled the fires of an ongoing nature–nurture debate about what intelligence tests actually measure. On one side were those who viewed intelligence test results as indicative of



**Figure 2-3**  
**Psychological Testing at Ellis Island**

*Immigrants coming to America via Ellis Island were greeted not only by the Statue of Liberty but also by immigration officials ready to evaluate them with respect to physical, mental, and other variables. Here, a block design test, one measure of intelligence, is administered to a would-be American. Immigrants who failed physical, mental, or other tests were returned to their country of origin at the expense of the shipping company that had brought them. Critics would later charge that at least some of the immigrants who had fared poorly on mental tests were sent away from our shores not because they were actually mentally deficient but simply because they did not understand English well enough to execute instructions. Additionally, the criteria on which these immigrants from many lands were being evaluated was questioned.*

some underlying native ability. On the other side were those who viewed such data as indicative of the extent to which knowledge and skills had been acquired. More details about the highly influential Henry Goddard and his most controversial career are presented in this chapter's *Close-up*.

If language and culture did indeed have an effect on mental ability test scores, then how could a more unconfounded or “pure” measure of intelligence be obtained? One way that early test developers attempted to deal with the impact of language and culture on tests of mental ability was, in essence, to “isolate” the cultural variable. So-called **culture-specific tests**, or tests designed for use with people from one culture but not from another, soon began to appear on the scene. Representative of the culture-specific approach to test development were early versions of some of the best-known tests of intelligence. For example, the 1937 revision of the Stanford-Binet Intelligence Scale, which enjoyed widespread use until it was revised in 1960, included no minority children in the research that went into its formulation. Similarly, the Wechsler-Bellevue Intelligence Scale, forerunner of a widely used measure of adult

**JUST THINK . . .**

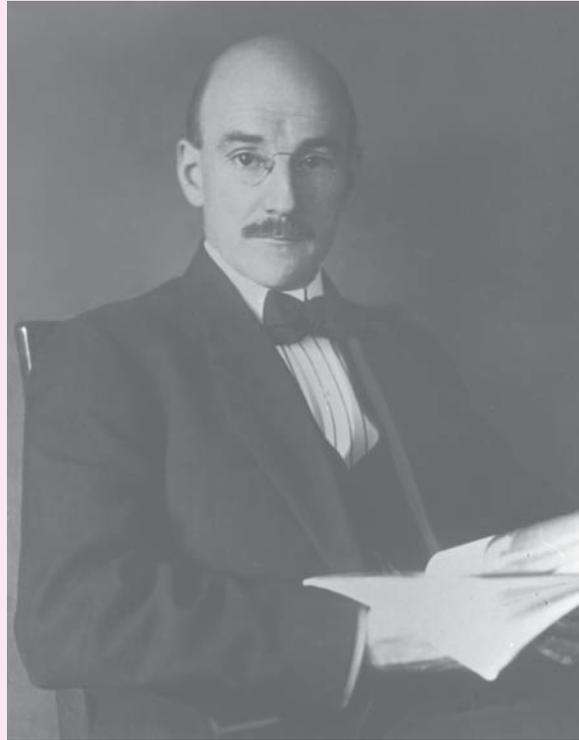
Try your hand at creating one culture-specific test item on any subject. Testtakers from what culture would probably succeed in responding correctly to the item? Testtakers from what culture would not?

## The Controversial Career of Henry Herbert Goddard

**B**orn to a devout Quaker family in Maine, Henry Herbert Goddard (1866–1957) was the fifth and youngest child born to farmer Henry Clay Goddard and Sarah Winslow Goddard. The elder Goddard was gored by a bull and succumbed to the injuries he sustained when young Henry was nine. Sarah would subsequently marry a missionary, and she and her new husband would travel the United States and abroad preaching. Young Henry attended boarding school at Oak Grove Seminary in Maine and the Friends School in Providence, Rhode Island. After earning his bachelor's degree from Haverford College, a Quaker-founded school just outside of Philadelphia, it was off to California to visit an older sister. While there he accepted a temporary teaching post at the University of Southern California (USC), which included coaching that school's football team. And so it came to pass that, among Herbert H. Goddard's many lifelong achievements, he could list the distinction of being USC's first football coach (along with a co-coach; see Pierson, 1974).

Goddard returned to Haverford in 1889 to earn a master's degree in mathematics and then took a position as a teacher, principal, and prayer service conductor at a small Quaker school in Ohio. In August of that year he married Emma Florence Robbins; the couple never had children. Goddard enrolled for study of psychology at Clark University and by 1899 had earned a doctorate under G. Stanley Hall. Goddard's doctoral dissertation, a blending of his interests in faith and science, was entitled, "The Effects of Mind on Body as Evidenced in Faith Cures."

Goddard became a professor at the State Normal School in West Chester, Pennsylvania, a teacher's college where he cultivated an interest in the growing child welfare movement. As a result of his interest in studying children, Goddard had occasion to meet Edward Johnstone, the superintendent of the New Jersey Home for Feeble-Minded Children in Vineland, New Jersey. In 1902, Goddard and Johnstone, along with educator Earl Barnes, founded a "Feeble-minded Club," which—despite its misleading name by current standards—served as an interdisciplinary forum for the exchange of ideas regarding special education. By 1906, Goddard felt frustrated in his teaching position. His friend Johnstone created the position of Director of Psychological Research at the Vineland facility and so Goddard moved to New Jersey.



In 1908, with a newfound interest in the study of "feeble-mindedness" (mental deficiency), Goddard toured psychology laboratories in Europe. It is a matter of historical interest that on this tour he did *not* visit Binet at the Sorbonne in Paris. Rather, it happened that a Belgian psychologist (Ovide Decroly) informed Goddard of Binet's work and gave him a copy of the Binet-Simon Scale. Few people at the time could appreciate just how momentous the Decroly–Goddard meeting would be nor how influential Goddard would become in terms of launching the testing movement. Returning to New Jersey, Goddard oversaw the translation of Binet's test and distributed thousands of copies of it to professionals working in various settings. Before long, Binet's test would be used in schools, hospitals, and clinics to help make diagnostic and treatment decisions. The military would use the test, as well as other newly created intelligence tests, to screen recruits. Courts would even begin to mandate the use of intelligence tests to aid in making determinations as to the intelligence of criminal

defendants. Such uses of psychological tests were very “cutting edge” at the time.

At the Vineland facility, Goddard found that Binet’s test appeared to work very well in terms of quantifying degrees of mental deficiency. Goddard devised a system of classifying assessee’s by their performance on the test, coining the term *moron* and using other such terms that today are out of favor and not in use. Goddard fervently believed that one’s placement on the test was revealing in terms of many facets of one’s life. He believed intelligence tests held the key to answers to questions about everything from what job one should be working at to what activities could make one happy. Further, Goddard came to associate low intelligence with many of the day’s most urgent social problems, ranging from crime to unemployment to poverty. According to him, addressing the problem of low intelligence was a prerequisite to addressing prevailing social problems.

Although previously disposed to believing that mental deficiency was primarily the result of environmental factors, Goddard’s perspective was radically modified by exposure to the views of biologist Charles Davenport. Davenport was a strong believer in the role of heredity to account for mental deficiency and a staunch advocate of **eugenics**, the science of improving qualities of a breed (in this case, the human race) through intervention with factors related to heredity. Davenport collaborated with Goddard in collecting hereditary information on children at the Vineland school. At Davenport’s urgings, the research included a component whereby a “eugenic field worker,” trained to identify mentally deficient individuals, would be sent out to research the mental capabilities of relatives of the residents of the Vineland facility.

The data Goddard and Davenport collected was used to argue the case that mental deficiency was caused by a recessive gene and could be inherited, much like eye color is inherited. Consequently, Goddard believed that—in the interest of the greater good of society at large—mentally deficient individuals should be segregated or institutionalized (at places such as Vineland) and not be permitted to reproduce. By publicly advocating this view, Goddard, along with Edward Johnstone, “transformed their obscure little institution in rural New Jersey into a center of international influence—a model school famous for its advocacy of special education, scientific research, and social reform” (Zenderland, 1998, p. 233).

Goddard traced the lineage of one of his students at the Vineland school back five generations in his first (and most famous) book, *The Kallikak Family: A Study in the Heredity of Feeble-Mindedness* (1912). In this book, Goddard sought

to prove how the hereditary “menace of feeble-mindedness” manifested itself in one New Jersey family. “Kallikak” was the fictional surname given to the Vineland student, Deborah, who had previous generations of relatives that were from distinctly “good” (from the Greek *kalos*) or “bad” (from the Greek *kakos*) genetic inheritance. The book traced the family lineages resulting from the legitimate and illegitimate unions of a Revolutionary War soldier with the pseudonym “Martin Kallikak.” Martin had fathered children both with a mentally defective waitress and with the woman he married—the latter being a socially prominent and reportedly normal (intellectually) Quaker. Goddard determined that feeble-mindedness ran in the line of descendants from the illegitimate tryst with the waitress. Deborah Kallikak was simply the latest descendant in that line of descendants to manifest that trait. By contrast, the line of descendants from Martin and his wife contained primarily fine citizens. But how did Goddard come to this conclusion?

One thing Goddard did *not* do was administer the Binet to all of the descendants on both the “good” and “bad” sides of Martin Kallikak’s lineage over the course of some hundred years. Instead, Goddard employed a crude case study approach ranging from analysis of official records and documents (which tended to be scarce) to reports of neighbors (which was later characterized by critics as unreliable gossip). Conclusions regarding the feeble-mindedness of descendants were likely to be linked to any evidence of alcoholism, delinquency, truancy, criminality, prostitution, illegitimacy, or economic dependence. Some of Martin Kallikak’s descendants, alive at the time the research was being conducted, were classified as feeble-minded solely on the basis of their physical appearance. Goddard (1912) wrote, for example,

The girl of twelve should have been at school, according to the law, but when one saw her face, one realized that it made no difference. She was pretty, with olive complexion and dark, languid eyes, but there was no mind there. (pp. 72–73)

Although well received by the public, the lack of sophistication in the book’s research methodology was a cause for concern for many professionals. In particular, psychiatrist Abraham Myerson (1925) attacked the Kallikak study, and the eugenics movement in general, as pseudoscience (see also Trent, 2001). Myerson reanalyzed data from studies purporting to support the idea that various physical and mental conditions could be inherited, and he criticized those studies on statistical grounds. He especially criticized Goddard for making sweeping and unfounded generalizations from questionable data. Goddard’s book became an

(continued)

## The Controversial Career of Henry Herbert Goddard (*continued*)

increasing cause for concern because it (as well as related writings on the menace of feeble-mindedness) was used to support radical arguments in favor of eugenics, forced sterilization, restricted immigration, and other social causes. Because so many people deemed feeble-minded by Goddard were so classified because of undesirable social status, illegitimacy, or “sinful” activity, it gives one cause to wonder whether Goddard’s own religious upbringing—and biblical teachings linking children’s problems with parents’ sins—was inappropriately emphasized in what was supposed to be strictly scientific writing.

After 12 years at Vineland, Goddard left under conditions that have been the subject of some speculation (Wehmeyer & Smith, 2006). From 1918 through 1922, Goddard was director of the Ohio Bureau of Juvenile Research. From 1922 until his retirement in 1938, Goddard was a psychology professor at the Ohio State University. In 1947, Goddard moved to Santa Barbara, California, where he lived out his life to the age of 90. His remains were cremated and interred at the Vineland school, along with those of his wife, who had died in 1936.

Goddard’s accomplishments were many. It was largely through his efforts that state mandates requiring special education services first became law. These laws worked to the benefit of many mentally deficient as well as many gifted students. Goddard’s introduction of Binet’s test to American society attracted other researchers, such as Lewis Terman, to see what they could do in terms of improving the test for various applications. Goddard’s writings certainly had a momentous heuristic impact on the nature–nurture question. His writings stimulated many others to research and write, if only to disprove Goddard’s conclusions. Goddard advocated for court acceptance of intelligence test data into evidence and for the limitation of criminal responsibility in the case of mentally defective defendants, especially with respect to capital crimes. He personally contributed his time to military screening efforts during the First World War. Of more dubious distinction, of course, was the Ellis Island intelligence testing program he set up to screen immigrants. Although ostensibly well intentioned, this effort resulted in the misclassification and consequential repatriation of countless would-be citizens.

Despite an impressive list of career accomplishments, the light of history has not shone favorably on Henry Goddard. Goddard’s (1912) recommendation for segregation of the mentally deficient and his calls for their sterilization tend to be viewed, at best, as misguided. The low esteem in which Goddard is generally regarded today is perhaps compounded by the fact that Goddard’s work has traditionally been held in very *high* esteem by some groups with radically offensive views, such as the Nazi party. During the late 1930s and early 1940s, over 40,000 people were euthanized by Nazi physicians simply because they were deemed mentally deficient. This action preceded the horrific and systematic mass murder of 6 million or so more innocent civilians by the Nazi military, also on a eugenic basis; the alleged “genetic defect” of these victims were that they were Jewish. Eugenicist propaganda fed to the German public by the Nazis spoke of “purifying” German blood by limiting or totally eliminating the ability of people from various groups to reproduce.

It is not a matter of controversy that Goddard used ill-advised research methods to derive many of his conclusions; he himself acknowledged this sad fact in later life. At the very least, Goddard could be criticized for being too easily influenced by the (bad) ideas of others, for being somewhat naïve in terms of how his writings were being used, and simply not being up to the task of executing methodologically sound research. He focused on the nature side of the nature–nurture controversy not because he was an ardent eugenicist at heart but rather because the nature side of the coin was where researchers at the time all tended to focus. In a letter to a critic some years later, Goddard (letter to Nicolas Pastore dated April 3, 1948, quoted in J. D. Smith, 1985) wrote, in part, that he had “no inclination to deemphasize environment . . . [but] in those days environment was not being considered.”

The conclusion of Leila Zenderland’s relatively sympathetic biography of Goddard leaves one with the impression that he was basically a decent and likeable man who was a product of his times. He harbored neither evil intentions nor right-wing prejudices. For her, a review of the life of Henry Herbert Goddard should serve as a warning not to reflexively jump to the conclusion that “bad science is usually the product of bad motives or, more broadly, bad character” (1998, p. 358).

intelligence, contained no minority members in the samples of testtakers used in its development. Although “a large number” of Blacks had, in fact, been tested (Wechsler, 1944), that data had been omitted from the final test manual because the test developers “did not feel that norms derived by mixing the populations could be interpreted without special provisos and reservations.” Hence, Wechsler (1944) stated at the outset that the Wechsler-Bellevue norms could not be used for “the colored populations of the United States.” In like fashion, the inaugural edition of the Wechsler Intelligence Scale for Children (WISC), first published in 1949 and not revised until 1974, contained no minority children in its development.

Even though many published tests were purposely designed to be culture-specific, it soon became apparent that the tests were being administered—improperly—to people from different cultures. Perhaps not surprisingly, testtakers from minority cultures tended to score lower as a group than people from the group for whom the test was developed. Illustrative of the type of problems encountered by test users was this item from the 1949 WISC: “If your mother sends you to the store for a loaf of bread and there is none, what do you do?” Many Hispanic children were routinely sent to the store for tortillas and so were not familiar with the phrase “loaf of bread.”

Today, test developers typically take many steps to ensure that a major test developed for national use is indeed suitable for such use. Those steps might involve administering a preliminary version of the test to a tryout sample of testtakers from various cultural backgrounds, particularly from those whose members are likely to be administered the final version of the test. Examiners who administer the test may be asked to describe their impressions with regard to various aspects of testtakers’ responses. For example, subjective impressions regarding testtakers’ reactions to the test materials or opinions regarding the clarity of instructions will be noted. All of the accumulated test scores from the tryout sample will be analyzed to determine if any individual item seems to be biased with regard to race, gender, or culture. In addition, a panel of independent reviewers may be asked to go through the test items and screen them for possible bias. A revised version of the test may then be administered to a large sample of testtakers that is representative of key variables of the latest U.S. Census data (such as age, gender, ethnic background, and socioeconomic status). Information from this large-scale test administration will also be used to root out any identifiable sources of bias. More details regarding the contemporary process of test development will be presented in Chapter 8.

#### JUST THINK . . .

If no minority testtakers are used in the development of a test, how appropriate is it to use the finished form of that test with minority testtakers?

### *Some Issues Regarding Culture and Assessment*

Communication between assessor and assessee is a most basic part of assessment. Assessors must be sensitive to any differences between the language or dialect familiar to assessee and the language in which the assessment is conducted. Assessors must also be sensitive to the degree to which assessee have been exposed to the dominant culture and the extent to which they have made a conscious choice to become assimilated. Next, we briefly consider issues of assessment and communication, both verbal and nonverbal, in a cultural context.

**Verbal communication** Language, the means by which information is communicated, is a key yet sometimes overlooked variable in the assessment process. Most obviously, the examiner and the examinee must speak the same language. This is necessary not

only for the assessment to proceed but also for the assessor's conclusions regarding the assessment to be reasonably accurate. If a test is in written form and includes written instructions, then the testtaker must be able to read and comprehend what is written. When the language in which the assessment is conducted is not the assessee's primary language, he or she may not fully comprehend the instructions or the test items. The danger of such misunderstanding may increase as infrequently used vocabulary or unusual idioms are employed in the assessment. All of the foregoing presumes that the assessee is making a sincere and well-intentioned effort to respond to the demands of the assessment. Although this is frequently presumed, it is not always the case. In some instances, assessees may purposely attempt to use a language deficit to frustrate evaluation efforts (Stephens, 1992).

When an assessment is conducted with the aid of a translator, different types of problems may emerge. Depending upon the translator's skill and professionalism, subtle nuances of meaning may be lost in translation, or unintentional hints to the correct or more desirable response may be conveyed. Whether translated "live" by a translator or in writing, translated items may be either easier or more difficult than the original. Some vocabulary words may change meaning or have dual meanings when translated.

In interviews or other situations in which an evaluation is made on the basis of an oral exchange between two parties, a trained examiner may detect through verbal or nonverbal means that the examinee's grasp of a language or a dialect is too deficient to proceed. Such is not the case when the evaluation occurs in written form. In the case of written tests, it is clearly essential that the examinee be able to read and comprehend what is written. Otherwise, the evaluation may be more about language or dialect competency than whatever the test purports to measure. Even when examiner and examinee speak the same language, miscommunication and consequential effects on test results may result owing to differences in dialect (Wolfram, 1971).

In the assessment of an individual whose proficiency in the English language is limited or nonexistent, some basic questions may need to be raised: What level of proficiency in English is required on the part of the testtaker, and does the testtaker have that proficiency? Can a meaningful assessment take place through a trained interpreter? Can an alternative and more appropriate assessment procedure be devised to meet the objectives of the assessment? In addition to linguistic barriers, the contents of tests from a particular culture are typically laden with items and material—some obvious, some very subtle—that draw heavily from that culture. Test performance may, at least in part, reflect not only whatever variables the test purports to measure but also one additional variable: the degree to which the testtaker has assimilated the culture.

**Nonverbal communication and behavior** Humans communicate not only through verbal means but also through nonverbal means. Facial expressions, finger and hand signs, and shifts in one's position in space may all convey messages. Of course, the messages conveyed by such body language may be different from culture to culture. In American culture, for example, one who fails to look another person in the eye when speaking may be viewed as deceitful or having something to hide. However, in other cultures, failure to make eye contact when speaking may be a sign of respect.

If you have ever gone on or conducted a job interview, you may have developed a firsthand appreciation of the value of nonverbal communication in an evaluative setting. Interviewees who show enthusiasm and interest have the edge over interviewees who appear to be drowsy or bored. In clinical settings, an experienced evaluator may develop hypotheses to be tested from the nonverbal behavior of the interviewee. For example, a person who is slouching, moving slowly, and exhibiting a sad facial expression may be depressed. Then again, such an individual may be experiencing physical

discomfort from any number of sources, such as a muscle spasm or an arthritis attack. It remains for the assessor to determine which hypothesis best accounts for the observed behavior.

Certain theories and systems in the mental health field go beyond more traditional interpretations of body language. For example, in **psychoanalysis**, a theory of personality and psychological treatment developed by Sigmund Freud, symbolic significance is assigned to many nonverbal acts. From a psychoanalytic perspective, an interviewee's fidgeting with a wedding band during an interview may be interpreted as a message regarding an unstable marriage. As evidenced by his thoughts on "the first chance actions" of a patient during a therapy session, Sigmund Freud believed he could tell much about motivation from nonverbal behavior:

The first . . . chance actions of the patient . . . will betray one of the governing complexes of the neurosis. . . . A young girl . . . hurriedly pulls the hem of her skirt over her exposed ankle; she has betrayed the kernel of what analysis will discover later; her narcissistic pride in her bodily beauty and her tendencies to exhibitionism. (Freud, 1913/1959, p. 359)

By the way, this quote from Freud is also useful in illustrating the influence of culture on diagnostic and therapeutic views. Freud lived in Victorian Vienna. In that time and in that place, sex was not a subject for public discussion. In many ways, Freud's views regarding a sexual basis for various thoughts and behaviors were a product of the sexually repressed culture in which he lived.

An example of a nonverbal behavior in which people differ is the speed at which they characteristically move to complete tasks. The overall pace of life in one geographic area, for example, may tend to be faster than in another. In a similar vein, differences in pace of life across cultures may enhance or detract from test scores on tests involving timed items (Gopaul-McNicol, 1993; Knapp, 1960). In a more general sense, Hoffman (1962) questioned the value of timed tests of ability, particularly those tests that employed multiple-choice items. He believed such tests relied too heavily on testtakers' quickness of response and as such discriminated against the individual who is characteristically a "deep, brooding thinker."

Culture exerts effects over many aspects of nonverbal behavior. For example, a child may present as noncommunicative and having only minimal language skills when verbally examined. This finding may be due to the fact that the child is from a culture where elders are revered and where children speak to adults only when they are spoken to—and then only in as short a phrase as possible. Clearly, it is incumbent upon test users to be knowledgeable about aspects of an assessee's culture that are relevant to the assessment. Dr. Diana D. Jeffery touched on such issues as she reminisced about past experiences (see this chapter's *Meet an Assessment Professional*).

#### JUST THINK . . .

Play the role of a therapist in the Freudian tradition and cite one example of public behavior that you believe may be telling about an individual's private motivation. Then, think about the significance *you* would attribute to "the first chance actions" of a patient or client as described in Freud's quotation.

#### JUST THINK . . .

What type of test is best suited for administration to people who are "deep, brooding thinkers"? How practical for group administration would such tests be?

**Standards of evaluation** Suppose that master chefs from more than a hundred nations entered a contest to discover the best chicken soup in the world. Who do you think would win? The answer to that question hinges on the evaluative standard to be employed. If the sole judge of the contest was the owner of a kosher delicatessen on the

## MEET AN ASSESSMENT PROFESSIONAL

### Meet Dr. Diana D. Jeffery

**W**hen I was a graduate student in the 1980s, my psychometrics professor at Yeshiva University, Joshua Fishman (he's in Wikipedia: Look him up!), said to our less-than-enthusiastic class, "If you learn this material it can be your bread and butter someday." He inspired me to read and re-read my psychometrics textbook and seek out supplemental readings. Learning "classical psychometrics" has opened numerous windows of opportunities in my professional life. . . .

[In one study exploring the reliability and factor structure of a particular instrument, we found that] *all* patients in the New Orleans sample responded to the item "I pray" at the top of the Likert scale which . . . caused us to throw out the item. . . . Was the community of New Orleans more religious than the participants from the other sites in Atlanta and San Francisco? [My mentor] discovered why during a site visit: The New Orleans interviewer was a nun!

. . . When the European Organization for Research and Treatment of Cancer (EORTC) group for research into QOL [Quality of Life] first met more than 25 years ago, there were many heated discussions about cultural differences in QOL. The Dutch, who took the lead on developing the test, argued for the inclusion of questions about bicycling and climbing flights of stairs. The French retorted that such activities were not a normal part of their everyday experience except, maybe, to go down a flight of stairs to get wine from their cellars. Where, they asked, were the questions on cheese? The Italians wanted many more questions about food, and the British



**Diana D. Jeffery, Ph.D., Program Director, National Cancer Institute, National Institutes of Health (NIH)**

maintained that mental health was getting short shrift . . .

. . . my advice to students of psychometrics is to learn this material so well that you wear out your textbook and have to buy another. Your knowledge of psychometrics may lead you to foreign countries, IRB [Institutional Review Board] membership, NIH Roadmap initiatives, or your own research niche. And, just maybe, it might become your bread and butter someday.

*Read more of what Dr. Jeffery had to say—her complete essay—at [www.mhhe.com/cohentesting7](http://www.mhhe.com/cohentesting7).*

Lower East Side of Manhattan, the entry that came closest to the "Jewish mother homemade" variety might well be declared the winner. However, other judges might have other standards and preferences. For example, soup connoisseurs from Arabic cultures might prefer chicken soup with fresh lemon juice in the recipe. Judges from India might be inclined to give their vote to a chicken soup flavored with curry and other exotic spices. For Japanese and Chinese judges, soy sauce might be viewed as an indispensable ingredient; any chicken soup prepared without this key ingredient might lose by

default. Ultimately, the judgment of which soup is best will probably be very much a matter of personal preference and the standard of evaluation employed.

Somewhat akin to judgments concerning the best chicken soup recipe, judgments related to certain psychological traits can also be culturally relative. For example, whether specific patterns of behavior are considered to be male- or female-appropriate will depend on the prevailing societal standards regarding masculinity and femininity. In some societies, for example, it is role-appropriate for women to fight wars and put food on the table while the men are occupied in more domestic activities. Whether specific patterns of behavior are considered to be psychopathological also depends on the prevailing societal standards. In Sudan, for example, there are tribes that live among cattle because they regard the animals as sacred. Judgments as to who might be the best employee, manager, or leader may differ as a function of culture, as might judgments regarding intelligence, wisdom, courage, and other psychological variables.

A challenge inherent in the assessment enterprise concerns tempering test- and assessment-related outcomes with good judgment regarding the cultural relativity of those outcomes. In practice, this means raising questions about the applicability of assessment-related findings to specific individuals. It therefore seems prudent to supplement questions such as “How intelligent is this person?” or “How assertive is this individual?” with other questions, such as: “How appropriate are the norms or other standards that will be used to make this evaluation?” “To what extent has the assessee been assimilated by the culture from which the test is drawn, and what influence might such assimilation (or lack of it) have on the test results?” “What research has been done on the test to support the applicability of findings with it for use in evaluating this particular assessee?” These are the types of questions that are being raised not only by responsible test users but also by courts of law.

### JUST THINK . . .

When considering tools of evaluation that purport to measure the trait of assertiveness, what are some culture-related considerations that should be kept in mind?

## Tests and Group Membership

Tests and other evaluative measures administered in vocational, education, counseling, and other settings leave little doubt that people differ from one another on an individual basis and also from group to group on a collective basis. What happens when groups systematically differ in terms of scores on a particular test? The answer, in a word, is *conflict*.

On the face of it, questions such as “What student is best qualified to be admitted to this school?” or “Which job candidate should get the job?” are rather straightforward. On the other hand, societal concerns about fairness both to individuals and to groups of individuals have made the answers to such questions matters of heated debate, if not lawsuits and civil disobedience. What happens when a person who happens to be a member of a particular group—cultural or otherwise—fails to obtain a desired outcome (such as attainment of employment or admission to a university) and it is observed that most other people from that same group have also failed to obtain the same prized outcome? What typically happens is that the criteria being used to judge attainment of the prized outcome become the subject of intense scrutiny, sometimes by a court or a legislature.

In vocational assessment, test users are sensitive to legal and ethical mandates concerning the use of tests with regard to hiring, firing, and related decision making. If a test is used to evaluate a candidate’s ability to do a job, one point of view is that the test should do just that—regardless of the group membership of the testtaker. According to

this view, scores on a test of job ability should be influenced only by job-related variables. That is, scores should not be affected by variables such as hair length, eye color, group membership, or any other variable extraneous to the ability to perform the job. Although this rather straightforward view of the role of tests in personnel selection may seem consistent with principles of equal opportunity, it has attracted charges of unfairness and claims of discrimination. Why?

Claims of test-related discrimination made against major test publishers may be best understood as evidence of the great complexity of the assessment enterprise rather than as a conspiracy to use tests to discriminate against individuals from certain groups. In vocational assessment, for example, conflicts may arise from disagreements about the criteria for performing a particular job. The potential for controversy looms over almost all selection criteria that an employer sets, regardless of whether the criteria are physical, educational, psychological, or experiential.

The critical question with regard to hiring, promotion, and other selection decisions in almost any work setting is: "What criteria must be met to do this job?" A state police department may require all applicants for the position of police officer to meet certain physical requirements, including a minimum height of 5 feet 4 inches. A person who is 5 feet 2 inches tall is barred from applying. Because such police force evaluation poli-

cies thus have the effect of systematically excluding members of a specific cultural group where the average height of adults is less than 5 feet 4 inches, the result may be a class-action lawsuit charging discrimination. Whether the police department's height requirement is reasonable and job related, and whether discrimination actually occurred, are complex questions that are usually left to be resolved in court. Compelling arguments may be presented on both

sides, as benevolent, fair-minded, knowledgeable, and well-intentioned people may have honest differences about the necessity of the prevailing height requirement for the job of police officer.

Beyond the variable of height, it would seem that variables such as appearance and religion should have little to do with what job one is qualified to perform. However, it is precisely such factors that keep some group members from entry into many jobs and careers. Consider in this context observant Jews. Their appearance and dress is not mainstream. The food they eat must be kosher. They are unable to work or travel on weekends. Given the established selection criteria for many positions in corporate America, candidates who are members of the group known as observant Jews are effectively excluded regardless of their ability to perform the work (Korman, 1988; Mael, 1991; Zweigenhaft, 1984).

General differences among groups of people also extend to psychological attributes such as measured intelligence. Unfortunately, the mere suggestion that such differences in psychological variables exist arouses skepticism if not charges of discrimination, bias, or worse. This is especially true when the observed group differences are deemed responsible for blocking one or another group from employment or educational opportunities.

If systematic differences related to group membership were found to exist on job ability test scores, then what, if anything, should be done? One view is that nothing needs to be done. According to this view, the test was designed to measure job ability, and it does what it was designed to do. In support of this view is evidence suggesting that group differences in scores on professionally developed tests do reflect differences in real-world performance (Gottfredson, 2000; Halpern, 2000; Hartigan & Wigdor, 1989; Kubiszyn et al., 2000; Neisser et al., 1996; Schmidt, 1988; Schmidt & Hunter, 1992).

#### JUST THINK . . .

Devise your own version of a fair and equitable process to determine the minimum required height, if any, for police officers in your community.

A contrasting view is that efforts should be made to “level the playing field” between groups of people. The term **affirmative action** refers to voluntary and mandatory efforts undertaken by federal, state, and local governments, private employers, and schools to combat discrimination and to promote equal opportunity in education and employment for all (American Psychological Association, 1996a, p. 2). Affirmative action seeks to create equal opportunity actively, not passively; inherent in it is the view that “policies that appear to be neutral with regard to ethnicity or gender can operate in ways that advantage individuals from one group over individuals from another group” (Crosby et al., 2003, p. 95).

In assessment, one way of implementing affirmative action is by altering test scoring procedures according to set guidelines. For example, an individual’s score on a test could be revised according to the individual’s group membership (McNemar, 1975). While proponents of such remedies see them as necessary to address past inequities, others condemn such manipulation of test scores as introducing “inequity in equity” (Benbow & Stanley, 1996).

As sincerely committed as they may be to principles of egalitarianism and fair play, test developers and test users must ultimately look to society at large—and, more specifically, to laws, administrative regulations, and other rules and professional codes of conduct—for guidance in the use of tests and test scores.

**Psychology, tests, and public policy** Few people would object to using psychological tests in academic and applied contexts that obviously benefit human welfare. Then again, few people are aware of the everyday use of psychological tests in such ways. More typically, members of the general public become acquainted with the use of psychological tests in high-profile contexts, such as when an individual or a group has a great deal to gain or to lose as a result of a test score. In such situations, tests and other tools of assessment are portrayed as instruments that can have a momentous and immediate impact on one’s life. In such situations, tests may be perceived by the everyday person as tools used to deny people things they very much want or need. Denial of educational advancement, job opportunity, parole, or custody are some of the more threatening consequences that the public may associate with psychological tests and assessment procedures.

Members of the public call upon government policy makers to protect them from perceived threats. Legislators pass laws, administrative agencies make regulations, judges hand down rulings, and citizens call for referenda to reflect and enforce prevailing public policy or to modify it. In the section that follows, we broaden our view of the assessment enterprise beyond the concerns of the profession to include also the concerns of the public.

## Legal and Ethical Considerations

**Laws** are rules that individuals must obey for the good of the society as a whole—or rules thought to be for the good of society as a whole. Some laws are and have been relatively uncontroversial. For example, the law that mandates driving on the right side of the road has not been a subject of debate, a source of emotional soul-searching, or a stimulus to civil disobedience. For safety and the common good, most people are

### JUST THINK . . .

What are your thoughts on the manipulation of test scores as a function of group membership to advance certain social goals? Should membership in a particular cultural group trigger an automatic increase (or decrease) in test scores?

willing to relinquish their freedom to drive all over the road. But what about laws pertaining to abortion? To busing? To capital punishment? To euthanasia? To deprogramming of religious cult members? To affirmative action in employment? Exactly how laws regulating matters such as these should be written and interpreted are issues of heated controversy, as are some of the laws that affect the way that psychological testing and assessment are conducted.

Whereas a body of laws is a body of rules, a body of **ethics** is a body of principles of right, proper, or good conduct. Thus, for example, an ethic of the Old West was “Never shoot ‘em in the back.” Two well-known principles subscribed to by seafarers are “Women and children leave first in an emergency” and “A captain goes down with his ship.”<sup>3</sup> The ethics of journalism dictate that reporters present all sides of a controversial issue. A principle of ethical research is that the researcher should never fudge data; all data must be reported accurately.

### JUST THINK . . .

List five ethical guidelines that you think should govern the professional behavior of psychologists involved in psychological testing and assessment.

To the extent that a **code of professional ethics** is recognized and accepted by members of a profession, it defines the standard of care expected of members of that profession. Members of the public and members of the profession have in recent times been on different sides of the fence with respect to issues of ethics and law. Let’s review how and why this has been the case.

## The Concerns of the Public

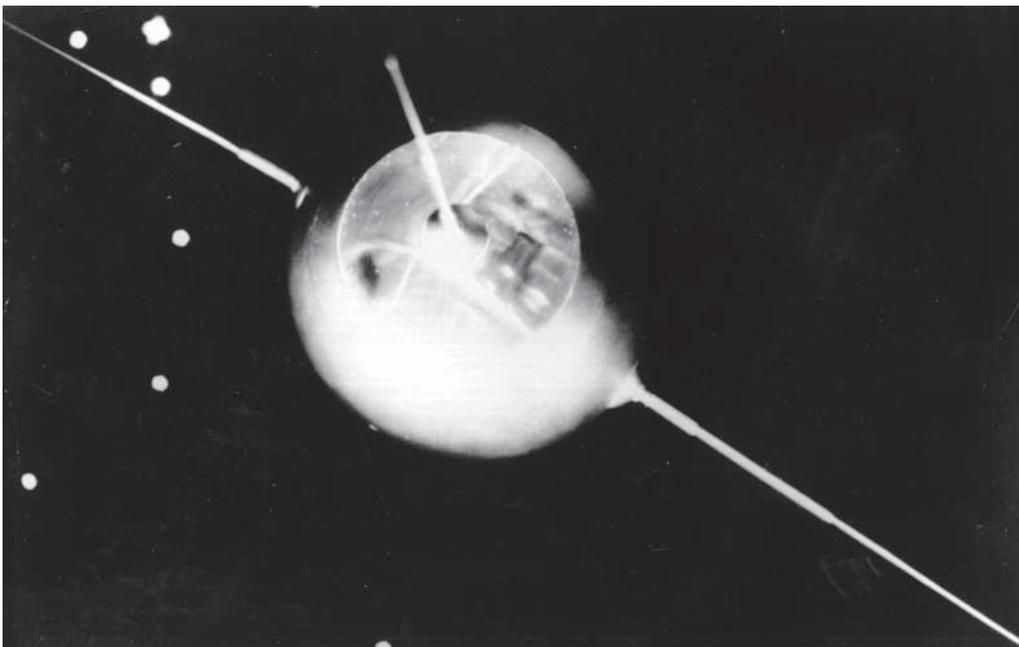
The assessment enterprise has never been very well understood by the public. Even today, it is unfortunate that we may hear statements symptomatic of misunderstanding with regard to tests (e.g., “The only thing tests measure is the ability to take tests”). Possible consequences of public misunderstanding include fear, anger, legislation, litigation, and administrative regulations.

Widespread concern about the use of psychological tests first became evident in the aftermath of World War I. At that time, various professionals (as well as nonprofessionals) sought to adapt group tests developed by the military for civilian use in schools and industry. Reflecting growing public discomfort with the burgeoning assessment industry were popular magazine articles featuring stories with titles such as “The Abuse of Tests” (see Haney, 1981). Less well known were voices of reason that offered constructive ways to correct what was wrong with assessment practices.

Anticipating the present-day *Standards*, Ruch (1925), a measurement specialist, proposed a number of standards for tests and guidelines for test development. He also wrote of “the urgent need for a fact-finding organization which will undertake impartial, experimental, and statistical evaluations of tests” (Ruch, 1933). History records that one team of measurement experts even took on the (overly) ambitious task of attempting to rank all published tests designed for use in educational settings. The result was a pioneering book (Kelley, 1927) that provided test users with information needed to compare the merits of published tests. However, given the pace at which test instruments were being published, this resource required regular updating. And so, Oscar Buros was not the first measurement professional to undertake a comprehensive testing of the tests. He was, however, the most tenacious in updating and revising the information.

---

3. We leave the question of what to do when the captain of the ship is a woman to a volume dedicated to an in-depth exploration of seafaring ethics.



**Figure 2-4**  
**The Launch of a Satellite . . . and Renewed Interest in Testing**

*On October 4, 1957, scientists in the country then known as the Union of Soviet Socialist Republics launched a satellite (they called it Sputnik) into space. The event was greeted with surprise if not shock by most Americans. The prospect of a cold war enemy having a satellite orbiting Earth 24 hours a day was most unsettling. The launch caused widespread concern about the ability of the United States to compete in the new frontier of space. More emphasis would have to be placed on education, particularly in subjects such as math, science, engineering, and physics. And greater efforts would have to be made to identify the gifted children who would one day apply such knowledge in the race to space.*

The widespread military testing during the 1940s as a result of World War II did not attract as much popular attention as the testing undertaken during World War I. Rather, an event that took place in a faraway land had a far more momentous effect on testing in the United States: the launching of a satellite into space (see Figure 2-4). About a year after the Soviet Union's launch of *Sputnik*, Congress passed the National Defense Education Act, which provided federal money to local schools for the purpose of ability and aptitude testing to identify gifted and academically talented students. This event triggered a proliferation of large-scale testing programs in the schools. At the same time, the use of ability tests and personality tests for personnel selection increased in government, the military, and business. The wide and growing use of tests led to renewed public concern, reflected in magazine articles such as "Testing: Can Everyone Be Pigeonholed?" (*Newsweek*, July 20, 1959) and "What the Tests Do Not Test" (*New York Times Magazine*, October 2, 1960). The upshot of such concern was congressional hearings on the subject of testing (Amrine, 1965).

The fires of public concern about testing were again fanned in 1969 when widespread media attention was given to the publication of an article, in the prestigious *Harvard Educational Review*, entitled "How Much Can We Boost IQ and Scholastic Achievement?" Its author, Arthur Jensen, argued that "genetic factors are strongly implicated in the average Negro-white intelligence difference" (1969, p. 82). What followed

was an outpouring of public and professional attention to nature-versus-nurture issues in addition to widespread skepticism about what intelligence tests were really measuring. By 1972, the U.S. Select Committee on Equal Education Opportunity was preparing for hearings on the matter. However, according to Haney (1981), the hearings “were canceled because they promised to be too controversial” (p. 1026).

The extent of public concern about psychological assessment is reflected in the extensive involvement of the government in many aspects of the assessment process in recent decades. Assessment has been affected in numerous and important ways by activities of the legislative, executive, and judicial branches of federal and state governments. A sampling of some landmark legislation and litigation is presented in Table 2–1.

**Table 2–1**  
**Some Significant Legislation and Litigation**

Legislation	Significance
Americans with Disabilities Act of 1990	Employment testing materials and procedures must be essential to the job and not discriminate against persons with handicaps.
Civil Rights Act of 1964 (amended in 1991), also known as the Equal Opportunity Employment Act	It is an unlawful employment practice to adjust the scores of, use different cutoff scores for, or otherwise alter the results of, employment-related tests on the basis of race, religion, sex, or national origin.
Family Education Rights and Privacy Act (1974)	Mandated that parents and eligible students be given access to school records. Also granted right to challenge findings in records by a hearing.
Health Insurance Portability and Accountability Act of 1996 (HIPAA)	Provided for federal privacy standards that limit the way that health care providers and others can use patients' personal information.
Education for All Handicapped Children (PL 94-142) (1975 and then amended several times thereafter, including IDEA of 1997 and 2004)	Mandated screening of children with suspected mental or physical handicaps. Once identified, individual child must be evaluated by a professional team qualified to determine that child's special educational needs. Child must be reevaluated periodically. Amended in 1986 to extend disability-related protections downward to infants and toddlers.
Individuals with Disabilities Education Act (IDEA) Amendments of 1997 (PL 105-17)	Deterred inappropriate placement in special education programs due to cultural differences. Encouraged accommodation of existing test instruments and other alternate means of assessment for the purpose of gauging the progress of special education students as measured by state- and district-wide assessments.
The No Child Left Behind (NCLB) Act of 2001	Known as the NCLB, the reauthorization of the Elementary and Secondary Education Act of 2001 was designed to “close the achievement gaps between minority and nonminority students and between disadvantaged children and their more advantaged peers” by, among other things, setting strict standards for school accountability and establishing periodic assessments to gauge the progress of school districts in improving academic achievement. The “battle cry” driving this legislation was “Demographics are not destiny!”
<b>Litigation</b>	
<i>Hobson v. Hanson</i> (1967)	Supreme Court ruled that ability tests developed on Whites could not lawfully be used to track Black students in the school system. To do so could result in resegregation of desegregated schools.
<i>Tarasoff v. Regents of the University of California</i> (1974)	Therapists (and presumably psychological assessors) must reveal privileged information if a third party is endangered. In the words of the Court, “Protective privilege ends where the public peril begins.”
<i>Larry P. v. Riles</i> (1979 and reaffirmed by the same judge in 1986)	California judge ruled that the use of intelligence tests to place Black children in special classes had a discriminatory impact because the tests were “racially and culturally biased.”
<i>Debra P. v. Turlington</i> (1981)	Federal court ruled that minimum competency testing in Florida was unconstitutional because it perpetuated the effects of past discrimination.
<i>Griggs v. Duke Power Company</i> (1971)	Black employees brought suit against a private company for discriminatory hiring practices. The Supreme Court found problems with “broad and general testing devices” and ruled that tests must “fairly measure the knowledge or skills required by a particular job.”
<i>Albemarle Paper Company v. Moody</i> (1976)	An industrial psychologist at a paper mill found that scores on a general ability test predicted measures of job performance. However, as a group, Whites scored better than Blacks on the test. The U.S. District Court found the use of the test to be sufficiently job related. An appeals court did not. It ruled that discrimination had occurred, however unintended.

<i>Regents of the University of California v. Bakke</i> (1978)	When Alan Bakke learned that his test scores were higher than those of some minority students who had gained admission to the University of California, Davis, medical school, he sued. A highly divided Supreme Court agreed that Bakke should be admitted, but it did not preclude the use of diversity considerations in admission decisions.
<i>Allen v. District of Columbia</i> (1993)	Blacks scored lower than Whites on a city fire department promotion test based on specific aspects of fire-fighting. The court found in favor of the fire department, ruling that "the promotional examination . . . was a valid measure of the abilities and probable future success of those individuals taking the test. . . ."
<i>Adarand Constructors, Inc. v. Peña et al.</i> (1995)	A construction firm competing for a federal contract brought suit against the federal government for losing a bid to a minority-controlled competitor which the government had retained instead in the interest of affirmative action. The Supreme Court, in a close (5–4) decision, found in favor of the plaintiff, ruling that the government's affirmative action policy violated the equal protection clause of the 14th Amendment. The Court ruled, "Government may treat people differently because of their race only for the most compelling reasons."
<i>Jaffee v. Redmond</i> (1996)	Communication between a psychotherapist and a patient (and presumably a psychological assessor and a client) is privileged in federal courts.
<i>Grutter v. Bollinger</i> (2003)	In a highly divided decision, the Supreme Court approved the use of race in admissions decisions on a time-limited basis to further the educational benefits that flow from a diverse student body ( <i>see Close-Up</i> ).

---

**Legislation** Although the legislation summarized in Table 2–1 was enacted at the federal level, states also have passed legislation that affects the assessment enterprise. In the 1970s, numerous states enacted **minimum competency testing programs**: formal testing programs designed to be used in decisions regarding various aspects of students' education. The data from such programs was used in decision making about grade promotions, awarding of diplomas, and identification of areas for remedial instruction. These laws grew out of grassroots support for the idea that high-school graduates should have, at the very least, "minimal competencies" in areas such as reading, writing, and arithmetic.

**Truth-in-testing legislation** was also passed at the state level beginning in the 1980s. The primary objective of these laws was to provide testtakers with a means of learning the criteria by which they are being judged. To meet that objective, some laws mandate the disclosure of answers to postsecondary and professional school admissions tests within 30 days of the publication of test scores. Some laws require that information relevant to a test's development and technical soundness be kept on file. Some truth-in-testing laws require providing descriptions of (1) the test's purpose and its subject matter, (2) the knowledge and skills the test purports to measure, (3) procedures for ensuring accuracy in scoring, (4) procedures for notifying testtakers of errors in scoring, and (5) procedures for ensuring the testtaker's confidentiality. Truth-in-testing laws create special difficulties for test developers and publishers, who argue that it is essential for them to keep the test items secret. They note that there may be a limited item pool for some tests and that the cost of developing an entirely new set of items for each succeeding administration of a test is prohibitive.

Some laws mandate the involvement of the executive branch of government in their application. For example, Title VII of the Civil Rights Act of 1964 created the Equal Employment Opportunity Commission (EEOC) to enforce the act. The EEOC has published sets of guidelines concerning standards to be met in constructing and using employment tests. In 1978, the EEOC, the Civil Service Commission, the Department of Labor, and the Justice Department jointly published the *Uniform Guidelines on Employee Selection Procedures*. Here is a sample guideline:

The use of any test which adversely affects hiring, promotion, transfer or any other employment or membership opportunity of classes protected by Title VII constitutes discrimination unless (a) the test has been validated and evidences a high degree of utility as hereinafter described, and (b) the person giving or acting upon the results of the particular test can demonstrate that alternative suitable hiring, transfer or promotion procedures are unavailable for . . . use.

Note that here the definition of discrimination as exclusionary coexists with the proviso that a valid test evidencing “a high degree of utility” (among other criteria) will not be considered discriminatory. Generally, however, the public has been quick to label a test as unfair and discriminatory regardless of its utility. As a consequence, a great public demand for proportionality by group membership in hiring and college admissions now coexists with a great lack of proportionality by group membership in skills. Gottfredson (2000) noted that, although selection standards can often be improved, the manipulation of such standards “will produce only lasting frustration, not enduring solutions.” She recommended that enduring solutions be sought by addressing the problem related to gaps in skills between groups and argued against addressing the problem by lowering hiring and admission standards or by legislation designed to make hiring and admissions decisions a matter of group quotas. Yet it is in this latter direction that the tide seems to be turning, at least according to recent legislation and court decisions.

In Texas, state law was enacted mandating that the top 10% of graduating seniors from all Texas high schools be admitted to a state university regardless of SAT scores. This means that, regardless of the quality of education in any particular Texas high school, a senior in the top 10% of the graduating class is guaranteed college admission regardless of how he or she might score on a nationally administered measure. There have been reports that in some Texas high schools as many as 25% of the students are in the top 10% of their class (Kronholz, 1998). In California, the use of skills tests in the public sector decreased following the passage of Proposition 209, which banned racial preferences (Rosen, 1998). One consequence has been the deemphasis on the Law School Admissions Test (LSAT) as a criterion for being accepted by the University of California, Berkeley, law school. Additionally, the law school stopped weighing grade point averages from undergraduate schools in their admission criteria, so that “a 4.0 from California State is now worth as much as a 4.0 from Harvard” (Rosen, 1998, p. 62).

Gottfredson (2000) makes the point that those who advocate reversal of achievement standards obtain “nothing of lasting value by eliminating valid tests.” For her, lowering standards amounts to hindering progress “while providing only the illusion of progress.” Rather than reversing achievement standards, society is best served by action to reverse other trends with deleterious effects (such as trends in family structure). In the face of consistent gaps between members of various groups, Gottfredson emphasized the need for skills training, not a lowering of achievement standards or an unfounded attack on tests.

State and federal legislatures, executive bodies, and courts have been involved in many aspects of testing and assessment. There has been little consensus about whether validated tests on which there are racial differences can be used to assist with employment-related decisions. Courts have also been grappling with the role of diversity in criteria for admission to colleges, universities, and professional schools. For example, in 2003, the question before the Supreme Court in the case of *Grutter v. Bollinger* was “whether diversity is a compelling interest that can justify the narrowly tailored use of race in selecting applicants for admission to public universities.” One of the questions to be decided in that case was whether or not the University of Michigan Law School

was using a **quota system**, a selection procedure whereby a fixed number or percentage of applicants from certain backgrounds were selected.<sup>4</sup>

**Litigation** Rules governing citizens' behavior stem not only from legislatures but also from interpretations of existing law in the form of decisions handed down by courts. This is why legal disputes and resolution of criminal and administrative matters handled by courts—referred to here simply as litigation—can impact our daily lives. Examples of some court cases that have affected the assessment enterprise were presented in Table 2–1 under the “Litigation” heading. It is also true that litigation can result in bringing an important and timely matter to the attention of legislators, thus serving as a stimulus to the creation of new legislation. This is exactly what happened in the cases of *PARC v. Commonwealth of Pennsylvania* (1971) and *Mills v. Board of Education of District of Columbia* (1972). In the PARC case, the Pennsylvania Association for Retarded Children brought suit because mentally retarded children in that state had been denied access to public education. In *Mills*, a similar lawsuit was filed on behalf of children with behavioral, emotional, and learning impairments. Taken together, these two cases had the effect of jump-starting similar litigation in several other jurisdictions and alerting Congress to the need for federal law to ensure appropriate educational opportunities for children with disabilities.

Litigation has sometimes been referred to as “judge-made law” because it typically comes in the form of a ruling by a court. And while it is true that judges do, in essence, create law by their rulings, these rulings are seldom made in a vacuum. Rather, judges typically rely on prior rulings and on other people—most notably, expert witnesses—to assist in their judgments. A psychologist acting as an expert witness in criminal litigation may testify on matters such as the competence of a defendant to stand trial, the competence of a witness to give testimony, or the sanity of a defendant entering a plea of “not guilty by reason of insanity.” A psychologist acting as an expert witness in a civil matter could conceivably offer opinions on many different types of issues ranging from the parenting skills of a parent in a divorce case to the capabilities of a factory worker prior to sustaining a head injury on the job. In a malpractice case, an expert witness might testify about how reasonable and professional the actions taken by a fellow psychologist were and whether any reasonable and prudent practitioner would have engaged in the same or similar actions (Cohen, 1979).

The number of different issues on which expert witnesses can be called upon to give testimony is as varied as the number of different issues that reach courtrooms for resolution. And so, some important questions arise with respect to expert witnesses. For example: Who is qualified to be an expert witness? How much weight should be given to the testimony of an expert witness? Questions such as these have themselves been the subject of litigation.

A landmark case heard by the Supreme Court in June 1993 has implications regarding the admissibility of expert testimony in court. The case was *Daubert v. Merrell Dow Pharmaceuticals*. The origins of this case can be traced to Mrs. Daubert's use of the prescription drug Bendectin to relieve nausea during pregnancy. The plaintiffs sued the manufacturer of this drug, Merrell Dow Pharmaceuticals, when their children were born with birth defects. They claimed that Mrs. Daubert's use of Bendectin had caused their children's birth defects.

Attorneys for the Dauberts were armed with research that they claimed would prove that Bendectin causes birth defects. However, the trial judge ruled that the research failed

---

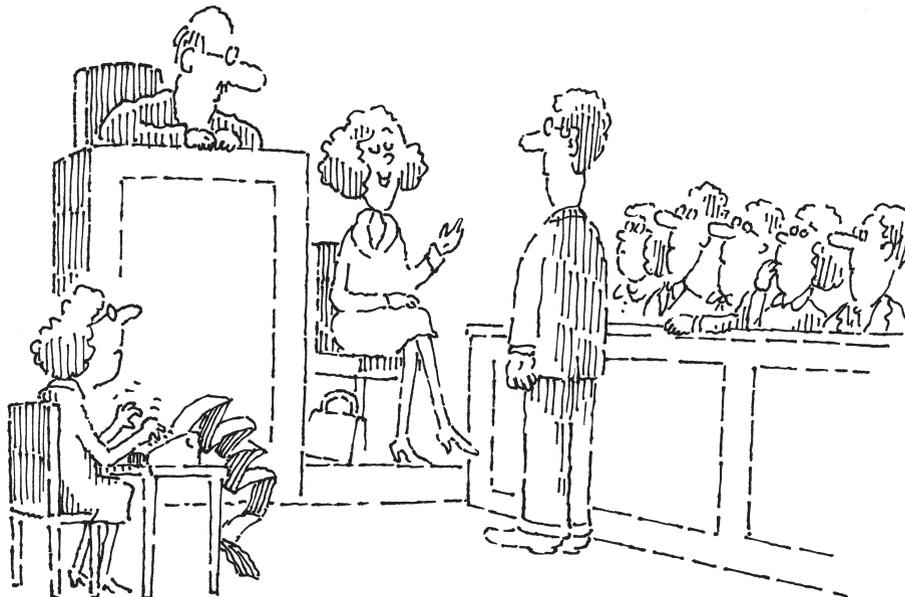
4. A detailed account of *Grutter v. Bollinger* is presented in the context of one of the exercises in the companion workbook to this text.

to meet the criteria for admissibility. In part because the evidence the Dauberts wished to present was not deemed admissible, the trial judge ruled against the Dauberts.

The Dauberts appealed to the next higher court. That court, too, ruled against them and in favor of Merrell Dow. Once again, the plaintiffs appealed, this time to the Supreme Court of the United States. A question before the Supreme Court was whether the judge in the original trial had acted properly by not allowing the plaintiffs' research to be admitted into evidence. To understand whether or not the trial judge acted properly, it is important to understand (1) a ruling that was made in the 1923 case of *Frye v. the United States* and (2) a law subsequently passed by Congress, Rule 702 in the *Federal Rules of Evidence* (1975).

In *Frye*, the Court held that scientific research is admissible as evidence when the research study or method enjoys general acceptance. General acceptance could typically be established by the testimony of other experts and by reference to publications in peer-reviewed journals. In short, if an expert claimed something that most other experts in the same field would agree with then, under *Frye*, the testimony could be admitted into evidence. Rule 702 changed that by allowing more experts to testify regarding the admissibility of the original expert testimony. In addition to the expert testimony or research that enjoyed general acceptance in the field, other experts could now testify about the admissibility of research or research methods. An expert might offer an opinion to a jury concerning the acceptability of a research study or method regardless of whether that opinion represented the opinions of other experts. Rule 702 was enacted to assist juries in their fact-finding by helping them to understand the issues involved.

Presenting their case before the Supreme Court, the attorneys for the Dauberts argued that Rule 702 had wrongly been ignored by the trial judge. The attorneys for the defendant, Merrell Dow Pharmaceuticals, countered that the trial judge had ruled appropriately. They argued that high standards of evidence admissibility were necessary to protect juries from "scientific shamans who, in the guise of their purported



...the Ancients measured facial beauty by the millihelen, a unit equal to that necessary to launch one ship....

expertise, are willing to testify to virtually any conclusion to suit the needs of the litigant with resources sufficient to pay their retainer.”

The Supreme Court ruled that the *Daubert* case be retried and that the trial judge should be given wide discretion in deciding what does and does not qualify as scientific evidence. In effect, federal judges were charged with a *gatekeeping* function with respect to what expert testimony is or is not admitted into evidence. The *Daubert* ruling superseded the long-standing policy, set forth in *Frye*, of admitting into evidence only scientific testimony that had won general acceptance in the scientific community. Opposing expert testimony, whether or not such testimony had won general acceptance in the scientific community, was to be admissible.

In *Daubert*, the Supreme Court viewed factors such as general acceptance in the scientific community or publication in a peer-reviewed journal as only some of many possible factors for judges to consider. Other factors judges might consider included the extent to which a theory or technique had been tested and the extent to which the theory or technique might be subject to error. In essence, the Supreme Court’s ruling in *Daubert* gave trial judges a great deal of leeway in deciding what juries could and could not hear.

Subsequent to *Daubert*, the Supreme Court has ruled on several other cases that in one way or another clarify or slightly modify its position in *Daubert*. For example, in the case of *General Electric Co. v. Joiner* (1997), the Court emphasized that the trial court had a duty to exclude unreliable expert testimony as evidence. In the case of *Kumho Tire Company Ltd. v. Carmichael* (1999), the Supreme Court expanded the principles expounded in *Daubert* to include the testimony of *all* experts, whether or not the experts claimed scientific research as a basis for their testimony. Thus, for example, a psychologist’s testimony based on personal experience in independent practice (rather than findings from a formal research study) may be admitted into evidence if the trial judge so chooses (Mark, 1999).

Whether or not *Frye* or *Daubert* will be relied on by the court depends on the individual jurisdiction in which a legal proceeding occurs. Some jurisdictions still rely on the *Frye* standard when it comes to admitting expert testimony, and some subscribe to *Daubert*. The implications of *Daubert* for psychologists and others who might have occasion to provide expert testimony in a trial are wide-ranging (Ewing & McCann, 2006). More specifically, discussions of the implications of *Daubert* for psychological experts can be found in cases involving mental capacity (Frolik, 1999; Poythress, 2004), claims of emotional distress (McLearn et al., 2004), personnel decisions (Landy, 2007), child custody and termination of parental rights (Bogacki & Weiss, 2007; Gould, 2006; Krauss & Sales, 1999), and numerous other matters (Grove & Barden, 1999; Lipton, 1999; Mossman, 2003; Posthuma et al., 2002; Saldanha, 2005; Saxe & Ben-Shakhar, 1999; Slobogin, 1999; Stern, 2001; Tenopyr, 1999).

## *The Concerns of the Profession*

As early as 1895, the infant American Psychological Association (APA) formed its first committee on mental measurement. The committee was charged with investigating various aspects of the relatively new practice of testing. Another APA committee on measurements was formed in 1906 to further study various testing-related issues and problems. In 1916 and again in 1921, symposia dealing with various issues surrounding the expanding uses of tests were sponsored (*Mentality Tests*, 1916; *Intelligence and Its Measurement*, 1921). In 1954, APA published its *Technical Recommendations for Psychological Tests and Diagnostic Tests*, a document that set forth testing standards and technical recommendations. The following year, another professional organization, the National Educational Association (working in collaboration with the National Council on

Measurements Used in Education—now known as the National Council on Measurement) published its *Technical Recommendations for Achievement Tests*. Collaboration between these professional organizations led to the development of rather detailed testing standards and guidelines that would be periodically updated in future years.

### JUST THINK . . .

Who should be privy to test data? Who should be able to purchase psychological test materials? Who is qualified to administer, score, and interpret psychological tests? What level of expertise in psychometrics qualifies someone to administer which types of test?

The APA and related professional organizations in the United States have made available numerous reference works and publications designed to delineate ethical, sound practice in the field of psychological testing and assessment.<sup>5</sup> Along the way, these professional organizations have tackled a variety of thorny questions, such as the questions cited in the next *Just Think*.

**Test-user qualifications** Should just anyone be allowed to purchase and use psychological test materials? If not, then who should be permitted to use psychological tests?

As early as 1950, an APA Committee on Ethical Standards for Psychology published a report called *Ethical Standards for the Distribution of Psychological Tests and Diagnostic Aids*. This report defined three levels of tests in terms of the degree to which the test's use required knowledge of testing and psychology.

*Level A:* Tests or aids that can adequately be administered, scored, and interpreted with the aid of the manual and a general orientation to the kind of institution or organization in which one is working (for instance, achievement or proficiency tests).

*Level B:* Tests or aids that require some technical knowledge of test construction and use and of supporting psychological and educational fields such as statistics, individual differences, psychology of adjustment, personnel psychology, and guidance (e.g., aptitude tests and adjustment inventories applicable to normal populations).

*Level C:* Tests and aids that require substantial understanding of testing and supporting psychological fields together with supervised experience in the use of these devices (for instance, projective tests, individual mental tests).

The report included descriptions of the general levels of training corresponding to each of the three levels of tests. Although many test publishers continue to use this three-level classification, some do not. In general, professional standards promulgated by APA (American Educational Research Association et al., 1999), the National Association of School Psychologists (2000), and other professional organizations state that psychological tests should be used only by qualified persons. Furthermore, there is an ethical mandate to take reasonable steps to prevent the misuse of the tests and the information they provide. The obligations of professionals to testtakers are set forth in a document called the *Code of Fair Testing Practices in Education*. Jointly authored and/or sponsored by the Joint Committee of Testing Practices (a coalition of APA, AERA, NCME, the American Association for Measurement and Evaluation in Counseling and Development, and the American Speech-Language-Hearing Association), this document presents standards for educational test developers in four areas: (1) developing/selecting tests, (2) interpreting scores, (3) striving for fairness, and (4) informing testtakers.

Beyond promoting high standards in testing and assessment among professionals, APA has initiated or assisted in litigation to limit the use of psychological tests to

---

5. Unfortunately, although organizations in many other countries have verbalized concern about ethics and standards in testing and assessment, relatively few organizations have taken meaningful and effective action in this regard (Leach & Oakland, 2007).

qualified personnel. Skeptics label such measurement-related legal action as a kind of jockeying for turf, done solely for financial gain. A more charitable and perhaps more realistic view is that such actions benefit society at large. It is essential to the survival of the assessment enterprise that certain assessments be conducted by people qualified to conduct them by virtue of their education, training, and experience.

A psychologist licensing law designed to serve as a model for state legislatures has been available from APA since 1987. The law contains no definition of psychological testing. In the interest of the public, the profession of psychology, and other professions that employ psychological tests, it may now be time for that model legislation to be rewritten—with terms such as *psychological testing* and *psychological assessment* clearly defined and differentiated. Terms such as *test-user qualifications* and *psychological assessor qualifications* must also be clearly defined and differentiated. It seems that part of the problem surrounding legal conflicts regarding psychological test usage stems from confusion of the terms *psychological testing* and *psychological assessment*. People who are not considered professionals by society may be qualified to use psychological tests (psychological testers). However, these same people may not be qualified to engage in psychological assessment. As we argued in the previous chapter, psychological assessment requires certain skills, talents, expertise, and training in psychology and measurement over and above that required to engage in psychological testing. In the past, psychologists have been lax in differentiating psychological testing from psychological assessment. However, continued laxity may prove to be a costly indulgence, given current legislative and judicial trends.

**Testing people with disabilities** Difficulties analogous to those concerning testtakers from linguistic and cultural minorities are present when testing people with disabling conditions. Specifically, these difficulties may include (1) transforming the test into a form that can be taken by the testtaker, (2) transforming the responses of the testtaker so that they are scorable, and (3) meaningfully interpreting the test data.

The nature of the transformation of the test into a form ready for administration to the individual with disabling conditions will, of course, depend on the nature of the disability. Then, too, some test stimuli do not translate easily. For example, if a critical aspect of a test item contains artwork to be analyzed, there may be no meaningful way to translate this item for use with testtakers who are blind. With respect to any test converted for use with a population for which the test was not originally intended, choices must inevitably be made regarding exactly how the test materials will be modified, what standards of evaluation will be applied, and how the results will be interpreted. Professional assessors do not always agree on the answers to such questions.

Another issue on which there is little consensus among professional assessors concerns a request by a terminally ill individual for aid in dying. Because such a request may only be granted contingent on the findings of a psychological evaluation, life or death literally hangs in the balance of such assessments. Presently, only Oregon has a law on the books dealing with this complex scenario. However, if other states adopt similar legislation, such situations will no doubt become more common, and many more psychological assessors will be a part of them. Some ethical and related issues surrounding this phenomenon are discussed in greater detail in this chapter's *Everyday Psychometrics*.

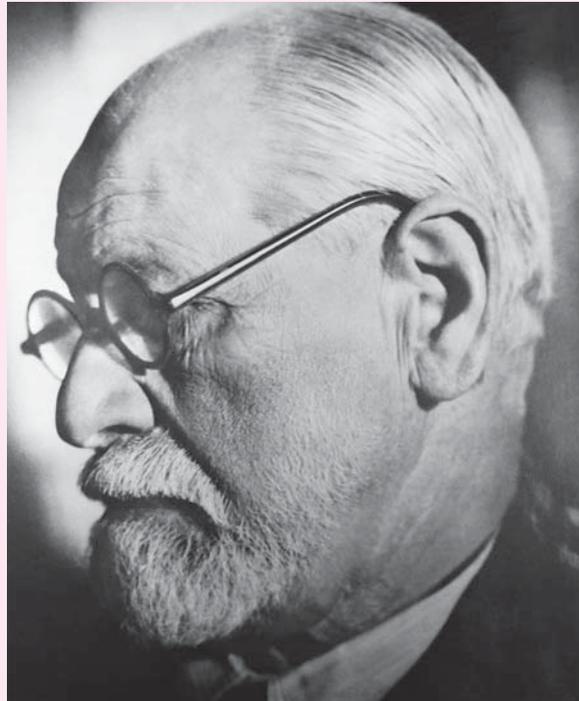
**Computerized test administration, scoring, and interpretation** Computer-assisted psychological assessment (CAPA) has become more the norm than the exception. An ever-growing number of psychological tests can be purchased on disc or even administered and scored online. In many respects, the relative simplicity, convenience, and range of

## Life-or-Death Psychological Assessment

The state of Oregon has the distinction—dubious to some people, depending on one’s values—of having enacted the nation’s first aid-in-dying law. Oregon’s Death with Dignity Act (ODDA) provides that a patient with a medical condition thought to give that patient six months or less to live may end his or her own life by voluntarily requesting a lethal dose of medication. The law requires that two physicians corroborate the terminal diagnosis and stipulates that either may request a psychological evaluation of the patient by a state-licensed psychologist or psychiatrist in order to ensure that the patient is competent to make the life-ending decision and to rule out impaired judgment due to psychiatric disorder. Aid-in-dying will be denied to persons “suffering from a psychiatric or psychological disorder, or depression causing impaired judgement” (ODDA, 1997).

The ODDA was hotly debated prior to its passage by referendum, and it remains controversial today. Critics of the law question whether suicide is ever a rational choice under any circumstances, and they fear that state-condoned aid in dying will serve to destigmatize suicide in general (Callahan, 1994; see also Richman, 1988). It is argued that the first duty of health and mental health professionals is to do no harm (Jennings, 1991). Some fear that professionals willing to testify to almost anything (so-called **hired guns**) will corrupt the process and accommodate those who can pay their fees with any professional opinion desired. Critics also point with concern to the experience of the Dutch death-with-dignity legislation. In the Netherlands, relatively few individuals requesting physician-assisted suicide are referred for psychological assessment. Further, the highest court of that land ruled that “in rare cases, physician-assisted suicide is possible even for individuals suffering only from mental problems rather than from physical illnesses” (Abeles & Barlev, 1999, p. 233). On moral and religious grounds, it has been argued that death should be viewed as the sole province of divine, not human, intervention.

Supporters of death-with-dignity legislation argue that life-sustaining equipment and methods can extend life beyond a time when it is meaningful and that the first obligation of health and mental health professionals is to relieve suffering (Latimer, 1991; Quill et al., 1992; Weir, 1992). Additionally, they may point to the dogged determination of people intent on dying and to stories of how many terminally ill people have struggled to end their lives using all kinds of



**Sigmund Freud (1856–1939)**

*It has been said that Sigmund Freud made a “rational decision” to end his life. Suffering from terminal throat cancer, having great difficulty in speaking, and experiencing increasing difficulty in breathing, the founder of psychoanalysis asked his physician for a lethal dose of morphine. For years it has been debated whether a decision to die, even on the part of a terminally ill patient, can ever truly be “rational.” Today, in accordance with death-with-dignity legislation, the responsibility for evaluating just how rational such a choice is falls on mental health professionals.*

less-than-sure methods, enduring even greater suffering in the process. In marked contrast to such horror stories, the first patient to die under the ODDA is said to have described how the family “could relax and say what a wonderful life we had. We could look back at all the lovely things because we knew we finally had an answer” (cited in Farrenkopf & Bryan, 1999, p. 246).

Professional associations such as the American Psychological Association and the American Psychiatric

Association have long promulgated codes of ethics requiring the prevention of suicide. The enactment of the law in Oregon has placed clinicians in that state in a uniquely awkward position. For years, many of these same clinicians have devoted their efforts to suicide prevention. Currently, they have been thrust into the position of being a potential party to, if not a facilitator of, physician-assisted suicide—regardless of how the aid-in-dying process is referred to in the legislation. Note that the Oregon law scrupulously denies that its objective is the legalization of physician-assisted suicide. In fact, the language of the act mandates that action taken under it “shall not, for any purpose, constitute suicide, assisted suicide, mercy killing or homicide, under the law.” The framers of the legislation perceived it as a means by which a terminally ill individual could exercise some control over the dying process. Couched in these terms, the sober duty of the clinician drawn into the process may be made more palatable or even ennobled.

The ODDA provides for various records to be kept regarding patients who die under its provisions. Each year since the Act first took effect, the collected data is published

in an annual report. So, for example, in the 2006 report, we learn that the reasons most frequently cited for seeking to end one’s life were loss of autonomy, decreasing ability to participate in activities that made life enjoyable, loss of dignity, and inadequate pain control.

Psychologists and psychiatrists called upon to make death-with-dignity competency evaluations may accept or decline the responsibility (Haley & Lee, 1998). Judging from one survey of 423 psychologists in clinical practice in Oregon (Fenn & Ganzini, 1999), many of the psychologists who could be asked to make such a life-or-death assessment might decline to do so. About one-third of the sample responded that an ODDA assessment would be outside the scope of their practice. Another 53% of the sample said they would either refuse to perform the assessment and take no further action or refuse to perform the assessment themselves and refer the patient to a colleague.

Although firm guidelines as to what an ODDA assessment should entail have yet to be established, Farrenkopf and Bryan (1999) offered several useful suggestions, which are summarized as follows.

### The ODDA Assessment Process

#### 1. Review of Records and Case History

With the patient’s consent, the assessor will gather records from all relevant sources, including medical and mental health records. A goal is to understand the patient’s current functioning in the context of many factors, ranging from the current medical condition and prognosis to the effects of medication and substance use.

#### 2. Consultation with Treating Professionals

With the patient’s consent, the assessor may consult with the patient’s physician and other professionals involved in the case to better understand the patient’s current functioning and current situation.

#### 3. Patient Interviews

Sensitive but thorough interviews with the patient will explore the reasons for the aid-in-dying request, including the pressures and values motivating the request. Other areas to explore include: (a) the patient’s understanding of his or her medical condition, the prognosis, and the treatment alternatives; (b) the patient’s experience of physical pain, limitations of functioning, and changes over time in cognitive, emotional, and perceptual functioning; (c) the patient’s characterization of his or her quality of life, including exploration of related factors including personal identity, role functioning, and self-esteem; and (d) external pressures on the patient, such as personal or familial financial inability to pay for continued treatment.

#### 4. Interviews with Family Members and Significant Others

With the permission of the patient, separate interviews should be conducted with the patient’s family and significant others. One objective is to explore from their perspective how the patient has adjusted in the past to adversity and how the patient has changed and adjusted to his or her current situation.

#### 5. Assessment of Competence

Like the other elements of this overview, this aspect of the assessment is complicated, and only the barest of guidelines can be presented here. In general, the assessor seeks to understand the patient’s reasoning and decision-making process, including all information relevant to the decision and its consequences. Some formal tests of competency are available (Appelbaum & Grisso, 1995a, 1995b; Lavin, 1992), but the clinical and legal applicability of such tests to an ODDA assessment has yet to be established.

#### 6. Assessment of Psychopathology

To what extent is the decision to end one’s life a function of pathological depression, anxiety, dementia, delirium, psychosis, or some other pathological condition? This is a question the assessor addresses using not only interviews but formal tests. Examples of the many possible instruments the assessor might employ include intelligence tests, personality tests, neuropsychological tests, symptom checklists, and depression and anxiety scales; refer to the appendix in Farrenkopf and Bryan (1999) for a complete list of these tests.

#### 7. Reporting Findings and Recommendations

Findings, including those related to the patient’s mental status and competence, family support and pressures, and anything else relevant to the patient’s aid-in-dying request, should be reported. If treatable conditions were found, treatment recommendations relevant to those conditions may be made. Nontreatment types of recommendations may include recommendations for legal advice, estate planning, or other resources. In Oregon, a Psychiatric/Psychological Consultant’s Compliance Form with the consultant’s recommendations should be completed and sent to the Oregon Health Division.

Adapted from Farrenkopf and Bryan (1999).

potential testing activities that computer technology brings to the testing industry have been a great boon. Of course, every rose has its thorns.

For assessment professionals, some major issues with regard to CAPA are as follows.

- *Access to test administration, scoring, and interpretation software.* Despite purchase restrictions on software and technological safeguards to guard against unauthorized copying, software may still be copied. Unlike test kits, which may contain manipulatable objects, manuals, and other tangible items, a computer-administered test may be easily copied and duplicated.
- *Comparability of pencil-and-paper and computerized versions of tests.* Many tests once available only in a paper-and-pencil format are now available in computerized form as well. In many instances, the comparability of the traditional and the computerized forms of the test has not been researched or has only insufficiently been researched.
- *The value of computerized test interpretations.* Many tests available for computerized administration also come with computerized scoring and interpretation procedures. Thousands of words are spewed out every day in the form of test interpretation results, but the value of these words in many cases is questionable.
- *Unprofessional, unregulated “psychological testing” online.* A growing number of Internet sites purport to provide, usually for a fee, online psychological tests. Yet the vast majority of the tests offered would not meet a psychologist’s standards. Assessment professionals wonder about the long-term effect of these largely unprofessional and unregulated “psychological testing” sites. Might they, for example, contribute to more public skepticism about psychological tests?

#### JUST THINK . . .

Use any search engine to find some Web sites purporting to administer quick and easy psychological tests. See if you can tell why a psychologist might consider the test to be more for entertainment purposes than for psychological insight. By the way, you may wish to revisit the test you selected after you have read Chapter 8 of this book and see if you can identify more reasons why it may not be considered a test in the eyes of professionals.

Imagine being administered what has been represented to you as a “psychological test” only to find that the test is not bona fide. The online availability of myriad tests of uncertain quality that purport to measure psychological variables increases the possibility of such events. To help remedy such potential problems, a Florida-based organization called the International Test Commission developed the “International Guidelines on Computer-Based and Internet-Delivered Testing” (Coyne, 2006). These guidelines address technical, quality, security, and related issues. Although not without limitations (Sale, 2006), these guidelines clearly represent a step forward in nongovernmental regulation.

Let’s now consider some other rights of testtakers.

### *The Rights of Testtakers*

As prescribed by the *Standards* and in some cases by law, some of the rights that test users accord to testtakers are the right of informed consent, the right to be informed of test findings, the right to privacy and confidentiality, and the right to the least stigmatizing label.

**The right of informed consent** Testtakers have a right to know why they are being evaluated, how the test data will be used, and what (if any) information will be released to whom. With full knowledge of such information, testtakers give their **informed**

**consent** to be tested. The disclosure of the information needed for consent must, of course, be in language the testtaker can understand. Thus, for a testtaker as young as 2 or 3 years of age or an individual who is mentally retarded with limited language ability, a disclosure before testing might be worded as follows: "I'm going to ask you to try to do some things so that I can see what you know how to do and what things you could use some more help with" (APA, 1985, p. 85).

If a testtaker is incapable of providing an informed consent to testing, such consent may be obtained from a parent or a legal representative. Consent must be in written rather than oral form. The written form should specify (1) the general purpose of the testing, (2) the specific reason it is being undertaken in the present case, and (3) the general type of instruments to be administered. Many school districts now routinely send home such forms before testing children. Such forms typically include the option to have the child assessed privately if a parent so desires. In instances where testing is legally mandated (as in a court-ordered situation), obtaining informed consent to test may be considered more of a courtesy (undertaken in part for reasons of establishing good rapport) than a necessity.

One gray area with respect to the testtaker's right of fully informed consent before testing involves research and experimental situations wherein the examiner's complete disclosure of all facts pertinent to the testing (including the experimenter's hypothesis and so forth) might irrevocably contaminate the test data. In some instances, deception is used to create situations that occur relatively rarely. For example, a deception might be created to evaluate how an emergency worker might react under emergency conditions. Sometimes deception involves the use of confederates to simulate social conditions that can occur during an event of some sort.

For situations in which it is deemed advisable not to obtain fully informed consent to evaluation, professional discretion is in order. Testtakers might be given a minimum amount of information before the testing. For example, "This testing is being undertaken as part of an experiment on obedience to authority." A full disclosure and debriefing would be made after the testing. Various professional organizations have created policies and guidelines regarding deception in research. For example, the APA *Ethical Principles of Psychologists and Code of Conduct* (2002) provides that psychologists (a) do not use deception unless it is absolutely necessary, (b) do not use deception at all if it will cause participants emotional distress, and (c) fully debrief participants.

**The right to be informed of test findings** In a bygone era, the inclination of many psychological assessors, particularly many clinicians, was to tell testtakers as little as possible about the nature of their performance on a particular test or test battery. In no case would they disclose diagnostic conclusions that could arouse anxiety or precipitate a crisis. This orientation was reflected in at least one authoritative text that advised testers to keep information about test results superficial and focus only on "positive" findings. This was done so that the examinee would leave the test session feeling "pleased and satisfied" (Klopfer et al., 1954, p. 15). But all that has changed, and giving realistic information about test performance to examinees is not only ethically and legally mandated but may be useful from a therapeutic perspective as well. Testtakers have a right to be informed, in language they can understand, of the nature of the findings with respect to a test they have taken. They are also entitled to know what recommendations are being made as a consequence of the test data. If the test results, findings, or recommendations made on the basis of test data are voided for any reason (such as irregularities in the test administration), testtakers have a right to know that as well.

Because of the possibility of untoward consequences of providing individuals with information about themselves—ability, lack of ability, personality, values—the

communication of results of a psychological test is a most important part of the evaluation process. With sensitivity to the situation, the test user will inform the testtaker (and the parent or the legal representative or both) of the purpose of the test, the meaning of the score relative to those of other testtakers, and the possible limitations and margins of error of the test. And regardless of whether such reporting is done in person or in writing, a qualified professional should be available to answer any further questions that testtakers (or their parents) have about the test scores. Ideally, counseling resources will be available for those who react adversely to the information presented.

**The right to privacy and confidentiality** The concept of the **privacy right** “recognizes the freedom of the individual to pick and choose for himself the time, circumstances, and particularly the extent to which he wishes to share or withhold from others his attitudes, beliefs, behavior, and opinions” (Shah, 1969, p. 57). When people in court proceedings “take the Fifth” and refuse to answer a question put to them on the grounds that the answer might be self-incriminating, they are asserting a right to privacy provided by the Fifth Amendment to the Constitution. The information withheld in such a manner is termed *privileged*; it is information that is protected by law from disclosure in a legal proceeding. State statutes have extended the concept of **privileged information** to parties who communicate with each other in the context of certain relationships, including the lawyer-client relationship, the doctor-patient relationship, the priest-penitent relationship, and the husband-wife relationship. In most states, privilege is also accorded to the psychologist-client relationship.

Privilege is extended to parties in various relationships because it has been deemed that the parties’ right to privacy serves a greater public interest than would be served if their communications were vulnerable to revelation during legal proceedings. Stated another way, it is for society’s good if people feel confident that they can talk freely to their attorneys, clergy, physicians, psychologists, and spouses. Professionals such as psychologists who are parties to such special relationships have a legal and ethical duty to keep their clients’ communications confidential.

**Confidentiality** may be distinguished from *privilege* in that, whereas “confidentiality concerns matters of communication outside the courtroom, privilege protects clients from disclosure in judicial proceedings” (Jagim et al., 1978, p. 459). Privilege is not absolute. There are occasions when a court can deem the disclosure of certain information necessary and can order the disclosure of that information. Should the psychologist or other professional so ordered refuse, he or she does so under the threat of going to jail, being fined, and other legal consequences.

Privilege in the psychologist-client relationship belongs to the client, not the psychologist. The competent client can direct the psychologist to disclose information to some third party (such as an attorney or an insurance carrier), and the psychologist is obligated to make the disclosure. In some rare instances, the psychologist may be ethically (if not legally) compelled to disclose information if that information will prevent harm either to the client or to some endangered third party. An illustrative case would be the situation where a client details a plan to commit suicide or homicide. In such an instance, the psychologist would be legally and ethically compelled to take reasonable action to prevent such an occurrence. Here, the preservation of life would be deemed an objective more important than the nonrevelation of privileged information.

A wrong judgment on the part of the clinician regarding the revelation of confidential communication may lead to a lawsuit or worse. A landmark Court case in this area was the 1974 case of *Tarasoff v. Regents of the University of California*. In that case, a therapy patient had made known to his psychologist his intention to kill an unnamed

but readily identifiable girl two months before the murder. The Court held that “protective privilege ends where the public peril begins,” and so the therapist had a duty to warn the endangered girl of her peril. Clinicians may have a duty to warn endangered third parties not only of potential violence but of potential AIDS infection from an HIV-positive client (Buckner & Firestone, 2000; Melchert & Patterson, 1999) as well as other threats to their physical well-being.

Another ethical mandate with regard to confidentiality involves the safekeeping of test data. Test users must take reasonable precautions to safeguard test records. If these data are stored in a filing cabinet then the cabinet should be locked and preferably made of steel. If these data are stored in a computer, electronic safeguards must be taken to ensure only authorized access. The individual or institution should have a reasonable policy covering the length of time that records are stored and when, if ever, the records will be deemed to be outdated, invalid, or useful only from an academic perspective. In general, it is not a good policy to maintain all records in perpetuity. Policies in conformance with privacy laws should also be in place governing the conditions under which requests for release of records to a third party will be honored.

Relevant to the release of assessment-related information is the Health Insurance Portability and Accountability Act of 1996 (HIPAA), which took effect in April 2003. These federal privacy standards limit the ways that health care providers, health plans, pharmacies, and hospitals can use patients’ personal medical information. For example, personal health information may not be used for purposes unrelated to health care.

In part due to the decision of the U.S. Supreme Court in the case of *Jaffee v. Redmond* (1996), HIPAA singled out “psychotherapy notes” as requiring even more stringent protection than other records. The ruling in *Jaffee* affirmed that communications between a psychotherapist and a patient were privileged in federal courts. The HIPAA privacy rule cited *Jaffee* and defined privacy notes as “notes recorded (in any medium) by a health care provider who is a mental health professional documenting or analyzing the contents of conversation during a private counseling session or a group, joint, or family counseling session and that are separated from the rest of the individual’s medical record.” Although “results of clinical tests” were specifically *excluded* in this definition, we would caution assessment professionals to obtain specific consent from assessees before releasing assessment-related information. This is particularly essential with respect to data gathered using assessment tools such as the interview, behavioral observation, and role play.

**The right to the least stigmatizing label** The *Standards* advise that the least stigmatizing labels should always be assigned when reporting test results. To better appreciate the need for this standard, consider the case of Jo Ann Iverson.<sup>6</sup> Jo Ann was 9 years old and suffering from claustrophobia when her mother brought her to a state hospital in Blackfoot, Idaho, for a psychological evaluation. Arden Frandsen, a psychologist employed part-time at the hospital, conducted an evaluation of Jo Ann, during the course of which he administered a Stanford-Binet Intelligence Test. In his report, Frandsen classified Jo Ann as “feeble-minded, at the high-grade moron level of general mental ability.” Following a request from Jo Ann’s school guidance counselor, a copy of the psychological report was forwarded to the school—and embarrassing rumors concerning Jo Ann’s mental condition began to circulate.

---

6. See *Iverson v. Frandsen*, 237 F. 2d 898 (Idaho, 1956) or Cohen (1979, pp. 149–150).

Jo Ann’s mother, Carmel Iverson, brought a libel (defamation) suit against Frandsen on behalf of her daughter.<sup>7</sup> Mrs. Iverson lost the lawsuit. The court ruled in part that the psychological evaluation “was a professional report made by a public servant in good faith, representing his best judgment.” But although Mrs. Iverson did not prevail in her lawsuit, we can certainly sympathize with her anguish at the thought of her daughter going through life with a label such as “high-grade moron”—this despite the fact that the psychologist had probably merely copied that designation from the test manual. We would also add that the Iversons may have prevailed in their lawsuit had the cause of action been breach of confidentiality and the defendant been the guidance counselor; there was uncontested testimony that it was from the guidance counselor’s office, and not that of the psychologist, that the rumors concerning Jo Ann first emanated.

While on the subject of the rights of testtakers, let’s not forget about the rights—of sorts—of students of testing and assessment. Having been introduced to various aspects of the assessment enterprise, you have the right to learn more about technical aspects of measurement. Exercise that right in the succeeding chapters.

---

## Self-Assessment

Test your understanding of elements of this chapter by seeing if you can explain each of the following terms, expressions, abbreviations, events, or names in terms of their significance in the context of psychological testing and assessment:

affirmative action	HIPAA	psychoanalysis
<i>Albemarle Paper Company v. Moody</i>	hired gun	Public Law 105-17
Alfred Binet	<i>Hobson v. Hansen</i>	quota system
James McKeen Cattell	imperial examinations	releasing the roll
<i>Code of Fair Testing Practices in</i>	informed consent	Hermann Rorschach
<i>Education</i>	<i>Jaffee v. Redmond</i>	self-report test
code of professional ethics	<i>Larry P. v. Riles</i>	<i>Sputnik</i>
confidentiality	laws	<i>Tarasoff v. Regents of California</i>
culture	minimum competency testing	truth-in-testing legislation
culture-specific test	programs	David Wechsler
Charles Darwin	Christiana D. Morgan	Lightner Witmer
<i>Debra P. v. Turlington</i>	Henry A. Murray	Robert S. Woodworth
ethics	ODDA	World War I
eugenics	Karl Pearson	World War II
Francis Galton	privacy right	Wilhelm Max Wundt
Henry H. Goddard	privileged information	
<i>Griggs v. Duke Power Company</i>	projective test	

---

7. An interesting though tangential aspect of this case was that Iverson had brought her child in with a presenting problem of claustrophobia. The plaintiff questioned whether the administration of an intelligence test under these circumstances was unauthorized and beyond the scope of the consultation. However, the defendant psychologist proved to the satisfaction of the Court that the administration of the Stanford-Binet was necessary to determine whether Jo Ann had the mental capacity to respond to psychotherapy.