1

WHAT ARE
PSYCHOLOGICAL TESTS?

LEARNING OBJECTIVES

After completing your study of this chapter, you should be able to do the following:

- Explain why you should care about psychological testing.

- Define what a psychological test is, including their similarities and differences.

- Trace the history of psychological testing from Alfred Binet and intelligence testing to the tests of today.

- Describe the three characteristics that are common to all psychological tests, and understand that psychological tests can demonstrate these characteristics to various degrees.

- Describe the assumptions that must be made when using psychological tests.

- Describe the different ways that psychological tests can be classified.

- Distinguish among four commonly used terms that students often confuse: *psychological assessment*, *psychological test*, *psychological measurement*, and *survey*.

- Locate print and online resources to obtain information about psychological tests.

"When I was in the second grade, my teacher recommended that I be placed in the school's gifted program. As a result, the school psychologist interviewed me and had me take an intelligence test."

"Last semester I took a class in abnormal psychology. The professor had all of us take several personality tests, including the MMPI [Minnesota Multiphasic Personality Inventory]. It was

awesome! We learned about different types of psychological disorders that the
MMPI can help diagnose."

"This year I applied for a summer job with a local bank. As a part of the selection
process, I had to participate in a structured interview and take a test that
measured my math and verbal skills."

"Yesterday I took my driving test—both the written and the road test. I couldn't
believe everything they made me do when taking the road test. I had to parallel
park, switch lanes, and make a three-point turn."

If your instructor asked whether you have ever taken a psychological test, you would probably talk about the intelligence test you took as an elementary school student or the personality test you took in your abnormal psychology class. If your instructor asked what the purpose of psychological testing is, you would probably say that it is to determine whether someone is gifted or has a psychological disorder. Intelligence tests and personality tests are indeed psychological tests—and they are indeed used to identify giftedness and diagnose psychological disorders. However, intelligence and personality tests are only two of many available tests. There are many more types of psychological tests, and they have many different purposes.

In this chapter, we introduce you to the concept of psychological testing. We discuss what a psychological test is and introduce some tests you might never have considered to be psychological tests. Then, after providing an overview of the history of psychological testing, we discuss the three defining characteristics of psychological tests and the assumptions we must make when using tests. We then turn our attention to the many ways of classifying tests. We also distinguish four concepts that students often confuse: psychological assessment, psychological tests, psychological measurement, and surveys. We conclude this chapter by sharing with you some of the resources (print and online) that are available for locating information about psychological testing and specific psychological tests.

WHY SHOULD YOU CARE ABOUT PSYCHOLOGICAL TESTING?

Before discussing what a psychological test is, we would like to increase your understanding of the importance of psychological testing and why you should care about it. Psychological testing is not just another subject that you may study in college; rather, it is a topic that personally affects many individuals, including yourself. Each day, different types of professionals administer psychological tests to many different individuals, and the results of these tests are used in ways that significantly affect you and those around you. For example, test scores are used to diagnose mental disorders, to determine whether medicines should be prescribed (and, if so, which ones), to treat mental and emotional illnesses, to certify or license individuals for professional practice, to select individuals for jobs, to make acceptance decisions for undergraduate and professional schools (e.g., medical school, law school), and to determine grades. Psychological tests may even be used in the future to combat terrorism (see In the News Box 1.1). Good tests facilitate high-quality decisions, and bad tests facilitate low-quality decisions.

IN THE NEWS BOX 1.1
COMBATING TERRORISM

Imagine if psychological tests could be used to help combat terrorism, one of the most malicious threats to our society.

In May 2017, BBC News highlighted a research study Baez conducted with 132 individuals (66 incarcerated individuals thought to be members of a terrorist organization and 66 demographically similar nonincarcerated individuals).

In May 2017, BBC News issued the following news release:

> A project aiming to "scientifically understand the mindset of terrorists" has published insights that the scientists say could have implications for terror prevention.

Researchers worked with a group of 66 incarcerated ex-combatants from a paramilitary terrorist group in Colombia, a country with one of the greatest insurgency rates in the world.

This unique experiment revealed what the team described as an "abnormal pattern of moral judgment" in terrorists.

The scientists say a psychological score based on this could be an accurate way to discriminate between the mindset of a terrorist and that of a noncriminal.

The researchers, based in Argentina, the United States, Colombia, and Chile, published their findings in the journal *Nature Human Behaviour*.

Agustín Ibanez and Adolfo García, from Favaloro University in Buenos Aires, who were part of the international research team, told BBC News they had spent 4 years working with Colombian law enforcers to secure permission to work with this large group of dangerous, incarcerated terrorists.

The study participants were former members of right-wing paramilitary groups, all of whom had been convicted of murder.

Many had been involved in massacres with hundreds of victims.

They took part in a series of psychological tests, including an assessment of moral cognition.

This involved presenting the subjects with a series of scenarios in which characters either deliberately or accidentally caused harm to others.

Each subject was then asked to rate the scenario on a scale from totally forbidden (1) to totally permissible (7).

Dr. Ibanez said, "The typical response is that attempted harm should be more objected to than accidental harm. [But] the pattern in terrorists was the opposite."

The pattern this research revealed was that "extreme terrorists judge other people's actions by focusing on the outcomes of an action rather than its underlying intentions.

"This is the first study to demonstrate this psychological trait, [and it suggests that] a terrorist's moral code actually approves of any action if it contributes to achieving a given aim."

Brutal cognition

The researchers hope the conclusions could help build a psychological profile for use in forensics and law enforcement.

But they say further research will need specifically to examine how predictive this measure of moral cognition is when it comes to "identifying dangerous insurgent individuals."

They also pointed out that there were likely to be differences in the "origins and psychological traits of different forms of terrorism."

"For example, in the population we studied, religion does not seem to be a relevant factor. [In fact], most ex-combatants in Colombia joined paramilitary groups for economic reasons—because they were paid a salary.

"But I would envisage forensic psychologists ultimately using a moral score like this to help assess how much of a threat a particular individual poses—in addition to other measures of aggression and emotions, as well as other cognitive and social tasks," Dr. Ibanez told BBC News.

Prof Seena Fazel, from the University of Oxford, a psychiatrist focusing on the relationship between mental illness and violent crime, told BBC News that the study was "a step forward."

He said there was value in the study's comparison of terrorists and noncriminals—the team carried out the same battery of tests on 66 healthy individuals from the same geographical region who had no terrorist background.

"I'd be interested in identifiable and modifiable factors that can either stop people repeating [a violent act] or stop them committing it in the first place. That would be where research could be very useful.

"These type of assessments rely on detailed interviews, so we're not at a point where we could scale up and implement this."

Source: BBC News Services (2017).

The consequences of low-quality decisions can be significant. For example, a poor hiring decision can dramatically affect both the person being hired and the hiring organization. From the organization's perspective, a poor hiring decision can result in increased absenteeism, reduced morale of other staff members, and lost productivity and revenue. From the employee's perspective, a poor hiring decision may result in a loss of motivation, increased stress leading to depression and anxiety, and perhaps loss of opportunity to make progress in his or her career. Although you might never administer, score, or interpret a test, it is very likely that you or someone you know may have a life-altering decision made about him or her based on test scores. Therefore, it is important that you understand the foundations of psychological testing, specifically how to tell whether a decision is a good or bad one. Being able to do this requires that you understand the foundations of psychological testing.

WHAT IS A PSYCHOLOGICAL TEST?

Each anecdote at the beginning of this chapter involves the use of a psychological test. Intelligence tests, personality tests, interest and vocational inventories, college entrance exams, classroom tests, structured interviews, workplace simulations (e.g., assessment centers), and driving tests all are examples of psychological tests. Even the self-scored tests that you find in magazines such as *Glamour* and *Seventeen* (tests that supposedly tell you how you feel about your friends, stress, love, and more) can be considered psychological tests. Although some are more typical, all meet the definition of a psychological test. Together, they convey the very different purposes of psychological tests. For a continuum of some of the most and least commonly recognized types of psychological tests, see Figure 1.1.

Similarities Among Psychological Tests

A review of professional association publications and psychological testing textbooks indicates that no one distinct definition of a psychological test exists. For example, the most recent version of the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association [APA], and the National Council on Measurement in Education, 2014) defines a psychological test as "a device or procedure in which a sample of an examinee's behavior in a specified domain is obtained and subsequently evaluated and scored using a standardized process" (p. 2). In its *Report of the Task Force on Test User Qualifications*, the APA (2000) defines a psychological test similarly as "a measurement procedure for assessing psychological characteristics in which a sample of an examinee's behavior is obtained and subsequently evaluated and scored using a standardized process" (p. 7).

FIGURE 1.1 ■ A Continuum of Psychological Tests			
More Typical	◄─────────────────────────►		**Less Typical**
Personality tests Intelligence tests	Vocational tests Interest inventories Achievement tests Ability tests	Self-scored magazine tests Classroom quizzes and exams	Road portion of driving test Structured employment interviews Assessment centers

The Association of Test Publishers (2017) refers to tests as "professionally developed instruments . . . [that are] constructed responsibly and are responsive to existing professional standards . . . begin[ning with] a set of questions . . . to assess or measure a specific ability or characteristic" (para. 8-9). On the other hand, textbook author Robert J. Gregory (2010) defined a test as "a standardized procedure for sampling behavior and describing it with categories or scores" (p. 16), while Kaplan and Saccuzzo (2013) defined a psychological test as "a measurement device or technique used to quantify behavior or aid in the understanding and prediction of behavior" (p. 7). Salkind (2013) provided an even broader definition of a test as "a tool, procedure, device, examination, investigation, assessment, or measure of an outcome (which is usually some kind of behavior)" (p. 9).

Although associations and testing professionals might have slight differences of opinion about what constitutes a psychological test, most definitions reference a procedure, an instrument, or a device and behaviors. In a broad sense, a psychological test is a procedure, an instrument, or a device that measures samples of behaviors in order to make inferences. To increase your understanding of what a psychological test is, let's consider what all psychological tests have in common.

First, all psychological tests require a person to perform some **behavior**—an observable and measurable action. For example, when students take multiple-choice midterm exams, they must read the various answers for each item and identify the best one. When individuals take intelligence tests, they may be asked to define words or solve math problems. When participating in structured job interviews, individuals must respond to questions from interviewers—questions such as, "Tell me about a time when you had to deal with an upset customer. What was the situation, what did you do, and what was the outcome?" In each of these cases, individuals are performing some observable and measurable behavior.

Second, the behavior an individual performs is used to make inferences about some **psychological construct**—an underlying, unobservable personal attribute, trait, or characteristic of an individual that is thought to be important in describing or understanding human behavior. Because we cannot directly measure psychological constructs, we must instead make inferences about constructs from the behaviors we observe. For example, how you answer questions on a multiple-choice exam might be used to make inferences about your knowledge of a particular subject area, such as psychological testing. The words you define or the math problems you solve might be used to make inferences about your verbal ability or quantitative reasoning. Likewise, the questions you answer during a structured job interview or on a cognitive ability test may be used make inferences about how successful you might be in a management position.

Given the discussion above, we define a **psychological test** as a measurement tool or technique that requires a person to perform one or more behaviors in order to make inferences about human attributes, traits, or characteristics or predict future outcomes. By **inference**, we mean using evidence to reach a conclusion. For example, you might take a test designed to measure your knowledge of the material presented in one of the chapters of this textbook. You answer multiple-choice questions (the behavior), and based on your test score, your instructor might reach a conclusion (the inference) about the knowledge you have of the material. For example, you might take the Hogan Personality Inventory when applying for a job. You answer questions (the behavior), and an employer may use your test results to draw a conclusion (the inference) about how likely it is that you will succeed in the job.

Differences Among Psychological Tests

Although all psychological tests require that you perform some behavior to make inferences about human attributes, traits, or characteristics or predict future outcomes, these tests can differ in various ways. For example, they can differ in terms of the behavior they require you

to perform, what they measure, their content, how they are administered and formatted, how they are scored and interpreted, and their psychometric quality (**psychometrics** is the quantitative and technical aspects of testing).

Behavior Performed

The one or more behaviors a test taker must perform vary by test. For example, a popular intelligence test, the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV), a general test of adult cognitive ability, requires test takers to (among other things) define words, repeat lists of digits, explain what is missing from pictures, and arrange blocks to duplicate geometric card designs (Pearson Education, 2018d). The Thematic Apperception Test (TAT), a widely used and researched projective personality test designed at Harvard University in the 1930s, requires test takers to look at ambiguous pictures showing a variety of social and interpersonal situations and to tell stories about each picture (Pearson Education, 2018c). The 2011 revised Graduate Record Examination (GRE) General Test, a graduate school admissions test that measures the verbal reasoning, quantitative reasoning, critical thinking, and analytical writing skills individuals have developed over time, requires test takers to do things such as answer multiple-choice questions, perform calculations, and respond to analytical writing tasks (Educational Testing Service, 2018). The road portion of an automobile driving test typically requires test takers to do things such as change lanes, make right and left turns, use turn signals properly, and parallel park. Assessment centers require job applicants to participate in simulated job-related activities (which mimic the activities they would perform on the job), such as engaging in confrontational meetings with disgruntled employees, processing email and paperwork, and conducting manager briefings.

Construct Measured and Outcome Predicted

What a test measures or predicts can vary. For example, the WAIS-IV requires individuals to explain what is missing from pictures to measure the construct of verbal intelligence. The TAT requires individuals to tell stories about pictures to identify conscious and unconscious drives, emotions, conflicts, and so on in order to ultimately measure the construct of personality. The road portion of a driving test requires individuals to perform various driving behaviors to measure the construct of driving ability. The GRE requires students to answer different types of questions to determine if they are ready for graduate-level work and predict success in graduate school.

Some of the constructs psychological tests may measure include personality, intelligence, cognitive ability, motivation, mechanical ability, vocational preference, achievement in a school subject, and anxiety—to name just a few. Some of the outcomes that tests typically predict include worker productivity, success in college, and who will benefit from specialized services such as clinical treatment programs.

More detail about the WAIS-IV can be found in **Test Spotlight 1.1** in Appendix A.

Content

Two tests that measure the same construct can require individuals to perform significantly different behaviors or to answer significantly different questions. Sometimes, how the test developers define the particular construct affects how the test is structured. For example, the questions on two intelligence tests may differ because one author may define intelligence as the ability to reason and another author may define intelligence in terms of **emotional intelligence**—"the abilities to perceive, appraise, and express emotions accurately and appropriately, to use emotions to facilitate thinking, to understand and analyze emotions, to use

emotional knowledge effectively, and to regulate one's emotions to promote both emotional and intellectual growth" (American Psychological Association, 2018b, para. 10).

The difference in content may also be due to the theoretical orientation of the test. (We talk more about theoretical orientation and its relation to test content in the validity chapters of this textbook.)

Administration and Format

Psychological tests can differ in terms of how they are administered and their format. A test can be administered in paper-and-pencil format (individually or in a group setting), on a computer, or verbally. Similarly, a psychological test may consist of multiple-choice items, agree/disagree items, true/false items, open-ended questions, or some mix of these. There are also tests that ask respondents to perform some behavior, such as sorting cards, playing a role, or writing an essay.

Scoring and Interpretation

Psychological tests can differ in terms of how they are scored and interpreted. Some tests require test takers to document answers on scannable sheets that are then computer-scored. Some tests are hand-scored by the person administering the test. Other tests are scored by the test takers themselves. In terms of interpretation, some tests generate results that can be interpreted easily by the test taker, and others require a knowledgeable professional to explain the results to the test taker.

Psychometric Quality

Last, but extremely important, psychological tests can differ in terms of their psychometric quality. For now, let us just say that there are a lot of really good tests out there that measure what they say they measure and do so consistently, but there are also a lot of really poor tests out there that do not measure what they say they measure. Good tests measure what they claim to measure, and any conclusions that are drawn from the test scores about the person taking the test are appropriate (they have evidence of validity). Good tests also measure whatever they measure consistently (they have evidence of reliability). The concepts of reliability and validity are central to determining whether a test is "good" or "bad" and are covered in detail later in this textbook. These concepts are so important that four chapters are devoted to them.

Because tests can differ in so many ways, to make informed decisions about tests, you must know how to properly critique a test. A critique of a test is an analysis of the test. A good critique answers many of the questions in Table 1.1. (These questions also appear in Appendix B.) Not all questions can be answered for all tests, as some questions are not relevant to all tests. Your instructor may have additional ideas about what constitutes a good critique.

THE HISTORY OF PSYCHOLOGICAL TESTING

Scholars report that the use of psychological tests can be traced back to 2200 BCE during ancient China's Xia Dynasty. The Xia Dynasty was the first government in traditional Chinese history, and this government instituted royal examinations. A summary of ancient China's royal examinations, leading to 20th-century uses of psychological tests, can be found in For Your Information Box 1.1.

TABLE 1.1   ■   Guidelines for Critiquing a Psychological Test

General descriptive information

- What is the title of the test?
- Who is the author of the test?
- Who publishes the test, and when was it published?
- How long does it take to administer the test?
- How much does it cost to purchase the test?
- Is the test proprietary or nonproprietary?

Purpose and nature of the test

- What does the test measure or predict?
- What behavior does the test require the test taker to perform?
- What population was the test designed for?
- What is the nature of the test (e.g., maximal performance, behavior observation, self-report, standardized or nonstandardized, objective or subjective)?
- What is the format of the test (e.g., paper and pencil or computerized, multiple choice or true/false)?

Practical evaluation

- Is there a test manual, and is it comprehensive? (Does it include information on how the test was constructed, its reliability and validity, and the composition of norm groups, and is it easy to read?)
- Is the test easy or difficult to administer?
- How clear are the administration directions?
- How clear are the scoring procedures?
- What qualifications and training does a test administrator need to have?
- Does the test have face validity?

Technical evaluation

- Is there a norm group?
- If there is a norm group, who constitutes the norm group?
- If there is a norm group, what types of norms are there (e.g., percentiles, standard scores)?
- If there is a norm group, how was the norm group selected?
- If there is a norm group, are there subgroup norms (e.g., by age, gender, region, occupation, and so on)?
- What evidence exists of test reliability?
- What evidence exists for the validity of the test?
- Is there a reported standard error of measurement, and if so, what is it?
- Are confidence intervals presented, and if so, what are they?

Test reviews

- What do reviewers say are the strengths and weaknesses of the test?
- What studies that used the test as a measurement instrument have been published in peer-reviewed journals?
- How did the test perform when researchers or test users, other than the test developer or publisher, used it?

Summary

- Overall, what are the strengths and weaknesses of the test?

FOR YOUR INFORMATION BOX 1.1

PSYCHOLOGICAL TESTS: FROM ANCIENT CHINA TO THE 20TH CENTURY

2200 BCE: Xia Dynasty

The use of psychological tests is thought to date back approximately 4,000 years to 2200 BCE, when the Chinese emperor Yushun administered examinations to officials every 3rd year to determine whether they were suitable to remain in office. However, modern scholars of ancient China say that little archaeological evidence supports the existence of such an examination process.

618–907 CE: T'ang Dynasty

The use of examinations in imperial China appeared to increase significantly during the T'ang dynasty. Individuals applying for the state bureaucracy had to take a civil service exam assessing their knowledge of classics and literary style (vs. technical expertise). However, some evidence exists that the examination results were not entirely used; rather, aristocrats occupied most positions throughout the T'ang dynasty.

1368–1644 CE: Ming Dynasty

During the Ming dynasty, examinations appear to have become a more formal part of the government official application process, as the emperor needed large numbers of government officials to help govern China. Government positions were widely sought because the positions were considered privileged and the salary enough for men (the only ones allowed to take the exam) to support their families for many years. Examination results were associated with the granting of formal titles, similar to today's university degrees. Upon passing different levels of the examination, people received more titles and increasingly more power in the civil service. Although the examinations were reportedly distressful, they appeared to keep talented men in the national government and kept members of the national government from becoming nobility because of their descent. Seeing the value of the examinations for making important decisions, European governments, and eventually the governments of the United Kingdom, the United States, Canada, and other countries, adopted such examination systems.

1791: France and Britain

France initially began using the same kind of examination system around 1791. However, soon after, Napoleon temporarily abolished the exams. The system adopted by France served as a model for a British system begun in 1833 to select trainees for the Indian civil service—the beginning of the British civil service.

1800s: The United States

Because of the success of the British system, Senator Charles Sumner and Representative Thomas Jenckes proposed to Congress in 1860 that the United States use a similar system. Jenckes's report, *Civil Service in the United States*, described the British and Chinese systems in detail. This report laid the foundation for the Civil Service Act of 1883, a federal law Senator George H. Pendleton sponsored, intended to improve the U.S. civil service. Among other things, the act required that individuals be awarded government jobs not on political affiliation but rather based on merit, which was to be determined by the results of competitive exams.

Just prior, in 1859, Charles Darwin published his book *On the Origin of Species*. Although it had long been recognized that no two human beings are the same, Darwin provided additional insight about individual differences and survival of the fittest—specifically how species with the most adaptive characteristics survive and then pass these characteristics on to their offspring. Applying Darwin's theory to humans, Sir Francis Galton (Darwin's cousin) published his own

(Continued)

(Continued)

book, *Hereditary Genius*, in which he shared his theory that some people are more fit than others because of their characteristics. He conducted subsequent experiments to test his theory of individual differences, focusing on motor and sensory functioning. Over time, Galton published extensively on individual differences and reaction time and first introduced the term *mental tests* to refer to a set of tests he had developed to measure various constructs, such as reaction time and human memory.

In 1879, Wilhelm Wundt introduced the first psychological laboratory, in Leipzig, Germany. At this time, psychology was still primarily the study of the similarities among people. For example, physiological psychologists studied how the brain and the nervous system functioned, and experimental psychologists conducted research to discover how people learn and remember. However, in the 1880s, James McKeen Cattell began

conducting additional research extending Galton's work, studying individual differences and mental processes. Strongly influenced by Cattell, psychologists began focusing more attention on exploring individual differences. Cattell and others realized that learning about the differences among people was just as important as learning about the similarities among people. They believed that developing formal psychological tests to measure individual differences could help solve many social problems, such as who should be placed in remedial programs, who should be sent to battlefields, and who should be hired for particular jobs.

Although Darwin's, Galton's, and Cattell's work on individual differences contributed to the development of the field of psychological testing, the experimental research of German researchers Ernst Weber, Gustav Fechner, and Wilhelm Wundt helped us understand the need for rigor with testing.

Source: Darwin (1859/1936), DuBois (1970), Eberhard (1977), Franke (1960), Galton (1869/2000), Ho (1962), Hucker (1978), Jenckes (1868), Kracke (1963), Miyazaki (1981), Pirazzoli-t'Serstevens (1982), Rodzinski (1979), Thorne and Henley (2001), and W. Martin (1870).

Most scholars agree that serious research efforts on psychological tests, particularly in the United States, did not begin until late in the 19th century, with the advent of intelligence testing. At this time, serious research efforts began on the use and usefulness of various testing procedures. One of the most important breakthroughs in modern-day testing was the work of Théodore Simon and Alfred Binet's research on intelligence in children.

Intelligence Tests

Alfred Binet and the Binet–Simon Scale

Late in the 19th century, Alfred Binet, whom many refer to as a self-educated psychologist, abandoned his law career to pursue his interests in psychology. Over the years, Binet became a major figure in the development of experimental psychology in France. He founded the first French psychology journal, and in his lab, he developed experimental techniques to measure intelligence and reasoning ability. He believed that intelligence was a complex characteristic that could be determined by evaluating a person's reasoning, judgment, and problem-solving abilities. Binet tried a variety of tasks to measure reasoning, judgment, and problem solving on his own children as well as on other children in the French school system. Binet was fascinated by the differences he observed among the children.

After accepting an invitation to become a member of the Free Society for the Psychological Study of the Child in 1899, and being appointed to the Commission for the Retarded in 1904, Binet and other society members began studying children more scientifically to address the problem of retardation among schoolchildren in Paris schools. Their hope was to be able to differentiate between normal children who could benefit from regular school programs and those with intellectual disabilities who, no matter how hard they tried, could not benefit from

regular school programs. They wanted to help school administrators determine how to best help slow learners.

Binet was successful in measuring intelligence, and in 1905, he and Théodore Simon published the first test of mental ability: the Binet–Simon Scale (Binet & Simon, 1905). The purpose of the scale was to identify intellectually subnormal individuals, and it consisted of 30 items. To determine if a person was intellectually subnormal, Binet and Simon compared individual test scores with the scores of 50 children who were administered the test with exactly the same instructions and format, or what we call standardized conditions. The scores of these 50 children served as a frame of reference (standardization sample) for interpreting the test scores of other children. When a child took the test, his or her test score could be compared with the average of the group of 50 children to provide meaning to the child's score.

Over the years, significant improvements were made to the Binet–Simon Scale. The standardization sample was significantly increased, the number of test items increased, and the indicator mental age was added to indicate the level at which a child performed on the test. To help interpret test results, a child's mental age, based on his or her test score, was compared with the actual, or chronological, age. For example, a 10-year-old child's test score might have indicated a mental age of 8, 10, or perhaps 12. A tested mental age of 8 would indicate that the child was behind others in intellectual development, whereas a mental age of 12 would indicate that the child was more intellectually advanced.

Lewis Terman and the Stanford–Binet Intelligence Scales

Binet's work influenced psychologists across the globe. Psychological testing became a popular method of evaluation, and the Binet–Simon Scale was adapted for use in many countries. In 1916, Lewis Terman, an American psychologist and Stanford University faculty member, produced the Stanford–Binet Intelligence Scales (Terman, 1916), an adaptation of Binet's original test. The standardization sample was increased, and test items were revised and added. Added was an intelligence quotient (IQ) index. Developed for use with Americans ages 3 years to adulthood, the test was used for many years. A revised edition of the Stanford–Binet Intelligence Scale remains one of the most widely used intelligence tests today.

> More detail about the Stanford–Binet Intelligence Scales can be found in **Test Spotlight 1.2** in Appendix A.

The Wechsler–Bellevue Intelligence Scale and the Wechsler Adult Intelligence Scale

By the 1930s, thousands of psychological tests were available, and psychologists and others were debating the nature of intelligence (what intelligence was all about). This dispute over defining intelligence prompted the development in 1939 of the original Wechsler–Bellevue Intelligence Scale (WBIS) for adults, which provided an index of general mental ability (as did the Binet–Simon Scale) and revealed patterns of a person's intellectual strengths and weaknesses. David Wechsler, the chief psychologist at Bellevue Hospital in New York City, constructed the WBIS, believing that intelligence is demonstrated based on an individual's ability to act purposefully, think logically, and interact and cope successfully with the environment (Hess, 2001; Rogers, 2001; Thorne & Henley, 2001). Wechsler published the second edition, the WBIS-II, in 1946.

In 1955, Wechsler revised the WBIS-II and renamed it the Wechsler Adult Intelligence Scale (WAIS). In 1981 and 1991, the WAIS was updated and published as the WAIS-Revised and the WAIS–Third Edition, respectively. In a continuing effort to improve the measurement of intelligence, as well as the clinical utility and user-friendliness of the test, the fourth edition was published in 2008 (Pearson Education, 2018d).

Personality Tests

In addition to intelligence testing, the early 1900s brought about an interest in measuring personality.

The Personal Data Sheet

During World War I, the U.S. military wanted a test to help detect soldiers who would not be able to handle the stress associated with combat. To meet this need, the APA commissioned an American psychologist and Columbia University faculty member, Robert Woodworth (1920), to design such a test, which came to be known as the Personal Data Sheet (PDS). The PDS was a paper-and-pencil test that required military recruits to respond "yes" or "no" to a series of 200 questions (eventually reduced to 116 questions) asked during a psychiatric interview that searched for emotional instability. The questions covered topics such as excessive anxiety, depression, abnormal fears, impulse problems, sleepwalking, nightmares, and memory problems (Segal & Coolidge, 2004). One question asked, "Are you troubled with the idea that people are watching you on the street?" During a pilot study of the test, new recruits on average showed 10 positive psychoneurotic symptoms; recruits who were deemed unfit for service generally showed 30 to 40 positive psychoneurotic symptoms (Segal & Coolidge, 2004). Unfortunately, because Woodworth did not complete the final design of this test until too late in the war, the PDS was never implemented or used to screen new recruits.

After World War I, Woodworth developed the Woodworth Psychoneurotic Inventory, a version of the PDS. Unlike the PDS, the Woodworth Psychoneurotic Inventory was designed for use with civilians and was the first self-report test. It was also the first widely used personality inventory.

Immediately following Woodworth's development of the PDS and the Woodworth Psychoneurotic Inventory, researchers continued development of structured personality tests. However, interest declined in structured personality testing, while interest increased in less structured measures of personality.

The Rorschach Inkblot Test and the TAT

During the 1930s, interest also grew in measuring personality by exploring the unconscious. With this interest came the development of two important projective tests based on the personality theories of Carl Jung: the Rorschach inkblot test and the TAT. The Rorschach, a projective personality test (described further in the "How Are Tests Used in Clinical and Counseling Settings?" chapter), was developed initially in 1921 by Swiss psychiatrist Hermann Rorschach and introduced in the United States in 1932. The Rorschach inkblot test was designed to identify personality disorders and assess mental functioning. Individuals look at inkblots and report their perceptions, which are then analyzed.

Two American psychologists at Harvard University, Henry A. Murray and C. D. Morgan, developed the TAT in the 1930s. The test was developed to measure an individual's patterns of thought, attitudes, observational capacity, and emotional responses to ambiguous test material. Individuals are shown ambiguous pictures and asked to tell a story for each picture, including such things as what led up to the event that is shown, what is happening, what the individuals in the picture are feeling and thinking, and what the outcome of the event is. As with the Rorschach test, individuals' responses are then analyzed.

Vocational Tests

During the 1940s, a need developed for **vocational tests** to help predict how successful an applicant would be in specific occupations. Such a test was needed because thousands of

people had lost their jobs during the Great Depression, and thousands more were coming out of school and seeking work. Because there were not enough jobs, people were forced to look for new lines of work. As a result, psychologists developed large-scale programs to design vocational aptitude tests that would predict how successful a person would be at an occupation before entering it. In 1947, the U.S. Employment Service developed the General Aptitude Test Battery (GATB) to meet this need (Dvorak, 1956). The GATB was used for a variety of purposes, including vocational counseling and occupational selection.

By the mid-20th century, numerous tests were available, and they were used by many to make important decisions about individuals. Because of the increased use of psychological tests, to help protect the rights of the test taker, the APA (1953) established the eight-member Committee on Ethical Standards for Psychology, which collaborated with more than 2,000 psychologists to publish the 170-page *Ethical Standards of Psychologists*. (We discuss these ethical standards in more detail in the "What Are the Ethical Responsibilities of Test Publishers, Test Users, and Test Takers?" chapter).

Testing in the 21st Century

Psychological testing is a big, multibillion-dollar business. There are thousands of published, commercially available tests, as well as thousands of unpublished tests. Tests are developed and published by hundreds of test publishing companies that market their tests very proactively—on the web and in catalogs. Over the years, test publishers' revenue has increased significantly, as the desire to use test scores to make important decisions has increased. For example, whereas in 1955, test sales by publishers totaled approximately $7 million, by 2001 sales had increased to approximately $250 million (Association of American Publishers, 2002). In 2002, the standardized testing market was valued somewhere between $400 million and $700 million per year. According to the Society for Human Resource Management, in 2015, workplace assessments alone were estimated to be a $500 million a year industry growing at about 10% per year.

Furthermore, following passage of the No Child Left Behind Act of 2001 (NCLB), the Pew Center on the States reported that spending on standardized tests went from $423 million in 2001 to $1.1 billion in 2008. In 2012, U.S. schools nationwide spent approximately $1.7 million per year on assessments for kindergarten through 12th grade (Chingos, 2012). According to the *2014 Global Assessment Trends Report*, 73% of employers used skills or knowledge tests as a part of their hiring processes (Kantrowitz, 2014). Other tests included personality tests (62%), cognitive ability and general problem-solving tests (59%), job fit tests (47%), situational judgment tests, (43%), culture fit tests (33%), and job simulations (32%).

With such significant spending on psychological tests and the numbers of schools and employers using psychological tests, the number of people who take psychological tests is not surprising. For example, in the early 1990s, 20 million Americans per year were taking psychological tests (Hunt, 1993), of whom 1 million were taking the SAT, a college admission test you may have taken as you prepared to apply to colleges. By 2013, the number of college-bound students taking the SAT had increased to 1.6 million (College Board, 2013). 2017 was a record year for the SAT, with other 1.8 million students taking the SAT (College Board, 2018). During the 2012–2013 academic year, over 238,000 individuals took the Graduate Management Admission Test, a graduate admission test most business schools require as a part of the application process (Graduate Management Admission Council, 2014). Furthermore, each year, as many as 1.5 million people take the Myers–Briggs Type Indicator (MBTI), a personality inventory. Testing is indeed a big business.

Although there are many unpublished tests, many psychological tests are commercially available, published by test publishers. For the names and web addresses of some of the most well-known test publishers, as well as some of the tests they publish, see On the Web Box 1.1.

Today, psychological testing is a part of American culture. Psychological tests are in use everywhere. For example, let us take a look at the Society for Human Resource Management (SHRM). As the world's largest association devoted to human resources management, SHRM provides human resources professionals with essential information and resources (Society for Human Resource Management, 2018). One of the resources SHRM provides, in partnership with the Performance Assessment Network, is access to the SHRM Talent Assessment Center, which includes more than 400 tests and surveys, from over 50 test publishers. SHRM members who are qualified testing professionals can identify one or more tests or surveys to help make selection and development decisions about current and potential employees. Tests and surveys can be identified based on assessment type (e.g., personality, cognitive ability, skills and knowledge, 360° feedback), purpose (e.g., selection, development), level (e.g., individual contributor, leader or manager), industry (e.g., education, construction), occupation (e.g., legal, computer and mathematical), or competency. The testing center allows qualified testing professionals to purchase individual tests, administer the tests online, and receive electronic reports.

Let's also take a look at the NCLB, which in 2002 raised significant awareness about testing in the United States. President George W. Bush signed the NCLB into law on January 8, 2002, intending to improve the performance of America's primary and secondary schools. The objective was to make all children proficient in reading and math by 2014. The act contained the four basic strategies for improving the performance of schools—strategies that were intended to change the culture of America's schools by defining a school's success in terms of the achievement of its students (U.S. Department of Education, 2004):

1. Increase the accountability that states, school districts, and schools have for educating America's children by requiring that all states implement statewide systems that (a) set challenging standards for what children in Grades 3 to 8 should know and learn in reading and math, (b) test students in Grades 3 to 8 on a yearly basis to determine the extent to which they know and have learned what they should have according to state standards, and (c) include annual statewide progress objectives to ensure that all students are proficient by 12th grade.

2. Ensure that all children have access to a quality education by allowing parents to send their children to better schools if their schools do not meet state standards.

3. Increase the amount of flexibility that high-performing states and school districts have for spending federal education dollars.

4. Place more emphasis on developing children's reading skills by making grants available to states to administer screening and diagnostic assessments to identify children who may be at risk for reading failure and by providing teachers with professional development and resources to help young children attain the knowledge and skills they need to be readers.

Requiring more accountability of U.S. schools, NCLB required mandatory testing of all students in Grades 3 to 12, as well as public access to overall aggregate results and subgroup results (e.g., for major racial and ethnic groups, low-income students, students with disabilities). The NCLB was replaced in 2015 by the Every Student Succeeds Act, retaining the most important parts of the NCLB (such as required testing), but providing states more control over accountability, teacher evaluation, and correcting issues with achievement gaps and failing schools (U.S. Department of Education, 2018).

Similarly focused on improving student achievement in the United States, in 2009 our nation's governors and education commissioners, with input from experts, school

ON THE WEB BOX 1.1
NAMES AND WEB ADDRESSES OF TEST PUBLISHERS

Open your web browser, go to your favorite search engine, and conduct a search for "test publishers" or "psychological test publishers." You will find pages and pages of websites dedicated to psychological testing and publishing. You will also find the websites of hundreds of test publishers. Although there are many different publishers, some of the most well known, including some of the widely known tests they publish, are listed here:

Publisher	Website	Popular Published Tests
Educational Testing Service	www.ets.org	• Graduate Record Examination General Test and Subject Tests • SAT Reasoning Test and SAT Subject Tests • Test of English as a Foreign Language • Graduate Management Admission Test
Pearson	www.pearsonassessments.com	• Bar-On Emotional Quotient Inventory • Bayley Scales of Infant and Toddler Development, Third Edition • Behavior Assessment System for Children, Second Edition, Behavioral and Emotional Screening System • Bender Visual-Motor Gestalt Test, Second Edition • Watson–Glaser Critical Thinking Appraisal
Hogan Assessment Systems	www.hoganassessments.com	• Hogan Personality Inventory • Hogan Development Survey • Hogan Business Reasoning Inventory • Motives, Values, Preferences Inventory
PAR	www.parinc.com	• Self-Directed Search • NEO Personality Inventory • Personality Assessment Inventory
Psytech International	www.psytech.com	• Occupational Interest Profile • Clerical Test Battery • Values & Motives Inventory
PSI	www.psionline.com	• Customer Service Battery • Firefighter Selection Test • Police Selection Test
Hogrefe	www.hogrefe.co.uk	• Rorschach Test • Trauma Symptom Inventory • Work Profile Questionnaire Emotional Intelligence Questionnaire
University of Minnesota Press Test Division	www.upress.umn.edu	• Minnesota Multiphasic Personality Inventory–2
Wonderlic	www.wonderlic.com	• Wonderlic Personnel Test

administrators, teachers, and parents, introduced the Common Core State Standards (Common Core State Standards Initiative, 2014). Led at the state and local levels, the Common Core includes high-quality academic expectations created to ensure that all students, regardless of where they live, graduate from high school with the knowledge and skills they need to succeed. As with NCLB, testing plays a critical role in measuring achievement.

THE THREE DEFINING CHARACTERISTICS OF PSYCHOLOGICAL TESTS

As we have already discussed, a psychological test is a measurement tool or technique that requires a person to perform one or more behaviors to make inferences about human attributes, traits, or characteristics or predict future outcomes. All good psychological tests have three characteristics in common:

1. **All good tests representatively sample the behaviors thought to measure an attribute or thought to predict an outcome.** For example, suppose we are interested in developing a test to measure your physical ability. One option would be to evaluate your performance in every sport you have ever played. Another option would be to have you run the 50-meter dash. Both of these options have drawbacks. The first option would be very precise, but not very practical. Can you imagine how much time and energy it would take to review how you performed in every sport you have ever played? The second option is too narrow and unrepresentative. How fast you run the 50-meter dash does not tell us much about your physical ability in general. A better method would be to take a representative sample of performance in sports. For example, we might require you to participate in some individual sports (e.g., running, tennis, gymnastics) and team sports (e.g., soccer, basketball) that involve different types of physical abilities (e.g., strength, endurance, precision). This option would include a more representative sample.

2. **All good tests include behavior samples that are obtained under standardized conditions.** That is, a test must be administered the same way to all people. When you take a test, various factors can affect your score besides the characteristic, attribute, or trait that is being measured. Factors related to the environment (e.g., room temperature, lighting), the examiner (e.g., examiner attitude, how the instructions are read), the examinee (e.g., illness, fatigue), and the test (e.g., understandability of questions) all can affect your score. If everyone is tested under the same conditions (e.g., in the same environment), we can be more confident that these factors will affect all test takers similarly. If all of these factors affect test takers similarly, we can be more certain that a person's test score accurately reflects the attribute being measured. Although it is possible for test developers to standardize factors related to the environment, the examiner, and the test, it is difficult to standardize examinee factors. For example, test developers have little control over what test takers do the night before they take a test.

3. **All good tests have rules for scoring.** These rules ensure that all examiners will score the same set of responses in the same way. For example, teachers might award 1 point for each multiple-choice question you answer correctly, and they might award or deduct points based on what you include in your response to an essay question. Teachers might then report your overall exam score either as the number correct or as a percentage of the number correct (the number of correct answers divided by the total number of questions on the test).

Although all psychological tests have these characteristics, not all exhibit these characteristics to the same degree. For example, some tests may include a more representative sample of behaviors than do others. Some tests, such as group-administered tests, may be more conducive to administration under standardized conditions than are individually administered tests. Some tests have well-defined rules for scoring, and other tests have general guidelines. Some tests have very explicit scoring rules, for example, "If Question 1 is marked true, then deduct 2 points." Other tests, such as those that include short answers, may have less explicit rules for scoring, for example, "Award 1 point for each concept noted and defined."

ASSUMPTIONS OF PSYCHOLOGICAL TESTS

There are many assumptions that must be made when using psychological tests. The following are what we consider the most important assumptions:

1. **Psychological tests measure what they purport to measure or predict what they are intended to predict.** In addition, any conclusions or inferences that are drawn about the test takers based on their test scores must be appropriate. This is also called test validity. If a test is designed to measure mechanical ability, we must assume that it does indeed measure mechanical ability. If a test is designed to predict performance on the job, then we must assume that it does indeed predict performance. This assumption must come from a personal review of the test's validity data.

2. **An individual's behavior, and therefore test scores, will typically remain stable over time.** This is also called test–retest reliability. If a test is administered at a specific point in time and then we administer it again at a different point in time (e.g., 2 weeks later), we must assume, depending on what we are measuring, that an individual will receive a similar score at both points in time. If we are measuring a relatively stable trait, we should be much more concerned about this assumption. However, there are some traits, such as mood, that are not expected to show high test–retest reliability.

3. **Individuals understand test items the same way.** For example, when asked to respond "true" or "false" to a test item such as "I am almost always healthy," we must assume that all test takers interpret "almost always" similarly.

4. **Individuals will report accurately about themselves** (e.g., about their personalities, about their likes and dislikes). When we ask people to remember something or to tell us how they feel about something, we must assume that they will remember accurately and that they have the ability to assess and report accurately on their thoughts and feelings. For example, if we ask you to tell us whether you agree or disagree with the statement "I have always liked cats," you must remember not only how you feel about cats now but also how you felt about cats previously.

5. **Individuals will report honestly their thoughts and feelings.** Even if people are able to report correctly about themselves, they may choose not to do so. Sometimes people respond how they think the tester wants them to respond, or they lie so that the outcome benefits them. For example, if we ask test takers whether they have ever taken a vacation, they may tell us that they have even if they really have not. Why? Because we expect most individuals to occasionally take vacations, and therefore the test takers think we would expect most individuals to answer "yes" to this question. Criminals may respond to test questions in a way that makes them appear neurotic or psychotic so that they can claim that they were insane when they committed crimes. When people report about

themselves, we must assume that they will report their thoughts and feelings honestly, or we must build validity checks into the test. We discuss some the practical implications of this assumption in the "How Are Tests Used in Organizational Settings?" chapter.

6. **The test score an individual receives is equal to his or her true score plus some error, and this error may be attributable to the test itself, the examiner, the examinee, or the environment.** That is, a test taker's score may reflect not only the attribute being measured but also things such as awkward question wording, errors in administration of the test, examinee fatigue, and the temperature of the room in which the test was taken. When evaluating an individual's score, we must assume that it will include some error.

Although we must accept some of these assumptions at face value, we can increase our confidence in others by following certain steps during test development. For example, in Section III of this textbook, where we cover developing and piloting psychological tests, we talk about how to design test questions that are more likely to be understood universally. We also talk about the techniques that are available to promote honest answering. In Section II, which covers psychometric principles, we discuss how to gather evidence of test reliability/precision and validity for intended use.

TEST CLASSIFICATION METHODS

As we have already discussed, there are tens of thousands of commercially available psychological tests, and professionals refer to these tests in various ways. Sometimes professionals refer to them as tests of maximal performance, behavior observation tests, or self-report tests. Sometimes professionals refer to tests as being standardized or nonstandardized, objective or projective. Other times professionals refer to tests based on what the tests measure. In this section, we discuss the most common ways that professionals classify and refer to psychological tests.

Maximal Performance, Behavior Observation, or Self-Report

Most psychological tests can be defined as being tests of maximal performance, behavioral observation tests, or self-report tests.

- **Tests of maximal performance** require test takers to perform a particular well-defined task, such as making a right-hand turn, arranging blocks from smallest to largest, tracing a pattern, or completing mathematical problems. Test takers try to do their best because their scores are determined by their success in completing the task. Intelligence tests, tests of specific abilities (e.g., mechanical ability), driving tests (road and written), and classroom tests all are good examples of tests of maximal performance.

- **Behavior observation tests** involve observing people's behavior and how people typically respond in a particular context. Unlike with tests of maximal performance, many times people do not know that their behavior is being observed, and there is no single defined task for them to perform. Many restaurants use this technique to assess food servers' competence in dealing with customers. Sometimes managers hire trained observers to visit their restaurants disguised as typical customers. In exchange for a free meal or some predetermined compensation, observers agree to record specific behaviors performed by food servers. For example, observers may document

whether food servers greeted them in a friendly manner. Other examples of behavior observations include documenting job performance for performance appraisals or clinical interviews.

- **Self-report tests** require test takers to report or describe their feelings, beliefs, opinions, or mental states. Many personality inventories, such as the Hogan Personality Inventory (HPI), are self-report tests. The HPI, a test used primarily for personnel selection and individualized assessment, asks test takers to indicate whether each of more than 200 statements about themselves is true or false.

Most psychological tests fit one of the above categories, and some tests contain features of more than one category. For example, a structured job interview (which involves asking all job applicants a standard set of interview questions) could include both technical questions and questions about one's beliefs or opinions. Technical questions, which are well defined for the interviewee, qualify the interview as a test of maximal performance. Questions about beliefs and opinions qualify it as a self-report test. An interviewer may also observe interviewees' behaviors, such as their greetings, which would qualify the interview as a behavioral observation.

Standardized or Nonstandardized

Standardized tests are those designed to measure a specific construct, and after development, are administered to a large group of individuals who are similar to the group for whom the test has been designed. To interpret test scores, an individual's test score is compared to others similar to the individual. For example, if a test is designed to measure the construct of writing ability for high school students, after development, the test would be administered to a large group of high school students. This group is called the **standardization sample**—people who are tested to obtain data to establish a frame of reference for interpreting individual test scores. These data, called **norms**, indicate the average performance of a group and the distribution of scores above and below this average.

For example, if you took the SAT, the interpretation of your score included comparing it with the SAT standardization sample to determine whether your score was high or low in comparison with others and whether you scored above average, average, or below average. In addition, standardized tests always have specific directions for administration and scoring.

Nonstandardized tests do not have standardization samples and are more common than standardized tests. Nonstandardized tests are usually constructed by a teacher or trainer in a less formal manner for a single administration. For example, in many cases, the exams you take in your college courses are nonstandardized tests.

Objective or Projective

Sometimes people make a distinction between objective and projective tests. **Objective tests** are tests where test takers choose a response or provide a response and there are predetermined correct answers, requiring little subjective judgment of the person scoring the test. Objective tests require test takers to respond to structured true/false questions, multiple-choice questions, or rating scales. What the test taker must do is clear, for example, answer "true" or "false," circle the correct multiple-choice answer, or circle the correct item on the rating scale. The GRE, Stanford-Binet Intelligence Scales, GATB, and most classroom tests are examples of objective tests.

Another example of an objective test is the NEO Personality Inventory–3, an objective self-report instrument designed to identify what makes individuals unique in their thinking, feeling, and interaction with others. Although there are two forms of the inventory, both

measure five broad personality dimensions: neuroticism, extroversion, openness, agreeableness, and conscientiousness (PAR, 2018a). Test takers are asked to indicate whether they strongly disagree, disagree, are neutral, agree, or strongly agree with each of 240 statements. These statements are about their thoughts, feelings, and goals. For sample questions from the NEO Personality Inventory, see For Your Information Box 1.2.

On the other hand, **projective tests** are those on which test takers view and are asked to respond to unstructured or ambiguous stimuli such as images or incomplete sentences. The role of the test taker is less clear than with an objective test, and more subjectivity is involved in interpreting the test taker's answer. People who use projective tests believe that test takers project themselves into the tasks they are asked to perform and that their responses are based on what they believe the stimuli mean and on the feelings they experience while responding. These tests tend to elicit highly personal concerns. They are often used to detect unconscious thoughts or personality characteristics, and they may be used to identify the need for psychological counseling. The TAT is an example of a projective test. (The "How Are Tests Used in Clinical and Counseling Settings?" chapter contains more information on the TAT and other projective tests.)

> More detail about the NEO Personality Inventory can be found in **Test Spotlight 1.3** in Appendix A.

Dimension Measured

Psychological tests are often discussed in terms of the dimensions they measure. For example, sometimes we distinguish among achievement tests, aptitude tests, intelligence tests, personality tests, and interest inventories. We refer to these as dimensions because they are broader than a single attribute or trait level. Often these types of tests measure various personal attributes or traits.

FOR YOUR INFORMATION BOX 1.2
SAMPLE ITEMS FROM THE NEO PERSONALITY INVENTORY

The NEO Personality Inventory, developed by Costa and McCrae (1992), is an objective, self-report instrument designed to identify what makes individuals unique in their thinking, feeling, and interaction with others. The inventory measures five broad personality dimensions: neuroticism, extroversion, openness, agreeableness, and conscientiousness. Test takers are asked to indicate whether they strongly disagree (SD), disagree (D), are neutral (N), agree (A), or strongly agree (SA) with each of 240 statements. These statements are about their thoughts, feelings, and goals. In the following, we list a sample item from three of the five scales:

Neuroticism

Frightening thoughts
sometimes come into my head.   SD   D   N   A   SA

Extroversion

I don't get much pleasure
from chatting with people.   SD   D   N   A   SA

Openness

I have a very active imagination.   SD   D   N   A   SA

Source: Reproduced by special permission of the publisher, Psychological Assessment Resources, Inc., 16204 North Florida Avenue, Lutz, FL 33549, from the NEO Personality Inventory–Revised by Paul T. Costa Jr. and Robert R. McCrae. Copyright 1978, 1985, 1989, 1991, 1992 by Psychological Assessment Resources, Inc. (PAR). Further reproduction is prohibited without permission of PAR.

Achievement Tests

Achievement tests measure a person's previous learning in a specific academic area (e.g., computer programming, German, trigonometry, psychology). A test that requires you to list the three characteristics of psychological tests would be considered an achievement test. Achievement tests are also referred to as tests of knowledge.

Achievement tests are used primarily in educational settings to determine how much students have learned or what they can do at a particular point in time. Many elementary schools and high schools rely on achievement tests to compare what students know at the beginning of the year with what they know at the end of the year, to assign grades, to identify students with special educational needs, and to measure students' progress.

Aptitude Tests

Achievement tests measure a test taker's knowledge in a specific area at a specific point in time. **Aptitude tests** assess a test taker's potential for learning or ability to perform in a new job or situation. Aptitude tests measure the product of cumulative life experiences—or what one has acquired over time. They help determine what "maximum" can be expected from a person.

Schools, businesses, and government agencies often use aptitude tests to predict how well someone will perform or to estimate the extent to which an individual will profit from a specified course of training. Vocational guidance counseling may involve aptitude testing to help clarify the test taker's career goals. If a person's score is similar to scores of others already working in a given occupation, the test will predict success in that field.

Intelligence Tests

Intelligence tests, like aptitude tests, assess a test taker's ability to cope with the environment, but at a broader level. Intelligence tests are often used to screen individuals for specific programs (e.g., gifted programs, honors programs) or programs for the mentally challenged. Intelligence tests are typically used in educational and clinical settings.

Interest Inventories

Interest inventories assess a person's interests in educational programs for job settings and provide information for making career decisions. Because these tests are often used to predict satisfaction in a particular academic area or employment setting, they are administered primarily to students by counselors in high schools and colleges. Interest inventories are not intended to predict success; rather, they are intended only to offer a framework for narrowing career possibilities.

Personality Tests

Personality tests measure human character or disposition. The first personality tests were designed to assess and predict clinical disorders. These tests remain useful today for determining who needs counseling and who will benefit from treatment programs. Newer personality tests measure "normal" personality traits. For example, the MBTI is often used by industrial and organizational psychologists to increase employees' understanding of individual differences and to promote better communication among members of work teams. Career counselors also use the MBTI to help students select majors and careers consistent with their personalities.

Personality tests can be either objective or projective. The MBTI is an example of an objective personality test. Projective personality tests, such as the TAT, serve the same purpose as some objective personality tests, but they require test takers to respond to unstructured or ambiguous stimuli.

> More detail about the MBTI can be found in **Test Spotlight 1.4** in Appendix A.

Subject Tests

Many popular psychological testing reference books also classify psychological tests by subject. For example, the *Twentieth Mental Measurements Yearbook* (*MMY*; Carlson, Geisinger, & Jonson, 2017) classifies thousands of tests into 18 major subject categories:

- Achievement

- Behavior assessment

- Developmental

- Education

- English and language

- Fine arts

- Foreign languages

- Intelligence and general aptitude

- Mathematics

- Neuropsychological

- Personality

- Reading

- Science

- Sensorimotor

- Social studies

- Speech and hearing

- Vocations

Reference books such as the *MMY* often indicate whether a test is (a) a test of maximal performance, a behavior observation test, or a self-report test; (b) standardized or nonstandardized; and (c) objective or projective. We discuss the *MMY*, as well as other reference books, later in this chapter.

PSYCHOLOGICAL ASSESSMENT, PSYCHOLOGICAL TESTS, MEASUREMENTS, AND SURVEYS

Before discussing much more, we should spend some time discussing some terms that students often confuse—*psychological assessment*, *psychological test*, *measurement*, and *survey*. Students often think of psychological assessment and psychological testing as one and the same. Similarly, students often do not understand the difference between a psychological test and a survey. This section is designed to help you distinguish among these terms that are commonly used in psychological testing.

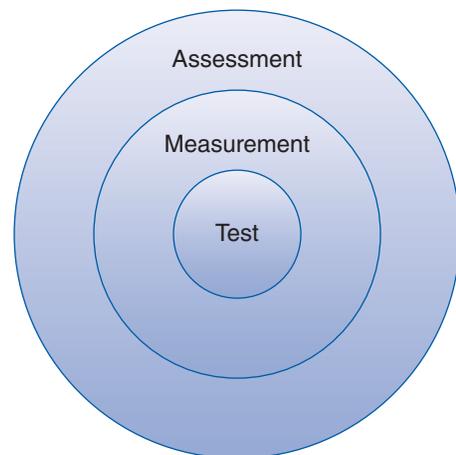## Psychological Assessment and Psychological Test

A psychological assessment and a psychological test are both evaluative methods of collecting important information about people. Both are used to help understand and predict behavior. While the terms are often used interchangeably, they are different. Many experts view psychological assessment as a broader concept than a psychological test (see Figure 1.2). With a **psychological assessment**, we use multiple methods, such as personal history interviews, behavioral observations, and psychological tests, for gathering information about an individual. A psychological test is only one tool in the psychological assessment process. For example, a clinical psychologist may conduct a psychological assessment of a patient by interviewing the patient, interviewing the patient's family members, observing the patient's behavior, and administering a psychological test such as the Minnesota Multiphasic Personality Inventory–2.

Psychological Test and Measurement

Although their meanings overlap, *psychological test* and *measurement* are not synonyms. Measurement, like assessment, is a broader concept than psychological test (see Figure 1.2). **Measurement** is the process of assessing the size, the amount, or the degree of an attribute using specific rules for transforming the attribute into numbers (Nunnally & Bernstein, 1994). For example, we measure a woman's dress size following specific rules—using a tape measure to measure the bust, waist, hips, and inseam. Or we might measure or quantify the size of an earthquake using the Richter or moment magnitude scale, focusing on the magnitude and intensity of ground movement. A **measurement instrument** is a tool or technique for assessing the size, amount, or degree of an attribute. So a psychological test can be considered a measurement or measurement instrument when the results of the test are expressed in terms of a derived score.

Even though some differences exist, you may find people who use the terms *psychological assessment*, *psychological test*, and *psychological measurement* interchangeably. Although we do not use the terms *assessment* and *psychological test* interchangeably throughout the remainder of this text, we do follow the common practice of referring to all psychological tests as measurements.

Psychological Tests and Surveys

Surveys, like psychological tests (and psychological assessments), are used to collect important information from individuals. Surveys differ from psychological tests in two important ways. First, psychological tests focus on individual outcomes, and surveys focus on group outcomes. Psychological tests provide important information about individual differences and help individuals and institutions make important decisions about individuals. For example, a psychological test may suggest that a child is unusually intelligent and therefore should be placed in a gifted or honors program. Surveys, on the other hand, provide important information about groups and help us make important decisions about groups. For example, an organizational

FIGURE 1.2 ■ Comparison of Psychological Assessment, Measurement, and Psychological Test

survey may suggest that employees are displeased with a company benefits program and that a new benefits program is needed.

Second, the results of a psychological test are often reported in terms of an overall derived score or scaled scores. Results of surveys, on the other hand, are often reported at the question level by providing the percentage of respondents who selected each answer alternative. Of course, in some cases, surveys focus on individual outcomes and are constructed using scales. In such a case, a survey approximates a psychological test. (The "How Do We Construct and Administer Surveys and Use Survey Data?" chapter is devoted to an in-depth discussion of surveys.)

LOCATING INFORMATION ABOUT TESTS

With so many psychological tests available, we are sure you can imagine that finding the most appropriate one for your specific purpose can be a difficult task. To choose an appropriate test for a particular circumstance, you must know the types of tests that are available and their merits and limitations. Prior to the 1950s, test users had few resources for obtaining such information. Today, however, numerous resources are available. Although all have the same general purpose—to help test users make informed decisions—the information such resources contain varies. Some resources provide only general descriptive information about psychological tests, such as the test's name, author, and publisher, and others contain detailed information, including test reviews and detailed bibliographies. Some resources focus on commercially available, standardized published tests, and others focus on unpublished tests. Some references include information about tests for particular groups (e.g., children), and others include a broad range of tests for various populations.

Some of the most commonly used resource books, including brief synopses of their contents, are described in For Your Information Box 1.3. The first four resource books, *Tests in Print* (*TIP*), the *MMY*, *Tests*, and *Test Critiques*, are often viewed as the most useful and popular.

Whether you are trying to locate tests that measure intelligence, self-esteem, or some other attribute, trait, or characteristic, we suggest that you begin your search with one of the first four resource books in For Your Information Box 1.3. *TIP* and the *MMY* are two of the most helpful references, and students often find it most helpful to begin with *TIP*. Figure 1.3 includes a descriptive guide of the type of information you will find in the *MMY*. Figure 1.4 includes a summary of how to use *TIP* to find tests. You can find more information on how to use both of these resources, as well as how to use the information contained in these resources to evaluate a test, on the Buros Center for Testing home page, discussed in On the Web Box 1.2.

Because there is a wealth of psychological tests available, there is a wealth of resources available for you to use in gathering information about psychological tests. You are not limited to print resources; advances in technology now allow you to access the Internet and gather information about psychological tests on demand. On the Web Box 1.2 discusses some websites you can access to locate information on psychological tests. For Your Information Box 1.4 discusses where you can locate unpublished psychological tests.

FOR YOUR INFORMATION BOX 1.3
COMMONLY USED RESOURCE BOOKS

Book Title	Contents
Tests in Print (multiple volumes)	*Tests in Print* (*TIP*) is published in multiple volumes. Each volume contains descriptive listings of commercially published tests that are available for purchase. *TIP* also serves as a comprehensive index to the contents of previously published editions of the *Mental Measurements Yearbook* (*MMY*). Each descriptive listing, or test entry, contains extensive information, including but not limited to the title of the test, the purpose of the test, the intended population, publication dates, the acronym used to identify the test, scores the test provides, whether the test is an individual test or group test, whether the test has a manual, the author(s), the publisher, the cost of the test, and available foreign adaptations. Each entry also contains brief comments about the test as well as cross-references to reviews in the *MMY*.
Mental Measurements Yearbook (multiple volumes)	The *MMY* is published in multiple volumes. Each volume contains descriptive information and test reviews of newly developed tests that are in English, commercially published tests, and tests that have been revised since the publication of the previous *MMY* edition. The *MMY* is cumulative, meaning that later volumes build on earlier ones rather than replacing them. Each descriptive listing, or test entry, contains extensive information about a particular test. If the test is a revision of a previous test, the entry also includes the volume of the *MMY* in which the test was originally described. Each entry also typically includes information about the test's reliability and validity, one or two professional reviews, and a list of references to pertinent literature.
Tests	*Tests* contains descriptions of a broad range of tests for use by psychologists, educators, and human resource professionals. Each entry includes the test's title, author, publisher, intended population, purpose, major features, administration time, cost, and availability.
Test Critiques (multiple volumes)	*Test Critiques* is published in multiple volumes. Each volume contains reviews of frequently used psychological, business, and educational tests. Each review includes descriptive information about the test (e.g., author, attribute measured, norms) and information on practical applications and uses. *Test Critiques* also contains in-depth information on reliability, validity, and test construction.
Personality Test and Reviews (multiple volumes)	*Personality Test and Reviews* is published in multiple volumes. Each volume contains a bibliography of personality tests that are contained in the *MMY*. Each entry contains descriptive information about the test as well as test reviews.
Tests in Education	*Tests in Education* contains descriptive and detailed information about educational tests for use by teachers, administrators, and educational advisers.
Testing Children	*Testing Children* contains descriptions of tests available for children. These descriptions include the knowledge, skills, and abilities measured by each test; the content and structure of the test; the time required to administer the test; the scores that are produced; the cost; and the publisher.
Tests and Measurements in Child Development: A Handbook	*Tests and Measurements in Child Development* contains a listing of unpublished measures for use with children as well as detailed information about each measure.
Measures for Psychological Assessment: A Guide to 3,000 Original Sources and Their Applications	*Measures for Psychological Assessment* is a guide that contains annotated references to thousands of less recognized assessment devices developed and described in journal articles.

FIGURE 1.3 ■ A Guide to Descriptive Entries in the Mental Measurements Yearbook

BUROS CENTER FOR TESTING

Mental Measurements Yearbook and *Tests in Print*
A Guide to the Descriptive Entries

Entry Number: The number cited in all indexes when referring to this test.

Title: Test titles are printed in boldface type; secondary or series titles are set off from main titles by colon.

Population: A description of the groups for which the test is intended.

Administration: Individual or group administration is indicated.

Distribution: This is noted only for tests that are put on a special market by the publisher.

Special Editions: Various types of special editions are listed here.

Author: All test authors' names are reported, exactly as printed on the test materials.

Cross References: For tests that have been previously listed in a Buros publication, cross references to the reviews, excerpts, and references will be noted here. "9:1410," for example, refers to test 1410 in the *Ninth Mental Measurements Yearbook*; "T4:3010" refers to test 3010 in *Tests in Print IV*.

[420]
The Hypothetical Test: Reading.
Purpose: Designed to "measure achievement in reading."
Population: Grades 9-12.
Publication Dates: 1989—1994.
Acronym: HYPE.
Scores, 3: Vocabulary, Comprehension, and Total.
Administration: Individual or group.
Forms, 3: Survey, Abbreviated, Complete Battery.
Restricted Distribution: Distribution of Survey Form restricted to school principals.
Price Data, 1995: $70 per complete kit including 100 tests, scoring key, and manual ('94, 120 pages); $9 per scoring key; $32 per manual.
Special Editions: Braille edition available.
Time: 50(60) minutes.
Comments: May be self-scored.
Author: Jane J. Doe.
Publisher: Hypothetical Tests, Inc.
Cross References: See T4:3010 (2 references); for reviews by John Roe and Robert Smith of an earlier edition, see 9:1410 (6 references).

Purpose: A brief, clear statement describing the purpose of the test; often these are quotations from the test manual.

Publication Date: The inclusive range of publication dates.

Acronym: Acronym by which the test may be commonly known.

Scores: The number and names of explicit scores are presented.

Forms: All available forms, parts, and levels are listed.

Price Data: Price information is reported for test packages, answer sheets, accessories, and specimen sets.

Time: This is the amount of time to take, and administer, the test. The first number is the actual working time examinees are allowed, and the second (parenthesized) number is the total time needed to administer the test.

Comments: Special notations and comments.

Publisher: The publisher's full address can be found in the Publishers Directory and Index.

Source: Buros Center for Testing, University of Nebraska–Lincoln (buros.org).

FIGURE 1.4 ■ How to Use Tests in Print

BUROS CENTER FOR TESTING

How to Use
Tests in Print

Tests in Print (TIP) *consists of descriptive listings, without reviews, of commercially published tests in print.* TIP *is also a comprehensive index to the contents of previously published* Mental Measurements Yearbooks.

1. If you know the TEST TITLE:

Use the "**Index of Titles**." The index lists all tests in that volume plus all tests out of print since last being listed. "2458," for example, refers to test 2458 in that volume; "9:1128" refers to now out-of-print test 1128 in the *Ninth Mental Measurements Yearbook. Citation numbers refer to entry numbers, not to page numbers.*

Example from "Index of Titles":

Short Tests of Clerical Ability, 2458
Shortened Edinburgh Reading Test, 2459
Shortened Aptitude Test, T4:2195
Signals Learning Test, 2461
Silver Burdett Music Competency Tests, 9:1128
Silver Drawing Test of Cognitive Skills and Adjustment, 2462
Simile Interpretations, T4:2198
Similes, T4:2199

2. If you know the TYPE OF TEST:

Use the "**Classified Subject Index**" to locate various categories of tests, such as achievement, intelligence, personality, etc. This index organizes all tests into 18 major categories; tests appear alphabetically within each category. *Citation numbers refer to entry numbers, not to page numbers.*

Example from "Classified Subject Index, Education":

Gifted Program Evaluation Survey, Gifted and talented programs, see 1040
Graduate Records Examinations Education Test, Graduate School candidates, see 1063
High School Characteristics Index, Grades 9-13, 4-13, see 1157
How a Child Learns, Classroom teachers, see 1175
Hudson Educational Skills Inventory, Grades K-12, see 1184

3. If you know the NAME OF THE TEST AUTHOR OR REVIEWER:

Use the "**Index of Names**." This index includes test authors (for example, "test, 1460"), review authors ("rev, 2589"), and authors of referenced articles ("ref, 2222"). (Parenthesized numbers indicate the reference number.) *Citation numbers refer to entry numbers, not to page numbers.*

Example from "Index of Names":

Caeglio, G.: test, 1460
Caffey, C. A.: ref, 2222(1)
Caggiula, A. A.: ref, 2563(449)
Cahalane, J.: ref, 268(395), 1043(39)
Cahen, L. S.: rev, 2589
Cahill, C.: ref, 1705(65), 2937(935)
Cahir, N.: ref, 1135(14), 2674(188)
Cahn, T. S.: ref, 268(90)
Cain, J.: ref, 93(84), 1690(84)
Cain, L. F.: test, 2844

Source: Buros Center for Testing, University of Nebraska–Lincoln (buros.org).

ON THE WEB BOX 1.2
LOCATING INFORMATION ABOUT TESTS ON THE WEB

Computer technology lets us connect to the Internet and locate websites containing valuable information about psychological tests. These websites include information such as the following:

- Frequently asked questions about psychological testing

- How to find a particular type of psychological test

- How to locate reviews of psychological tests

- How to select an appropriate test

- What qualifications are necessary to purchase psychological tests

- How to contact test publishers

- How to obtain copies of specific psychological tests

Although there are many available websites, here are four that we have found to be extremely valuable:

Website	Description
American Psychological Association www.apa.org/science/ programs/testing/find-tests .aspx#findinfo	Although the APA does not sell or endorse specific testing instruments, it does provide guidance on testing resources and how to find psychological tests. This website contains answers to the most frequently asked questions about psychological testing. One section focuses on questions about published psychological tests (those that can be purchased from a test publisher); here you will find advice on how to find information about a particular test and about the proper use of tests, how to contact test publishers and purchase tests, and available software and scoring services. Another section focuses on unpublished psychological tests and measures (those that are not commercially available); here you will find advice on how to find unpublished tests in your area of interest and important information regarding your responsibilities as a user of unpublished tests.
Buros Center for Testing buros.org	The Buros Center for Testing promotes the appropriate use of tests and provides professional assistance, expertise, and information to those who use commercially published tests. This website contains a number of instructional resources, tools, and links. For example, it contains detailed instructions on what information can be found in two popular Buros publications that we have already discussed: the *Mental Measurements Yearbook* (*MMY*) and *Tests in Print* (*TIP*). This site also contains some great "how-to" resources, such as how to use *TIP* and the *MMY* and how to use the information in these resources to evaluate a test. In addition, it contains a link to Test Reviews Online, a service that provides access to more than 2,000 test reviews, beginning with those that were published in the *Ninth MMY*. Likewise, there are links to the Code of Fair Testing Practices (discussed further in the "What Are the Ethical Responsibilities of Test Publishers, Test Users, and Test Takers?" chapter) and the APA's frequently asked questions website mentioned previously.

(Continued)

(Continued)

Website	Description
Test Collection at ETS www.ets.org/test_link/about	The Test Collection at ETS is the world's largest database of tests and measurement instruments that have been available since the early 1900s. This online database contains descriptions of more than 20,000 tests (published and unpublished) and research instruments, collected from test publishers and test authors from around the world. Each description includes the title of the test or instrument, the author, the publication date, availability (how to obtain the test or measurement), the intended population, and specific uses of the test or instrument. In addition to providing information about specific tests, this database contains valuable information on how to order tests.
O*NET Resource Center www.onetcenter.org/guides.html	O*NET is sponsored by the U.S. Department of Labor and is a primary source for occupational information. Consisting of a comprehensive database of worker attributes and job characteristics, O*NET also provides valuable resources on testing and assessment—resources intended to support public and private sector efforts to identify and develop the skills of the American workforce. This website provides access to three extremely valuable testing and assessment guides: • *Testing and Assessment: A Guide to Good Practices for Workforce Investment Professionals* includes information on how assessment instruments can be used to promote talent development in career counseling, training, and other talent development activities. It discusses how to evaluate and select assessment instruments, administer and score assessments to meet business and individual client needs, and accurately and effectively interpret assessment results. It also lists the professional and legal standards related to assessment use in talent development. • *Tests and Other Assessments: Helping You Make Better Career Decisions* includes an explanation of how assessment instruments are used in employment selection and career counseling and provides tips and strategies for taking tests and other assessments. • *Testing and Assessment: An Employer's Guide to Good Practices* helps managers and workforce development professionals understand and use employment testing and assessment practices to meet their organizations' human resources goals.

FOR YOUR INFORMATION BOX 1.4
LOCATING UNPUBLISHED PSYCHOLOGICAL TESTS

Although there are thousands of commercially available tests, there are just as many, if not more, unpublished tests designed and used by researchers. A number of print and nonprint resources are available for locating information on unpublished tests.

Two of the most popular print resources are the *Directory of Unpublished Experimental Mental Measures* (Goldman & Mitchell, 2007) and *Measures for Psychological Assessment: A Guide to 3,000 Original Sources and Their Applications* (Chun, Cobb, & French, 1975).

Three of the most popular nonprint resources for locating information about unpublished or noncommercial tests are Tests in Microfiche, the PsycINFO database, and the Health and Psychosocial Instruments (HaPI) database.

PsycTESTS

This subscription-based database contains thousands of psychological measures, scales, surveys, and other instruments that were originally developed, but are not available commercially. The psychometric properties are reported for many of the instruments, and the instruments are updated monthly. All of the research instruments are available for download.

Directory of Unpublished Experimental Mental Measures

This directory provides easy access to more than 5,000 experimental mental measures, tests, and surveys that have been used by other researchers but are not commercially available. Topics range from educational adjustment and motivation to personality and perception. The measures, tests, and surveys are arranged in a 24-category system and grouped according to function and content, noting purpose, format, psychometric information (where available), and related research. First published in 1974 and currently in its ninth edition, this resource is updated periodically by the publisher.

Measures for Psychological Assessment: A Guide to 3,000 Original Sources and Their Applications

This guide includes annotated references to psychological measures that have appeared in journal articles and other publications. Although a bit outdated, it can be a useful resource. It has two sections: primary references and applications. The primary references section includes the name of each measure, the reference in which the measure originally appeared, and one or more other researchers who have used the measure in experimental research. The applications section includes other research studies that have used the original measures and references other experimental tests.

Tests in Microfiche

This resource can be accessed through the Test Collection at ETS. It contains a variety of educational and psychological instruments that are cited in the literature but are either out of date or unpublished. It contains more than 800 tests, and new tests are added each year. For more information, go to www.ets.org/test_link/about or check with your college's library.

PsycINFO Database

This bibliographic database indexes published studies in psychology. By using the Form/Content field "Tests & Measures" to search the PsycINFO database, you can find tests that have been used in research and written about in the literature. For more information, go to www.apa.org/pubs/databases/psycinfo/index.aspx.

HaPI Database

This computerized database includes citations to unpublished health and psychosocial evaluation and measurement tools (e.g., questionnaires, interviews, tests, checklists, rating scales) that have appeared in journals and technical reports since 1985. HaPI is updated quarterly and contains more than 15,000 measurement instruments. HaPI is provided online by Ovid Technologies, which typically must be accessed through BRS Information Technologies at your college's library. Some libraries maintain the database on CD-ROM. For more information, see www.ovid.com/site/catalog/DataBase/866.jsp.

Chapter Summary

By now, we hope you understand that psychological testing extends well beyond the use of intelligence and personality tests. Any measurement tool or technique that requires a person to perform one or more behaviors to make inferences about human attributes, traits, or characteristics or predict future outcomes can be considered a psychological test. The quizzes and exams you take in class are psychological tests. The written and road portions of driving exams are psychological tests. Even the structured job interviews you have participated

(Continued)

in, or will participate in as you conduct your job search, qualify as psychological tests.

Psychological tests have various similarities and many differences. All psychological tests require an individual to perform one or more behaviors, and these behaviors are used to measure some personal attribute, trait, or characteristic thought to be important in describing or understanding behavior or to predict an outcome. However, psychological tests can and do differ in terms of the behaviors they require individuals to perform, the attributes they measure, their content, how they are administered and formatted, how they are scored and interpreted, and their psychometric quality.

Although the use of psychological tests can be traced to ancient China, most scholars agree that the advent of formal psychological testing did not begin until Binet published the first test of intelligence in 1905. Today, psychological testing is a big business, with tens of thousands of commercially available, standardized psychological tests as well as thousands of unpublished tests.

All good tests have three defining characteristics in common. First, they include a representative sample of behaviors. Second, they collect the sample under standardized conditions. Third, they have rules for scoring. When using psychological tests, we must make some assumptions. We must assume that a test measures what it says it measures, that any inferences that are drawn about test takers from their scores on the test are appropriate, that an individual's behavior (and therefore test scores) will remain stable over time, that individuals understand test items similarly, that individuals can and will report accurately about their thoughts and feelings, and that the test score an individual receives is equal to his or her true behavior or ability in the real world plus some error.

Testing professionals refer to psychological tests in various ways. Sometimes they refer to them as tests of maximal performance, behavior observations, or self-reports. Sometimes they refer to them as standardized or nonstandardized. Other times they refer to them as objective or projective. Professionals also refer to tests based on the dimensions they measure.

It is important to remember the distinctions among four commonly misunderstood terms: *psychological assessment*, *psychological test*, *measurement*, and *survey*. First, although both psychological assessments and psychological tests are used to gather information, a psychological test is only one of many tools in the psychological assessment process. Second, a psychological test can be considered to be a measurement when the sampled behavior can be expressed in a derived score. Third, psychological tests are different from surveys in that psychological tests focus on individual differences and often report one overall derived score (or scaled scores), and surveys focus on group similarities and typically report results at the question or item level.

Last, but not least, a number of resources are available, in print and online, to locate information about published and unpublished psychological tests and measures. The *MMY* and *TIP* are two of the most popular references for learning more about available tests.

Engaging in the Learning Process

Learning is the process of gaining knowledge and skills through schooling or studying. Although you can learn by reading the chapter material, attending class, and engaging in discussion with your instructor, more actively engaging in the learning process may help you better learn and retain chapter information. To help you actively engage in the learning process, we encourage you to access our new supplementary student workbook. The workbook contains critical thinking activities to help you understand and apply information and help you make progress toward learning and retaining material. If you do not have a copy of the workbook, you can purchase a copy through sagepub.com.

Key Concepts

After completing your study of this chapter, you should be able to define each of the following terms. These terms are bolded in the text of this chapter and defined in the Glossary.

achievement tests

aptitude tests

behavior

behavior observation tests

emotional intelligence

inference

intelligence tests

interest inventories

measurement

measurement instrument

nonstandardized tests

norms

objective tests

personality tests

projective tests

psychological assessments

psychological construct

psychological test

psychometrics

self-report tests

standardization sample

standardized tests

surveys

tests of maximal performance

vocational tests

Critical Thinking Questions

The following are some critical thinking questions to support the learning objectives for this chapter.

Learning Objectives	Critical Thinking Questions
Explain why you should care about psychological testing	• What are three specific reasons you should personally care about psychological testing?
Define what a psychological test is, and understand that psychological tests extend beyond personality and intelligence tests.	• Locate and read about two psychological tests that measure the same construct. How would you classify each test and why? How would you describe the similarities and differences between the two tests?
Trace the history of psychological testing from Alfred Binet and intelligence testing to the tests of today.	• What is the value of understanding the history of psychological testing? • What approach could you use to visually explain the history of psychological testing?
Describe the three characteristics that are common to all psychological tests, and understand that psychological tests can demonstrate these characteristics to various degrees.	• Imagine you were making a presentation to other students on the WAIS-IV. What opinion would you share about the degree to which the test meets the three defining characteristics of a good psychological test? • What are the implications of a psychological test not exhibiting all the three defining characteristics of a good psychological test?

(Continued)

(Continued)

Learning Objectives	Critical Thinking Questions
Describe the assumptions that must be made when using psychological tests.	• What might be the implications of using a psychological test without understanding the fundamental assumptions of psychological tests?
Describe the different ways that psychological tests can be classified.	• What are all the different ways you could classify the NEO Personality Inventory? What is the value of classifying tests? • What might some consequences be if a test publisher or user incorrectly classified a psychological test?
Distinguish the differences among four commonly used terms that students often get confused: *psychological assessment*, *psychological test*, *psychological measurement*, and *survey*.	• How are psychological assessments, psychological tests, psychological measurements, and surveys similar and different?
Locate print and online resources to obtain information about psychological tests.	• Which online resource would you find most helpful for locating information about commercially available tests an organization might use as part of a selection process? • What are some of the reasons that knowing about these resources are important to users of psychological tests?