All fields of human endeavor use measurement in some form, and each field has its own set of measuring tools and measuring units. If you’re recently engaged or thinking about becoming engaged, you may have learned about a unit of measure called the carat. If you’ve been shopping for a computer, you may have learned something about a unit of measurement called a byte. And if you’re in need of an air conditioner, you’ll no doubt want to know about the Btu (British thermal unit). Other units of measure you may or may not be familiar with include a mile, a nautical mile, miles per hour, and cycles per second. Professionals in the fields that employ these units know the potential uses, benefits, and limitations of such units in the measurements they make. So, too, users and potential users of psychological measurements need a working familiarity with the commonly used units of measure, the theoretical underpinnings of the enterprise, and the tools employed.

Testing and Assessment

The roots of contemporary psychological testing and assessment can be found in early twentieth-century France. In 1905, Alfred Binet and a colleague published a test designed to help place Paris schoolchildren in appropriate classes. Binet’s test would have consequences well beyond the Paris school district. Within a decade, an English-language version of Binet’s test was prepared for use in schools in the United States. When the United States declared war on Germany and entered World War I in 1917, the military needed a way to screen large numbers of recruits quickly for intellectual as well as emotional problems. Psychological testing provided this methodology. During World War II, the military would depend even more on psychological tests to screen recruits for service. Following the war, more and more tests purporting to measure an ever-widening array of psychological variables were developed and used.

Psychological Testing and Assessment Defined

The world’s receptivity to Binet’s test in the early twentieth century spawned not only more tests but more test developers, more test publishers, more test users, and the emergence of what, logically enough, has become known as a testing industry. Testing was the term used to refer to everything from the administration of a test (as in “Testing in
During World War I, the process of testing aptly described the group screening of thousands of military recruits. We suspect it was at that time that testing gained a powerful foothold in the vocabulary of professionals and lay people. The use of testing to denote everything from test administration to test interpretation can be found not only in postwar textbooks (such as Chapman, 1921; Hull, 1922; Spearman, 1927) but in varied test-related writings for decades thereafter. However, by World War II a semantic distinction between testing and a more inclusive term, assessment, began to emerge.

During World War II, the United States Office of Strategic Services (OSS) used a variety of procedures and measurement tools—psychological tests among them—in selecting military personnel for highly specialized positions involving espionage, intelligence gathering, and the like. As summarized in Assessment of Men (OSS, 1948) and elsewhere (Murray & MacKinnon, 1946), the assessment data generated were subjected to thoughtful integration and evaluation by highly trained assessment center staff. The OSS model—using an innovative variety of evaluative tools along with data from the evaluations of highly trained assessors—would later inspire what is now referred to as the assessment center approach to personnel evaluation (Bray, 1982).

Military, clinical, educational, and business settings are but a few of the many contexts that entail behavioral observation and active integration by assessors of test scores and other data. In such situations, the term assessment may be preferable to testing. The term assessment acknowledges that tests are only one type of tool used by professional assessors, and that a test’s value is intimately linked to the knowledge, skill, and experience of the assessor. As Sundberg and Tyler (1962) observed, “Tests are tools. In the hands of a fool or an unscrupulous person they become pseudoscientific perversion” (p. 131, emphasis in the original). In most evaluation contexts, it is the process of assessment that breathes life and meaning into test scores.

Psychological Assessment, a measurement textbook by Maloney and Ward (1976), echoed the uneasiness of psychologists with the anachronistic use of “psychological testing” to describe their many varied assessment-related activities. By articulating several differences between testing and assessment, Maloney and Ward clarified the rich texture of the thoughtful, problem-solving processes of psychological assessment—“unclumping” it from the more technician-like tasks of psychological testing.

Maloney and Ward conceived of assessment as a problem-solving process that could take many different forms. How an assessment proceeds depends on many factors, not the least of which is the reason for assessing. Different tools of evaluation—psychological tests among them—might be marshaled in the process of assessment, depending on the particular objectives, people, and circumstances involved as well as on other variables unique to the particular situation. By contrast, psychological testing was seen as much narrower in scope, referring only to “the process of administering, scoring, and interpreting psychological tests” (Maloney & Ward, 1976, p. 9). The examiner is more key to the process of assessment, in which decisions, predictions, or both are made on the basis of many possible sources of data (including tests).

Maloney and Ward also distinguished testing from assessment in regard to their respective objectives. In testing, a typical objective is to measure the magnitude of some psychological trait or attribute. For example, one might speak of intelligence testing if the purpose of administering a test was confined to obtaining a numerical gauge of the intelligence of a testee or group of testees. In assessment, which is always conducted on a one-to-one basis, the objective more typically extends beyond obtaining a number. In this context, it should not come as a surprise that the use of the term intelligence test may be out of vogue. Certainly this seems the trend among the folks who create and develop the major instruments to measure intelligence.
Published in 2002, the third edition of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III, Wechsler, 2002) was introduced in its manual as “an individually administered clinical instrument for assessing the intelligence of children” (p. 1). The fifth edition of the Stanford-Binet (SB5, Roid, 2003a) was introduced by its author, Gale H. Roid (2003b, p. 2) as “an individually administered assessment of intelligence and cognitive abilities.” The fourth edition of the Wechsler Intelligence Scale for Children (WISC-IV, Wechsler, 2003) was introduced as “an individually administered, comprehensive clinical instrument for assessing the intelligence of children” (p. 1). In each of these three introductory self-descriptions, assessment or assessing is a key word, and the word test is notable for its absence.

The term assessment is preferable to testing for various evaluation situations. Consider, for example, an evaluation of a student’s intelligence designed to answer referral questions about the student’s ability to function in a regular classroom. Such an evaluation might explore not only the student’s intellectual strengths and weaknesses but also social skills and judgment. By contrast, testing “could take place without being directed at answering a specific referral question and even without the tester actually seeing the client or testee” (Maloney & Ward, 1976, p. 9).

In testing, a tester will typically add up “the number of correct answers or the number of certain types of responses . . . with little if any regard for the how or mechanics of such content” (Maloney & Ward, 1976, p. 39). Assessment is more apt to focus on how the individual processes rather than the results of that processing. Thus, very different goals and purposes are served.

Regarding the collection of psychological assessment data, Maloney and Ward (1976) urged that, far beyond the use of psychological tests alone, “literally, any method the examiner can use to make relevant observations is appropriate” (p. 7). Years later, Roberts and Magrab (1991) argued that assessment was not an activity to be confined to the consulting room. For them, assessment involved less emphasis on the measurement of the strength of traits and more emphasis on the understanding of problems in their social contexts. To achieve such understanding, assessment might entail routine home visits or other community observations.

The semantic distinction between psychological testing and psychological assessment is blurred in everyday conversation, even in many published textbooks that make little distinction between the two terms. Yet the distinction is important. Society at large is best served by clear definition of and differentiation between these two terms as well as related terms such as psychological test user and psychological assessor. In the section “Test-User Qualifications” in Chapter 2, the point is made that clear distinctions between such terms not only serves the public good but might also help avoid the turf wars now brewing between psychology and various users of psychological tests. Admittedly, the line between what constitutes testing and what constitutes assessment is not always as clear as we might like it to be. However, by acknowledging that such ambiguity exists, we can work to sharpen our definition and use of these terms; denying or ignoring their distinctiveness provides no hope of a satisfactory remedy.

We define psychological assessment as the gathering and integration of psychology-related data for the purpose of making a psychological evaluation, accomplished through the use of tools such as tests, interviews, case studies, behavioral observation, and specially designed apparatuses and measurement procedures. We define psychological testing as the process of measuring psychology-related variables by means of devices or procedures designed to obtain a sample of behavior.

JUST THINK . . .

Describe a situation in which testing is more appropriate than assessment. Then describe a situation in which assessment is more appropriate than testing.
The process of assessment  In general, the process of assessment begins with a referral for assessment from a source such as a teacher, a school psychologist, a counselor, a judge, a clinician, or a corporate human resources specialist. Typically, one or more referral questions are put to the assessor about the assessee. Some examples of referral questions are “Can this child function in a regular classroom?” “Is this defendant competent to stand trial?” and “How well can this employee be expected to perform if promoted to an executive position?”

The assessor may meet with the assessee or others before the formal assessment to clarify aspects of the reason for referral. Then comes the formal assessment, during which tests and other tools will typically be employed by the assessor to help answer the referral question(s). After the assessment, the assessor writes a report of the findings. More personal feedback sessions with the assessee and/or interested third parties (such as the assessee’s parents and the referring professional) may also be scheduled.

Different assessors may approach the assessment task in different ways. Some assessors approach the assessment with minimal input from assessees themselves. In this approach to assessment, the assessor’s primary focus is on test scores, interview data, case history data, and other available data derived from the formal assessment. Other assessors view the process of assessment as more of a collaboration between the assessor and the assessee. For example, in the process of collaborative psychological assessment described by Constance Fischer (1978), the assessor and assessee may work as “partners” from initial contact through final feedback. In this approach, the assessee is viewed as “an expert about his or her current views and remembered life events” (Fischer, 2004, p. 14).

Another variety of collaborative assessment may include an element of therapy as part of the process. Stephen Finn and his colleagues (Finn, 2003; Finn & Martin, 1997; Finn & Tonsager, 2002) have described therapeutic psychological assessment as an approach that encourages therapeutic self-discovery and new understandings through the assessment process. A term increasingly used with regard to testing and assessment in the schools is dynamic assessment. Dynamic psychological assessment may be defined as a model and philosophy of interactive evaluation involving various types of assessor intervention during the assessment process. For example, an assessor may intervene with increasingly more explicit prompts, feedback, or hints in order to not only evaluate what the assessee knows but to effectively modify and improve the way the assessee thinks about the problem or subject matter. Although aspects of the dynamic assessment model have been written about at least since the 1920s (Lidz, 1987), it was not until the 1970s and 1980s that a number of tools incorporating this approach were published (Lidz, 1991, 1996).

Alternate assessment  The Individuals with Disabilities Education Act (IDEA) Amendments, PL 105-17, became law in 1997. Many of the provisions of the IDEA amendments are discussed elsewhere in this book. For now, let’s focus on a section of this law that introduces the term alternate assessment. Specifically, this section provides that the State or local educational agency “(i) develops guidelines for the participation of children with disabilities in alternate assessments for those children who cannot participate in State and district-wide assessment programs; and (ii) develops and . . . conducts those alternate assessments.”

PL 105-17 does not define “alternate assessments.” However, a look at past practice by assessors involved in evaluating students with special needs will illustrate the concept. For example, the student who has difficulty reading the small print of a particular test may be accommodated with a large-print version of the same test or with a specially lit test environment. A student with a hearing impairment may be administered the test
in sign language. A child with attention deficit disorder (ADD) might have an extended evaluation time, with frequent breaks during periods of evaluation.

So far, the process of alternate assessment may seem fairly simple and straightforward; in practice, however, it may be anything but. Consider, for example, the case of a student with a vision impairment who is scheduled to be given a written, multiple-choice test using an alternate procedure. There are several possible alternate procedures. For instance, the test could be translated into Braille and administered in that form, or it could be administered by means of audiotape. Whether the test is administered by Braille or audiotape may affect the test scores; some students may do better with a Braille administration and others with audiotape. Students with superior short-term attention and memory skills for auditory stimuli would seem to have an advantage with the audiotaped administration. Students with superior haptic (sense of touch) and perceptual-motor skills might have an advantage with the Braille administration.

Some alternative methods may take the form of performance-based tasks rather than paper-and-pencil tasks. For example, students whose math skills cannot be assessed by the administration of paper-and-pencil questions might be evaluated through tasks such as making change or making purchases in a “real-life” context. Another alternative method of assessment entails the evaluation of a collection of the assessee’s work samples over time.

A number of important questions can be raised about the equivalence of various alternate and traditional assessments. To what extent does each method really measure the same thing? How equivalent is the alternate test to the original test? How does modifying the format of a test, the time limits of a test, or any other aspect of the way a test was originally designed to be administered, affect test scores? And taking a step back from such complex issues, how shall we define alternate assessment?

Keeping in mind the complexities involved, we propose this definition of this somewhat elusive process: Alternate assessment is an evaluative or diagnostic procedure or process that varies from the usual, customary, or standardized way a measurement is derived, either by virtue of some special accommodation made to the assessee or by means of alternative methods designed to measure the same variable(s). This definition avoids the thorny issue of equivalence of methods. Unless the alternate procedures have been thoroughly researched, there is no reason to expect them to be equivalent. In most cases, because the alternate procedures have been individually tailored, there is seldom compelling research to support equivalence. Governmental guidelines for alternate assessment will evolve to include ways of translating measurement procedures from one format to another. Other guidelines may suggest substituting one tool of assessment for another.

All this talk about assessment might lead one to wonder how assessments are typically conducted and what tools are used. Before reading on, however, try the “Just Think” exercise.

The Tools of Psychological Assessment

The test  A test may be defined simply as a measuring device or procedure. When the word test is prefaced with a modifier, it refers to a device or procedure designed to measure a variable related to that modifier. Consider, for example, the term medical test, which refers to a device or procedure designed to measure some variable related to the practice of medicine (including a wide range of tools and procedures such as X-rays, blood tests, and testing of reflexes). In a like manner, the term psychological test refers
to a device or procedure designed to measure variables related to psychology (for example, intelligence, personality, aptitude, interests, attitudes, and values). And whereas a medical test might involve the analysis of a sample of blood, tissue, or the like, a psychological test almost always involves the analysis of a sample of behavior. The behavior sample could range from responses to a pencil-and-paper questionnaire to oral responses to questions to performance of some task. The behavior sample could be elicited by the stimulus of the test itself, or it could be naturally occurring behavior (under observation).

Psychological tests and other tools of assessment may differ on a number of variables such as content, format, administration procedures, scoring and interpretation procedures, and technical quality. The content (subject matter) of the test will, of course, vary with the focus of the particular test. But even two psychological tests purporting to measure the same thing—for example, personality—may differ widely in item content because of factors such as the test developer’s definition of personality and the theoretical orientation of the test. For example, items on a psychoanalytically oriented personality test may have little resemblance to those on an existentially oriented personality test, yet both are personality tests.

The term format pertains to the form, plan, structure, arrangement, and layout of test items as well as to related considerations such as time limits. Format is also used to refer to the form in which a test is administered: computerized, pencil-and-paper, or some other form. When making specific reference to a computerized test, format may further refer to the form of the software: IBM- or Apple-compatible. The term format is not confined to tests; it is also used to denote the form or structure of other evaluative tools and processes, such as the specific procedures used in obtaining a particular type of work sample.

Tests differ in their administration procedures. Some tests, particularly those designed for administration on a one-to-one basis, may require a very active and knowledgeable test administrator. The test administration may involve demonstration of various kinds of tasks on the part of the assessees as well as trained observation of an assessees’s performance. Alternatively, some tests, particularly those designed for administration to groups, may not even require the test administrator to be present while the testtakers independently do whatever it is the test requires.

Tests differ in their scoring and interpretation procedures. To better understand how and why, let’s define score and scoring. Sports enthusiasts are no strangers to these terms. For them, these terms refer to the number of points accumulated by competitors and the process of accumulating those points. In testing and assessment, we may formally define score as a code or summary statement, usually but not necessarily numerical in nature, that reflects an evaluation of performance on a test, task, interview, or some other sample of behavior. Scoring is the process of assigning such evaluative codes or statements to performance on tests, tasks, interviews, or other behavior samples. As we will see in the chapters that follow, there are many different types of scores. Some scores result from the simple summing of responses (such as the summing of correct/incorrect or agree/disagree responses), and some scores result from the application of more elaborate procedures.

Scores themselves can be described and categorized in many different ways. Here, let’s consider one such category of scores, the cut score. A cut score (also referred to as a cutoff score or simply a cutoff) is a reference point, usually numerical, derived by judgment and used to divide a set of data into two or more classifications. Some action will be taken or some inference will be made on the basis of these classifications. Cut scores on tests, usually in combination with other data, are used in schools in many contexts, such as grading and making decisions about the class or program to which a particular child will be assigned. Cut scores are used by employers as aids to decision making.
about personnel hiring and advancement. State agencies use cut scores to help determine who shall be licensed as a professional in a given field. There are probably upwards of 20 different methods that can be used to formally derive cut scores (Dwyer, 1996).

Sometimes, no formal method is used to arrive at a cut score. Some teachers use an informal, “eyeball” method to proclaim, for example, that a score of 65 or more on a test means “pass” and a score of 64 or below means “fail.” Whether formally or informally derived, cut scores typically take into account, to at least some degree, the values of those who set them. Further, there is another side to the human equation as it relates to cut scores, one seldom written about in measurement texts. Human judgment is very much a part not only of setting cut scores but of reacting to them. Some consequences of being “cut” by cut scores have been explored in innovative research; see Figure 1-1.

Tests differ widely in terms of their guidelines for scoring and interpretation. Some tests are designed to be scored by testtakers themselves, and others are designed to be scored by trained examiners. Still other tests may be scored and fully interpreted in
seconds by computer. Some tests, such as most tests of intelligence, come with test manuals that are very explicit not only about scoring criteria but also about the nature of the interpretations that can be made from the calculated score. Other tests, such as the Rorschach Inkblot Test (discussed in Chapter 12), are sold with no manual at all. The (qualified) purchaser buys the stimulus materials and then selects and uses one of many available guides for administration, scoring, and interpretation.

Tests differ with respect to their technical quality. More commonly, reference is made to what is called the psychometric soundness of a test. Synonymous with the antiquated term psychometry, psychometrics may be defined as the science of psychological measurement.1 The adjective psychometric refers to measurement of a psychological nature. And the psychometric soundness of a test is a reference to how consistently and how accurately a psychological test measures what it purports to measure.

We have much more to say about what constitutes quality in a test or other tool of assessment. Throughout this book, consistent with common practice, we sometimes use the word test (as well as related terms such as test score) in a generic sense when discussing general principles applicable to various measurement procedures. These measurement procedures range from those widely labeled as tests (such as paper-and-pencil examinations) to procedures that measurement experts might label with more specific terms (such as situational performance measures). But now let’s get back to the tools of assessment and introduce one that probably, as they say, “needs no introduction.”

The interview Another widely used tool in psychological assessment is the interview, a word that may conjure images of face-to-face talk. But the interview as a tool of psychological assessment involves more than talk. If the interview is conducted face to face, the interviewer probably notes nonverbal as well as verbal behavior, such as the interviewee’s dress, manner, and eye contact. An interview may be conducted over the telephone, in which case the interviewer might make inferences about what is said as a function of changes in the interviewee’s voice quality. Interviews need not involve speech, as when they are conducted in sign language. Interviews may be conducted by means of electronic media, such as e-mail. In its broadest sense, then, we can define an interview as a method of gathering information through direct communication involving reciprocal exchange.

Interviews differ with regard to many variables, such as their purpose, their length, or other restrictions under which they are conducted, and the willingness of the interviewee to provide information candidly. Interviews may be used by psychologists and others in clinical, counseling, forensic, or neuropsychological settings to help make diagnostic or treatment decisions. School psychologists and others in educational settings may use interviews to help make decisions about the appropriateness of various educational interventions or class placements. An interview may be used to help human resources professionals make more informed recommendations about the hiring, firing, and advancement of personnel. In some instances, the process takes the form of a panel interview, wherein more than one interviewer participates in the assessment of personnel. A presumed advantage of the panel interview (also sometimes referred to as a board interview) is that any idiosyncratic biases of a lone interviewer will be minimized by the use of two or more interviewers (Dipboye, 1992). A disadvantage of the panel interview

JUST THINK . . .

What are the strengths of the interview as a tool of assessment? What are the weaknesses of the interview as a tool of assessment?

1. Variants of these words include the adjective psychometric and the nouns psychometrist and psychometrician. Traditionally, a psychometrist holds a master’s degree and is qualified to administer specific tests. A psychometrician holds a doctoral degree in psychology or some related field (such as education) and specializes in areas such as individual differences, quantitative psychology, or theory of assessment.
Interviews are used by psychologists who study consumer behavior to answer corporate America’s questions about the market for various products and services and how best to advertise and promote them. Researchers in psychology and related fields use interviews to explore myriad other topics. A casual survey of the literature reveals recent research employing the interview to explore topics as diverse as food choice negotiation by newly married couples (Bove et al., 2003), the experience of hearing voices from the perspective of those who hear them (Jones et al., 2003), and a conception of what constitutes “masculinity” from the perspective of teenage boys (Pascoe, 2003).

The popularity of the interview as a method of gathering information extends far beyond psychology. Just try to think of one day when you were not exposed to an interview on television, radio, or the Internet! Regardless of the forum, the quality, if not the quantity, of useful information produced by an interview depends to some degree on the interviewer. An interview is a reciprocal affair. The interviewee reacts to the interviewer, and the interviewer reacts to the interviewee. Interviewers differ in many ways; for example, their pacing of interviews, their rapport with interviewees, and their ability to convey genuineness, empathy, and humor. With these differences between interviewers in mind, look at Figure 1-2. Think about how attributes of these well-known celebrities might affect responses of interviewees. More generally, think about other dimensions on

**Figure 1–2**
On Interviewing and Being Interviewed

Different interviewers have different styles of interviewing. How would you characterize the interview style of Howard Stern versus that of Jay Leno?

is the additional cost of using multiple interviewers, especially when the return on this investment is questionable (Dixon et al., 2002).

Interviews are used by psychologists who study consumer behavior to answer corporate America’s questions about the market for various products and services and how best to advertise and promote them. Researchers in psychology and related fields use interviews to explore myriad other topics. A casual survey of the literature reveals recent research employing the interview to explore topics as diverse as food choice negotiation by newly married couples (Bove et al., 2003), the experience of hearing voices from the perspective of those who hear them (Jones et al., 2003), and a conception of what constitutes “masculinity” from the perspective of teenage boys (Pascoe, 2003).

The popularity of the interview as a method of gathering information extends far beyond psychology. Just try to think of one day when you were not exposed to an interview on television, radio, or the Internet! Regardless of the forum, the quality, if not the quantity, of useful information produced by an interview depends to some degree on the interviewer. An interview is a reciprocal affair. The interviewee reacts to the interviewer, and the interviewer reacts to the interviewee. Interviewers differ in many ways; for example, their pacing of interviews, their rapport with interviewees, and their ability to convey genuineness, empathy, and humor. With these differences between interviewers in mind, look at Figure 1-2. Think about how attributes of these well-known celebrities might affect responses of interviewees. More generally, think about other dimensions on

**JUST THINK . . .**

What types of interviewing skills must the host of a talk show possess to be considered an effective interviewer? Do these skills differ from those needed by a professional in the field of psychological assessment?
which you might characterize interviewers you have seen and come to know in the
media. “Juvenile versus adult” and “eager-to-speak versus eager-to-listen” are two such
dimensions that may come to mind.

The portfolio In recent years, the popularity of portfolio (work sample) assessment in
many fields, including education, has been rising. Some have argued, for example, that
the best evaluation of a student’s writing skills can be accomplished not by the adminis-
tration of a test but by asking the student to compile a selec-
tion of writing samples. From the perspective of education administrators, portfolio assessment also has distinct
advantages in assessing the effectiveness of teachers. By exam-
ining teachers’ portfolios and seeing how teachers ap-
proach their coverage of various topics, educational evalu-
ators have another tool to help anchor judgments to work
samples.

Case history data In a general sense, case history data refers to records, transcripts, and
other accounts in written, pictorial, or other form, in any media, that preserve archival
information, official and informal accounts, and other data and items relevant to an as-
seesee. Case history data may include files or excerpts from files maintained at institu-
tions and agencies such as schools, hospitals, employers, religious institutions, and
criminal justice agencies. Other examples of case history data are letters and written cor-
respondence; photos and family albums; newspaper and magazine clippings; and home
videos, movies, and audiotapes. Work samples, artwork, doodlings, and accounts and
pictures pertaining to interests and hobbies are yet other examples.

Case history data can be a very useful tool in a wide variety of assessment contexts.
In a clinical evaluation, for example, case history data can shed light on an individual’s past and current adjustment, as well as the events and circumstances that may have con-
tributed to any changes in adjustment. Case history data can be of critical value in neu-
ropsychological evaluations, where it often provides information about neuropsycho-
logical functioning prior to the occurrence of a trauma or other event that results in a
deficit. School psychologists rely on case history data to, among other things, answer
questions about the course of a student’s developmental history.

Another use of the term case history, one synonymous with case study, has to do with
the assembly of case history data into an illustrative ac-
count. For example, a case study might detail how a num-
ber of aspects of an individual’s personality combined with
environmental conditions to produce a successful world
leader. A case study of an individual who attempted to as-
sassinate a high-ranking political figure might shed light
on what types of individuals and conditions might lead to
similar attempts in the future. A now classic work on the
subject of groupthink contains rich case history material on collective decision making
that did not always result in the best decisions (Janis, 1972).

Behavioral observation If you want to know how someone behaves in a particular situ-
ation, observe his or her behavior in that situation. Such “down-home” wisdom under-
lies at least one approach to evaluation. Behavioral observation as it is employed by as-
seessment professionals may be defined as monitoring the actions of others or oneself by
visual or electronic means while recording quantitative and/or qualitative information
regarding the actions. Behavioral observation may be used in a variety of settings for a
variety of assessment objectives. It may be used, for example, as a diagnostic aid in a clinical setting or as a means of data collection in basic research. Observations may be made in laboratory or otherwise structured settings. An example of this is a researcher’s observation of a child who is asked to perform some task as part of an experiment. Observation may also occur in the natural setting in which the behavior would typically be elicited or expected to occur. This variety of behavioral observation is referred to as **naturalistic observation**.

Behavioral observation as an aid to designing therapeutic intervention has proven extremely useful in institutional settings such as schools, hospitals, prisons, and group homes. Using published or self-constructed lists of targeted behaviors, staff can observe firsthand the behavior of the person under observation and design interventions accordingly. In a school situation, for example, naturalistic observation on the playground of a culturally different child suspected to have linguistic problems might reveal that the child does have English language skills but is unwilling—for reasons of shyness, cultural upbringing, or whatever—to demonstrate those abilities to adults.

Despite the potential value of behavioral observation, it tends to be used infrequently outside institutional settings. For private practitioners, it is typically not economically feasible to spend hours out of the consulting room observing clients. Just think about the time it would take to administer a test of grocery-shopping skills if the assessee/shopper required a couple of price checks (see Figure 1-3).

**Figure 1–3**

**Price (and Judgment) Check in Aisle 5**

*Hamera and Brown (2000) described the development of a context-based Test of Grocery Shopping Skills. Designed primarily for use with persons with psychiatric disorders, this assessment tool may be useful in evaluating a skill necessary for independent living.*

**JUST THINK . . .**

What are the strengths of behavioral observation as a tool of assessment? What are the weaknesses of behavioral observation as a tool of assessment?
Role-play tests  If you have ever enjoyed the television program Whose Line Is It Anyway, you may appreciate just how entertaining improvisation can be. Beyond entertainment, however, improvisational acting has a place in the context of psychological assessment. In this context, role play may be defined as acting an improvised or partially improvised part in a simulated situation. A role-play test is a tool of assessment wherein assessees are directed to act as if they were in a particular situation. Assessees may then be evaluated with regard to their expressed thoughts, behaviors, abilities, and other variables. (Note that role play is hyphenated when used as an adjective or a verb but not as a noun.)

An individual being evaluated in a corporate, industrial, organizational, or military context for managerial or leadership ability, for example, might be asked to mediate a hypothetical dispute between personnel at a work site. The context of the role play may be created by various techniques ranging from live actors to computer-generated simulation. Outcome measures for such an assessment might include ratings related to various aspects of the individual’s ability to resolve the conflict, such as effectiveness of approach, quality of resolution, and number of minutes to resolution.

Role play as a tool of assessment may be used in various clinical contexts. For example, it is routinely employed in many interventions with substance abusers. Clinicians may attempt to obtain a baseline measure of abuse, cravings, or coping skills by administering a role-play test prior to therapeutic intervention and then again at the completion of a course of treatment.

Computers as tools  Professionals who specialize in psychological and educational assessment have long recognized the value of computers in administering, scoring, and interpreting tests. As early as 1930, electromechanical scoring was available for at least one psychological test, the Strong Vocational Interest Blank (SVIB) (Campbell, 1971). By 1946, thanks to the efforts of a Minneapolis engineer named Elmer Hankes, profiling of the SVIB could be done by machine. And by the late 1950s, computers were used not only for scoring and profiling but also for test interpretation (Rome et al., 1965). With the advent of the personal computer in the 1970s, office-based test administration, scoring, and interpretation became reality. As technology has flourished, the use of computers has burgeoned.

Today, computers, whether desktop, laptop, or palm-held, are a part of the essential office of clinicians, consultants, and other test users. From the standpoint of test users, computer-assisted psychological assessment (CAPA) refers to the convenience and economy of time in administering, scoring, and interpreting tests. Thus, the “assistance” in the term computer-assisted is assistance to test users, not testtakers. CAPA allows testtakers to work independently, responding to items presented on a video screen. The computer may then score the test, analyze response patterns, and even provide some sort of report (see Close-up).

For many test users, CAPA represents a great advance over the not-too-distant past when they had to personally administer tests and possibly even place the responses in some other form for analysis (manually using a scoring template or other device) before beginning the often laborious tasks of scoring and interpreting the resulting data. CAPA opened a world of possibilities for test users, enabling the building into tests of complex scoring and data combination strategies that would not otherwise be practical. CAPA has also enabled the measurement of abilities that could not be measured by more
Types of Computer-Generated Psychological Reports

Have you ever taken a test and had the results given in a computer-generated report? What type of report was it? And how did it get there?

Computer-generated psychological reports may be categorized as scoring reports, interpretive reports, or integrative reports. Here we define each of these types of reports and describe the type of information that might be found in them.

**Scoring Reports**

In general, a **scoring report** may be defined as a formal or official computer-generated account of test performance, usually presented in numeric form. One type of scoring report, a **simple scoring report**, simply lists test scores. Another type of scoring report is an **extended scoring report**. Beyond a simple listing of test scores, an extended scoring report may contain more detailed information, such as a statistical analysis of how test-takers performed on individual items.

**Interpretive Reports**

In general, an **interpretive report** is a computer-generated account of test performance, presented in numeric as well as narrative form, including an explanation of the findings. There are three varieties of interpretive report: a descriptive report, a screening report, and a consultative report.

A **descriptive report** is a type of interpretive report that features brief narrative summaries. In fact, the "description" in a descriptive report may be so brief as to amount to a one-sentence comment regarding where a particular score stands from a normative perspective. Descriptive reports can help a test user determine which of many scores on a test need to be focused on.

A **screening report** provides more information than a descriptive report but less than a consultative report. It provides narrative information as well as analysis or commentary regarding relationships between the scores. As its name implies, a screening report is particularly useful for screening purposes. Programmed into the software are various criteria that must be met before the software causes a line of narrative text to be automatically printed on the report.

More than a bare-bones descriptive report and less tentative in its conclusions than a screening report is a consultative report. A **consultative report** provides a detailed analysis of test data in language appropriate for communication between assessment professionals. It provides the expert opinion of an individual or group of individuals who may have devoted years of study to the interpretation of a particular instrument.

**Integrative Reports**

An **integrative report** provides a level of description and analysis found in interpretive reports but integrates into the report data from other sources such as behavioral observations or medication records. From a report that integrates behavioral observation data with medication records, for example, a clinician might receive valuable assistance concerning optimal medications and dosages for a client.

**CAPA Processing**

Regardless of its nature, a report may be created in different ways at different sites. Here is a “short course” in CAPA processing.

The term **central processing** is used to refer to the sending of test protocols completed on paper or some other form at one location to some other central location for scoring or interpretation. The results may then be returned to the test user by e-mail, disk, mail, fax, or telephone.

One variety of central processing is **teleprocessing**. **Teleprocessing** refers to the computerized scoring, interpretation, or other conversion of raw test data that have been sent for processing over telephone or wireless lines from the test site to a central facility.

**Local processing** may be defined simply as on-site computerized scoring, interpretation, or other conversion of raw test data. With the appropriate hardware and software, the test user may use the same computer to administer a test and then score it. In an era of relatively inexpensive computer hardware, most tests today are locally processed.

**JUST THINK . . .**

What are the pros and cons of the various types of CAPA processing?
traditional methods (Figure 1-4). Of course, every rose has its thorns; some of the pros and cons of CAPA are presented in Table 1-1.

Other tools Varied instruments can be applied as tools of assessment. Psychologists and others who devise tools to assess people with disabilities and members of other special populations have been most innovative. For example, Wilson et al. (1982) developed a mechanism for test response involving a dental plate activated by the tongue. Useful to testtakers who lack the capacity for speech or control of their limbs, the device permits five kinds of response.

On your next trip to the video store to rent a DVD, take a moment to consider the role that video can play in assessment. In fact, specially created videos are widely used in training and evaluation contexts. For example, corporate personnel may be asked to respond to a variety of video-presented incidents of sexual harassment in the workplace. Police personnel may be asked about how they would respond to various types of emergencies, which are presented either as reenactments or as video recordings of actual occurrences. Psychotherapists may be asked to respond with a diagnosis and a treatment plan for each of several clients presented to them on videotape. The list of potential applications of video to assessment is endless.

Figure 1–4
A Method of Quantifying Back Stress

The innovative application of computer technology has facilitated the measurement of traits or abilities by techniques that could not be measured by more traditional methods. For example, Mirka et al. (2000) described an assessment methodology that employs video, computer, and other components to obtain continuous assessment of back stress. It involves capturing an image with a video camera (in this illustration, the act of sawing at ground level), computerized representation of the action, and laboratory simulation.

JUST THINK . . .
In general, when is assessment using videotape a good idea? What are the drawbacks, if any, to using videotape in assessment?
The list of tools used in the service of psychological assessment includes, for example, many tools traditionally associated with medical health, such as thermometers to measure body temperature and gauges to measure blood pressure. Biofeedback equipment is sometimes used to obtain measures of bodily reactions (such as muscular tension or galvanic skin response) to various sorts of stimuli. An instrument called a penile plethysmograph, which gauges male sexual arousal, may be very helpful in the diagnosis and treatment of sexual predators. Impaired ability to identify odors is common in many disorders in which there is central nervous system involvement, and simple tests of smell may be administered to help determine if such impairment is present. In general, there has been no shortage of innovation on the part of psychologists in devising measurement tools, or adapting existing tools, for use in psychological assessment.

To this point, our introduction has focused on some basic definitions and a look at some of the “tools of the trade.” We now raise some very fundamental questions regarding the who, what, why, and where of testing and assessment.

Table 1-1
CAPA: Some Pros and Cons

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPO saves professional time in test administration, scoring, and interpretation.</td>
<td>Professionals must still spend significant time reading software and hardware documentation and even ancillary books on the test and its interpretation.</td>
</tr>
<tr>
<td>CAPO results in minimal scoring errors resulting from human error or lapses of attention or judgment.</td>
<td>With CAPO, the possibility of software or hardware error is ever present, from difficult-to-pinpoint sources such as software glitches or hardware malfunction.</td>
</tr>
<tr>
<td>CAPO assures standardized test administration to all testtakers with little, if any, variation in test administration procedures.</td>
<td>CAPO leaves those testtakers at a disadvantage who are unable to employ familiar test-taking strategies (previewing test, skipping questions, going back to previous question, etc.).</td>
</tr>
<tr>
<td>CAPO yields standardized interpretation of findings due to elimination of unreliability traceable to differing points of view in professional judgment.</td>
<td>CAPO’s standardized interpretation of findings based on a set, unitary perspective may not be optimal; interpretation could profit from alternative viewpoints.</td>
</tr>
<tr>
<td>Computers’ capacity to combine data according to rules is more accurate than that of humans.</td>
<td>Computers lack the flexibility of humans to recognize the exception to a rule in the context of the “big picture.”</td>
</tr>
<tr>
<td>Nonprofessional assistants can be used in the test administration process, and the test can typically be administered to groups of testtakers in one sitting.</td>
<td>Use of nonprofessionals leaves diminished, if any, opportunity for the professional to observe the assesse’s testtaking behavior and note any unusual extra-test conditions that may have affected responses.</td>
</tr>
<tr>
<td>Professional groups such as APA develop guidelines and standards for use of CAPO products.</td>
<td>Profit-driven nonprofessionals may also create and distribute tests with little regard for professional guidelines and standards.</td>
</tr>
<tr>
<td>Paper-and-pencil tests may be converted to CAPO products with consequential advantages, such as a shorter time between the administration of the test and its scoring and interpretation.</td>
<td>The use of paper-and-pencil tests that have been converted for computer administration raises questions about the equivalence of the original test and its converted form.</td>
</tr>
<tr>
<td>Security of CAPO products can be maintained not only by traditional means (such as locked filing cabinets) but by high-tech electronic products (such as firewalls).</td>
<td>Security of CAPO products can be breached by computer hackers, and integrity of data can be altered or destroyed by untoward events such as introduction of computer viruses.</td>
</tr>
<tr>
<td>Computers can automatically tailor test content and length based on responses of testtakers.</td>
<td>Not all testtakers take the same test or have the same test-taking experience.</td>
</tr>
</tbody>
</table>
Who, What, Why, and Where?

Who are the parties in the assessment enterprise? In what types of settings are assessments conducted? Why is assessment conducted? Where does one go for authoritative information about tests? Think about the answer to each of these important questions before reading on. Then check your own ideas against those that follow.

Who Are the Parties?

Parties in the assessment enterprise include developers and publishers of tests, users of tests, and people who are evaluated by means of tests. A fourth and frequently overlooked party is society at large.

The test developer  Test developers and publishers create tests or other methods of assessment. The American Psychological Association (APA) estimates that more than 20,000 new psychological tests are developed each year (APA, 1993). Among these new tests are some that were created for a specific research study, some that were created in the hope that they would be published, and some that represent refinements or modifications of existing tests. Test creators bring a wide array of backgrounds and interests to the test development process. For an intriguing glimpse at biographical information on a sampling of test developers, visit the “Test Developer Profiles” section of our Web site: www.mhhe.com/cohentesting6.

Test developers and publishers appreciate the significant impact that test results can have on people’s lives. Accordingly, a number of professional organizations have published standards of ethical behavior that specifically address aspects of responsible test development and use. Perhaps the most detailed document addressing such issues is one jointly written by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (NCME). Referred to by many psychologists simply as “the Standards,” Standards for Educational and Psychological Testing covers issues related to test construction and evaluation, test administration and use, and special applications of tests, such as special considerations when testing linguistic minorities. Initially published in 1954, revisions of the Standards were published in 1966, 1974, 1985, and 1999. The Standards is an indispensable reference work not only for test developers but for test users as well.

The test user  Tests are used by a wide range of professionals, including clinicians, counselors, human resources personnel, teachers, and other school personnel. The Standards, as well as the official guidelines of various other professional organizations, have much to impart to test users about how, why, and under what conditions tests should be used. For example, the principles of professional ethics promulgated by the National Association of School Psychologists (Jacob-Timm & Hartshorne, 1998) stress that school psychologists should select and use the test or tests that are most appropriate for each individual student. NASP (2000) further emphasizes that any questions that serve to prompt the psychological assessment of students should be answered in as comprehensive a manner as possible—that is, with as much background information and other data as possible, including data from behavioral observation.

No matter how sound a test is, its purpose will be defeated if the test user fails to competently manage all phases of the testing or assessment process. For this reason, a responsible test user has obligations before, during, and after a test is administered. Ethical guidelines dictate that before a test is administered, it should be stored in a way that
reasonably ensures that its specific contents will not be made known in advance. Another obligation of the test user before the test’s administration is to ensure that a prepared and suitably trained person administers the test properly. The test administrator (or examiner) must be familiar with the test materials and procedures and must have at the test site all the materials needed to properly administer the test. Materials needed might include a stopwatch, a supply of pencils, and a sufficient number of test protocols.²

In addition to having sufficient supplies, the examiner also ensures that the room in which the test will be conducted is suitable and conducive to the testing (Figure 1-5). To the extent that it is possible, distracting conditions such as excessive noise, heat, cold, interruptions, glaring sunlight, crowding, inadequate ventilation, and so forth should be avoided.

During test administration, and especially in one-on-one or small-group testing, rapport between the examiner and the examinee can be critically important. In the

---

² In everyday, nontest-related conversation, protocol refers to diplomatic etiquette. A less common use of the word is as a synonym for the first copy or rough draft of a treaty or other official document before its ratification. This second meaning comes closer to the way the word is used with reference to psychological tests. Protocol refers to the form or sheet on which the testtaker’s responses have been entered. Protocols are typically single sheets of paper or booklets.
context of the testing situation, **rapport** may be defined as a working relationship between the examiner and the examinee. Such a working relationship can sometimes be achieved with a few words of small talk when examiner and examinee are introduced. If appropriate, some words about the nature of the test and why it is important for examinees to do their best may also be helpful. In other instances, with a frightened child, the achievement of rapport might involve more elaborate techniques such as engaging the child in play or some other activity until the child has acclimated to the examiner and the surroundings. It is important that attempts to establish rapport with the testtaker not compromise any rules of the test administration instructions.

After a test administration, test users have many obligations as well. These obligations may range from safeguarding the test protocols to conveying the test results in a clearly understandable fashion. In between, there are other obligations such as those related to scoring the test. If a test is to be scored by people, scorers need to be in agreement about scoring criteria—even if that is not always the case in everyday, nontest, evaluation situations (see *Everyday Psychometrics*). Interpreting the test results and seeing to it that the test data are used in accordance with established procedures and ethical guidelines are additional obligations of test users.

**The testtaker** Testtakers approach an assessment situation in different ways, and test users must be sensitive to the diversity of possible responses to a testing situation. On the day of test administration, testtakers may vary on a continuum with respect to numerous variables, including:

- The amount of test anxiety they are experiencing and the degree to which that test anxiety might significantly affect the test results
- Their capacity and willingness to cooperate with the examiner or to comprehend written test instructions
- The amount of physical pain or emotional distress they are experiencing
- The amount of physical discomfort brought on by not having had enough to eat, having had too much to eat, or other physical conditions
- The extent to which they are alert and wide-awake as opposed to nodding off
- The extent to which they are predisposed to agreeing or disagreeing when presented with stimulus statements
- The extent to which they have received prior coaching
- The importance they may attribute to portraying themselves in a good (or bad) light
- The extent to which they are, for lack of a better term, “lucky” and can “beat the odds” on a multiple-choice achievement test (even though they may not have learned the subject matter).

In the broad sense in which we are using the term “testtaker,” anyone who is the subject of an assessment or an evaluation can be a testtaker or an assessee. As amazing as it sounds, this means that even a deceased individual can be considered an assessee. True, it’s the exception to the rule, but there is such a thing as a **psychological autopsy**. A **psychological autopsy** may be defined as a reconstruction of a deceased individual’s psychological profile on the basis of archival records, artifacts, and interviews previously conducted with the deceased assessee or people who knew him or her. For interested readers, a fascinating case study that employed the technique of psychological autopsy is presented by Neagoe (2000).
EVERYDAY PSYCHOMETRICS


The Motion Picture Association of America (MPAA) ratings—we have all heard them: “The following film is rated PG-13.” The PG-13 stands for “parents strongly cautioned” with respect to allowing children younger than 13 to see the movie. There’s also G for “general audiences,” PG for “parental guidance suggested,” and NC-17 for “no one 17 and under admitted.” Have you ever wondered who actually evaluates movies and gives them one of these five age-based ratings? It is actually a group of 8 to 13 parents who are employed full-time by MPAA to watch and evaluate movies. Some criteria used for rating films PG through NC-17 are published on the MPAA Web site, http://www.mpaa.org. There you will find, for example, that “an R-rated film may include hard language or tough violence, or nudity within sensual scenes, or drug abuse or other elements, or a combination of some of the above, so that parents are counseled, in advance, to take this advisory rating very seriously.”

The group of parents who do the rating sees a film submitted by a producer, discusses the film, and then votes on a rating. A majority vote wins. Beyond that description, few details regarding the actual rating process are provided on the Web site. We do know that regardless how the board of paid parents votes, their rating can be overturned by the two-thirds vote of an appeals board that is made up of 14 to 18 entertainment industry members.

If there is mystery about the process of rating movies, it is only the tip of the iceberg with regard to mystery about ratings in the entertainment industry in general. There are television rating systems presented with age-range recommendations and content summaries in icon form. There are rating systems for music, electronic software, and coin-operated video games. There is even a rating system for Internet content promulgated by the Internet Content Rating Association. Reviewing these systems and the literature available on them, one research team concluded:

The efforts of a variety of independent media industries have resulted in a dizzying array of ratings, icons, definitions, and procedures that are, in many cases, difficult to understand and remember. Almost all of these rating systems have been prompted by the threat of government intervention, and each industry has attempted to balance the provision of information against its own economic concerns. Although creating a rating system that works well for parents is not an easy task, it is clear that the preferences of parents have not often prevailed. (Bushman & Cantor, 2003, pp. 138–139)

As you learn more about psychological testing and assessment, think back on occasion to this description of evaluation in the entertainment industry. Contrast evaluation in that industry with evaluation in psychology. Think about issues. For example, think about issues regarding the definition of terms in psychological evaluation. How clearly is whatever is being measured by a psychological test defined? Think about the process of evaluation. For example, think about differences among raters or scorers and how these differences might figure into an ultimate rating or score. And think about the utility of the evaluations. In the entertainment industry, the “end user” is a parent or guardian making a decision regarding an entertainment-related choice for a minor. Who are the potential “end users” when it comes to psychological and educational evaluations? What types of decisions may have to be made on the basis of such information? How useful is the information developed in making these decisions?

Society at large

The uniqueness of individuals is one of the most fundamental characteristic facts of life. . . . At all periods of human history men have observed and described differences between individuals. . . . But educators, politicians, and administrators have felt a need for some way of organizing or systematizing the many-faceted complexity of individual differences. (Tyler, 1965, p. 3)
The societal need for “organizing” and “systematizing” has historically manifested itself in such varied questions as “Who is a witch?” “Who is schizophrenic?” and “Who is qualified?” The specific questions asked have shifted with societal concerns. The methods used to determine the answers have varied throughout history as a function of factors such as intellectual sophistication and religious preoccupation. Palmistry, podoscopy, astrology, and phrenology, among other pursuits, have had proponents who argued that the best means of understanding and predicting human behavior was through the study of the palms of the hand, the feet, the stars, bumps on the head, tea leaves, and so on. Unlike such pursuits, the assessment enterprise has roots in science. Through systematic and replicable means that can produce compelling evidence, the assessment enterprise responds to what Tyler (1965, p. 3) referred to as the societal “need for some way of organizing or systematizing the many-faceted complexity of individual differences.”

Other parties Beyond the four primary parties we have focused on here, let’s briefly make note of others who may participate in varied ways in the testing and assessment enterprise. Organizations, companies, and governmental agencies sponsor the development of tests for various reasons, such as to certify personnel. Companies and services offer test scoring or interpretation services. In some cases, these companies and services are simply extensions of test publishers, and in other cases they are independent. There are people whose sole responsibility is the marketing and sales of tests. Sometimes these people are employed by the test publisher; sometimes they are not. There are academicians who review tests and evaluate their psychometric soundness. All of these people, as well as many others, are also parties to a greater or lesser extent in the assessment enterprise, an enterprise that clearly has few boundaries in terms of the settings in which it can be found.

In What Types of Settings Are Assessments Conducted, and Why?

Educational settings You are probably no stranger to the many types of tests administered in the classroom. As mandated by law, tests are administered early in school life to help identify children who may need special services or accommodations. In addition to school ability tests, you are by now no stranger to achievement tests: evaluations of accomplishment or the degree of learning that has taken place. Some of the achievement tests you have taken in school were constructed by your teachers; others were constructed for more widespread use by other educators and measurement professionals. In the latter category, acronyms such as SAT and GRE may ring a bell (and if they do not, they will after you have read Chapter 10).

You know from your own experience that a diagnosis may be defined as a description or conclusion reached on the basis of evidence and opinion. Typically, this conclusion is reached through a process of distinguishing the nature of something and ruling out alternative conclusions. As its name implies, a diagnostic test is a tool of assessment used to help narrow down and identify areas of deficit to be targeted for intervention. Diagnostic tests of reading, mathematics, and other academic subjects may be administered in educational settings by teachers, school counselors, and school psychologists to assess the need for educational interventions as well as eligibility for special education programs.

Schoolchildren receive grades on their report cards that are not based on any formal assessment. For example, the grade next to “Works and plays well with others” is probably based more on the teacher’s informal evaluation in the classroom than on scores.
on any published measure of social interaction. We may define informal evaluation as a typically nonsystematic assessment that leads to the formation of an opinion or attitude.

Informal evaluation is, of course, not limited to educational settings; it is very much a part of everyday life. In fact, many of the types of tests we have listed as being administered in educational settings (achievement tests, diagnostic tests, etc.) are also administered in various other settings. And some of the types of tests we discuss in the context of the settings that follow are also administered in educational settings. Thus, it is important to remember that the tools discussed in one context may overlap tools discussed in another. At this point we are simply introducing a sampling of the types of tests used in different settings, not providing a comprehensive list.

**Geriatric settings** In the United States, more than 12 million adults are currently in the age range of 75 to 84; that is about 16 times more people in this age range than there were in 1900. Four million adults in the United States are currently 85 years old or older; that is a 33-fold increase in the number of people in that age range in 1900 (Administration on Aging, 1999). Clearly, people in the United States are living longer, and the population as a whole is getting older.

Older Americans may live at home, in special housing designed for independent living, in housing designed for assisted living, or in long-term care facilities such as hospitals and hospices. Wherever older individuals reside, they may at some point require psychological assessment to evaluate cognitive, psychological, adaptive, or other functioning.

**Counseling settings** Assessment in a counseling context may occur in environments as diverse as schools, prisons, and government- or privately-owned institutions. Regardless of the particular tools used, the ultimate objective of many such assessments is the improvement of the assessee in terms of adjustment, productivity, quality of life, or some related variable. Measures of social and academic skills and measures of personality, interest, attitudes, and values are among the many types of tests that a counselor might administer to a client. Referral questions to be answered range from “How can this child better focus on tasks?” to “For what career is the client best suited?” to “What activities are recommended for retirement?” Because the testtaker is in many instances the primary recipient and user of the data from a test administered by a counselor, it is imperative that the counselor understand the strengths and limitations of the findings and be able to competently convey the test results to the client.

**Clinical settings** Tests and many other tools of assessment are widely used in clinical settings such as public, private, and military hospitals, inpatient and outpatient clinics, private-practice consulting rooms, schools, and other institutions. These tools are used to help screen for or diagnose behavior problems. What types of situations might prompt the employment of such tools? Here’s a small sample.

- A private psychotherapy client wishes to be evaluated to see if the assessment can provide any nonobvious clues regarding his maladjustment.
- A school psychologist clinically evaluates a child experiencing learning difficulties to determine what factors are primarily responsible for it.
A psychotherapy researcher uses assessment procedures to determine if a particular method of psychotherapy is effective in treating a particular problem.

A psychologist-consultant retained by an insurance company is called on to give an opinion as to the reality of a client’s psychological problems; is the client really experiencing such problems, or just malingering?

A court-appointed psychologist is asked to give an opinion as to a defendant’s competency to stand trial.

A prison psychologist is called on to give an opinion regarding the extent of a convicted violent prisoner’s rehabilitation.

The tests employed in clinical settings may be intelligence tests, personality tests, neuropsychological tests, or other specialized instruments, depending on the presenting or suspected problem area. The hallmark of testing in clinical settings is that the test or measurement technique is employed with only one individual at a time. Group testing is used primarily for screening; that is, identifying those individuals who require further diagnostic evaluation. In Chapter 13 and elsewhere, we will look at the nature, uses, and benefits of assessment in both clinical and counseling settings.

Business and military settings  In business, as in the military, tests are used in many ways, perhaps most notably in decision making about the careers of personnel. As we will see in Chapter 16, a wide range of achievement, aptitude, interest, motivational, and other tests may be employed in the decision to hire, as well as in related decisions regarding promotions, transfer, job satisfaction, and eligibility for further training. For a prospective air traffic controller, successful performance on a test of sustained attention to detail may be one requirement of employment. For promotion to the rank of officer in the military, successful performance on a series of leadership tasks may be essential.

Another application of psychological tests involves the engineering and design of products and environments. Engineering psychologists employ a variety of existing and specially devised tests in research designed to help people at home, in the workplace, and in the military. Products ranging from home computers to office furniture to jet cockpit control panels benefit from the work of such research efforts.

Using tests, interviews, and other tools of assessment, psychologists who specialize in the marketing and sale of products are involved in taking the pulse of consumers—helping to predict the public’s receptivity to a new product, a new brand, or a new advertising or marketing campaign.

Governmental and organizational credentialing  One of the many applications of measurement is in governmental licensing, certification, or general credentialing of professionals. Before they are legally entitled to practice medicine, physicians must pass an examination. Law school graduates cannot hold themselves out to the public as attorneys until they pass their state’s bar examination. Psychologists, too, must pass an examination entitling them to present themselves to the public with the title “psychologist.”

Members of some professions have formed organizations with requirements for membership that go beyond those of licensing or certification requirements. For example, physicians can take further specialized training and a specialty examination to earn the distinction of being “board certified” in a particular specialty area of medicine. Psychologists specializing in certain areas may be evaluated for a diploma from the American Board of Professional Psychology (ABPP) to recognize excellence in the practice of psychology. Another organization, the American Board of Assessment Psychology (ABAP), awards its diploma on the basis of an examination to test users, test
developers, and others who have distinguished themselves in the field of testing and assessment.

**Other settings**  Many different kinds of measurement procedures find application in a wide variety of settings. For example, the courts rely on psychological test data and related expert testimony as one source of information to help answer important questions such as “Is this defendant competent to stand trial?” and “Did this defendant know right from wrong at the time the criminal act was committed?”

Measurement may play an important part in program evaluation, be it a large-scale government program or a small-scale, privately funded one. Is the program working? How can the program be improved? Are funds being spent in the areas where they ought to be spent? How sound is the theory on which the program is based? These are the types of general questions that tests and measurement procedures used in program evaluation are designed to answer.

Tools of assessment can be found in use in research and practice in every specialty area within psychology. For example, consider health psychology, a specialty area that focuses on understanding the role of psychological variables in the onset, course, treatment, and prevention of illness, disease, and disability (Cohen, 1994). Health psychologists are involved in teaching, research, or direct-service activities designed to promote good health. Individual interviews, surveys, and paper-and-pencil tests are some of the tools that may be employed to help assess a current state of affairs with regard to some disease or condition, gauge treatment progress, and evaluate outcome of intervention.

One research approach in health psychology entails reporting on the nature of the psychological adjustment, the nature of coping measures, or the nature of the quality of life of members of targeted groups. Various measures of adjustment, coping, and quality of life may be employed in research with a wide variety of populations ranging from middle-aged women who have just given birth to elderly men afflicted with debilitating medical conditions. Another general line of research in health psychology focuses on aspects of personality, behavior, or lifestyle as they relate to variables ranging from good physical health and longevity to sudden death by heart attack. For example, Hill and Pargament (2003) reviewed advances in the measurement of spirituality and the possible implications of those advancements for physical and mental health. Using a test called the Drinking Motives Measure (DMM), Martens et al. (2003) studied college athletes’ motivations for alcohol use. Consistent with prior research, these investigators concluded that athletes involved in intercollegiate sports may be particularly susceptible to using alcohol as well as other drugs as a coping mechanism, due to elevated stressors. The researchers concluded that the DMM was effective in predicting alcohol consumption and might therefore have an application in prevention or intervention programs.

What personality traits, if any, are predictive of smoking initiation and cessation? Compliance or noncompliance with physicians’ instructions? Strengthened or compromised immune functioning in AIDS patients? These questions are representative of many asked by health psychologists. All such questions require sound techniques of evaluation if answers are to be forthcoming.

Of course, psychological testing and assessment is not confined to health psychology. It is very much a part of all specialty areas within psychology and education. Further, what constitutes a “test” may take many different forms ranging from paper-and-pencil form to . . . well, just take a look at Figure 1-6. There you will find a very small sample of the tens of thousands of measurement methods that have been used in one situation or another. They are not presented here to illustrate the most typical kinds of assessment procedures, but rather to illustrate the diversity of measuring tools that have
Evidence suggests that some people with eating disorders may actually have a self-perception disorder; that is, they see themselves as heavier than they really are (Thompson & Smolak, 2001). J. Kevin Thompson and his associates devised the adjustable light beam apparatus to measure body image distortion. Assessees adjust four beams of light to indicate what they believe is the width of their cheeks, waist, hips, and thighs. A measure of accuracy of these estimates is then obtained.

Herman Witkin and his associates (Witkin & Goodenough, 1977) studied personality-related variables in some very innovative ways. For example, they identified field (or context) dependent and field independent people by means of this specially constructed tilting room/tilting chair device. Assessees were asked questions designed to evaluate their dependence or independence of visual cues.

At least since the beginning of the nineteenth century, military units throughout the world have relied on psychological and other tests for personnel selection, program validation, and related reasons (Hartmann et al., 2003). In some cultures, where military service is highly valued, students take preparatory courses with hopes of being accepted into elite military units. This is the case in Israel, where rigorous training such as that pictured here prepares high school students for physical and related tests that only 1 in 60 military recruits will pass.
Pictures such as these sample items from the Meier Art Judgment Test might be used to evaluate people’s aesthetic perception. Which of these two renderings do you find more aesthetically pleasing? The difference between the two pictures has to do with the positioning of the objects on the shelf.

Impairment of certain sensory functions can indicate neurological deficit. For purposes of diagnosis, as well as in measuring progress in remediation, the neurodevelopment training ball can be useful in evaluating one’s sense of balance.
been created for varied uses. In short, if a need exists to measure a particular variable, a way to measure that variable will be devised.

Having considered some of the who, what, and why of assessment, it remains for us to raise the question of where to go for more information. Actually, we think you will find this book to be most useful when future questions arise. But in addition to a book such as this one, where does one go for up-to-date information about tests, testing, and assessment?

Where to Go for Authoritative Information: Reference Sources

Many reference sources exist for learning more about published tests and assessment-related issues. These sources vary with respect to detail. Some merely provide descriptions of tests, whereas others provide very detailed information regarding technical aspects.

Test catalogues Perhaps one of the most readily accessible sources of information about a test is a catalogue distributed by the publisher of the test. Because most test publishers make available catalogues of their offerings, this source of test information can be tapped by a simple telephone call, e-mail, or note. As you might expect, however, publishers' catalogues usually contain only a brief description of the test and seldom contain the kind of detailed technical information that a prospective user of the test might require. Further, the objective of the catalogue is to sell the test. For this reason, highly critical reviews of a test are seldom, if ever, found in a publisher's test catalogue.

Test manuals Detailed information concerning the development of a particular test and technical information relating to it should be found in the manual for the test itself. Test manuals are usually available from test publishers. However, for security purposes, the test publisher will typically require documentation of professional training before filling an order for a test manual. Beyond purchasing a manual from the publisher, the chances are good that somewhere within your university (be it the library or the counseling center), a collection of popular test manuals is maintained. If the test manual you are interested in looking at is not available there, ask your instructor about how best to obtain a reference copy.

Reference volumes The Buros Institute of Mental Measurements provides “one-stop shopping” for a great deal of test-related information. The initial version of what would evolve into the Mental Measurements Yearbook (MMY) was compiled by Oscar Buros (Figure 1-7) in 1933. At this writing, the latest edition of this authoritative compilation of test reviews is the 15th Mental Measurements Yearbook (Plake et al., 2003), although the 16th cannot be far behind. The Buros Institute also publishes Tests in Print (Murphy et al., 2002) as well as a number of other test-related reference works. For a list of its latest offerings, as well as links to a number of other useful test-related test databases, visit the Institute's Web site at http://www.unl.edu/buros/index/simm.html.

Journal articles Articles in current journals may contain reviews of the test, updated or independent studies of its psychometric soundness, or examples of how the instrument

3. We sincerely hope this thought occurs to you when you have completed your coursework and find yourself in line at the bookstore to sell used textbooks.
was used in either research or an applied context. Such articles may appear in a wide array of behavioral science journals such as Psychological Bulletin, Psychological Review, Professional Psychology: Research and Practice, Journal of Personality and Social Psychology, Psychology & Marketing, Psychology in the Schools, School Psychology Quarterly, and School Psychology Review. There are also journals that focus more specifically on matters related to testing and assessment. For example, take a look at journals such as the Journal of Psychoeducational Assessment, Psychological Assessment, Educational and Psychological Measurement, Applied Measurement in Education, and the Journal of Personality Assessment. Journals such as Psychology, Public Policy, and Law and Law and Human Behavior frequently contain highly informative articles on legal and ethical issues and controversies as they relate to psychological testing and assessment.

In addition to articles relevant to specific tests, journals are a rich source of information on important trends in testing and assessment. For example, with reference to clinical psychological assessment, the negative impact of managed health care and the reluctance or refusal of insurers to pay for assessment services have spurred a great deal of self-evaluation on the part of those in the business of evaluation (Camara et al., 2000; Sanchez & Turner, 2003). While critics of clinical assessment argue that testing and assessment is too expensive, too time consuming, and of too little value (Griffith, 1997), more informed reviews of the issues find abundant empirical support for the value of the enterprise (Kubiszyn et al., 2000).

Online databases One of the most widely used bibliographic databases for test-related publications is that maintained by the Educational Resources Information Center (ERIC). Funded by the United States Department of Education and operated out of the University of Maryland, the ERIC Web site at www.eric.ed.gov/searchdb/index.html contains a wealth of resources and news about tests, testing, and assessment. There are abstracts of articles, original articles, and links to other useful Web sites. ERIC strives to provide balanced information concerning educational assessment and to provide resources to encourage responsible test use.

The American Psychological Association (APA) maintains a number of databases useful in locating psychology-related information in journal articles, book chapters, and doctoral dissertations. PsycINFO is a database of abstracts dating back to 1887.
ClinPSYC is a database derived from PsycINFO that focuses on abstracts of a clinical nature. PsycSCAN: Psychopharmacology contains abstracts of articles having to do with psychopharmacology. PsycARTICLES is a database of full-length articles dating back to 1988. Health and Psychosocial Instruments (HAPI) contains a listing of measures created or modified for specific research studies but not commercially available. It is available at many college libraries through BRS Information Technologies, and also available on CD-ROM (updated twice a year). PsycLAW is a free database available to everyone that contains discussions of selected topics having to do with psychology and law. It can be accessed at http://www.apa.org/psyclaw. For more information on any of these databases, visit APA’s Web site at http://www.apa.org.

Educational Testing Service (ETS), “the world’s largest and most influential testing organization” (Frantz & Nordheimer, 1997), maintains its own Web site at http://www.ets.org. The site contains a wealth of information about college and graduate school admission and placement tests, as well as many related resources. If you wanted to try your hand at some practice questions for a test such as the Graduate Record Examination (GRE), for example, this is the place to go. For more information, ETS can be contacted by e-mail at etsinfo@ets.org. A list of Web sites for publishers of other educational and psychological tests is presented in Table 1-2. A number of other Web sites may be of interest to students of assessment, and we have listed a sample of them in Table 1-3.

**Other sources** Your school library contains a number of other sources that may be used to acquire information about tests and test-related topics. For example, two sources for exploring the world of unpublished tests and measures are the Directory of Unpublished Experimental Measures (Goldman & Mitchell, 1997) and Tests in Microfiche, available from Test Collections. APA makes available Finding Information About Psychological Tests (1995), its own guide to locating test-related information. And now for one source of information about these various sources of information . . . see Table 1-4.

Armed with a wealth of background information about tests and other tools of assessment, we’ll explore historical, cultural, and legal/ethical aspects of the assessment enterprise in the following chapter.
### Table 1-3
**Web Sites Having To Do With Testing and Assessment**

<table>
<thead>
<tr>
<th>Web Site Address</th>
<th>Reason to Visit</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://edres.org/scripts/cat">http://edres.org/scripts/cat</a></td>
<td>To obtain some firsthand experience with computerized assessment, learn its logic, and take a “behind-the-scenes” look at it.</td>
</tr>
<tr>
<td><a href="http://www.apa.org/journals/pas.html">www.apa.org/journals/pas.html</a></td>
<td>This is the home page of the APA journal Psychological Assessment. There you will find a table of contents for the current issue and even be able to access the articles.</td>
</tr>
<tr>
<td><a href="http://www.gre.org">www.gre.org</a></td>
<td>The Graduate Record Exam (GRE) is a test in the future of many readers of this book. This is the official site for GRE authoritative information about this test.</td>
</tr>
<tr>
<td><a href="http://edres.org/irt">http://edres.org/irt</a></td>
<td>After you have read Chapter 7 of this book, you may wish to read this advanced, in-depth material on an approach to measurement called item response theory.</td>
</tr>
<tr>
<td><a href="http://edres.org/mdt">http://edres.org/mdt</a></td>
<td>After you have read Chapter 7 of this book, you may wish to read this advanced, in-depth material on an approach to classifying examinees on the basis of statistical decision theory.</td>
</tr>
</tbody>
</table>

### Table 1-4
**Sources of Information About Tests: Some Pros and Cons**

<table>
<thead>
<tr>
<th>Information Source</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test catalogues available from the publisher of the test as well as affiliated distributors of the test</td>
<td>Contains general description of test, including what it is designed to do and who it is designed to be used with. Readily available to most anyone who requests a catalogue.</td>
<td>Primarily designed to sell the test to test users and seldom contains any critical reviews. Information not detailed enough for basing a decision to use the test.</td>
</tr>
<tr>
<td>Test manual</td>
<td>Usually the most detailed source available for information regarding the standardization sample and test administration instructions. May also contain useful information regarding the theory the test is based on, if that is the case. Typically contains at least some information regarding psychometric soundness of the test.</td>
<td>Details regarding the test's psychometric soundness are usually self-serving and written on the basis of studies conducted by the test author and/or test publisher. Test manual itself may be difficult to obtain by students, as its distribution may be restricted to qualified professionals.</td>
</tr>
<tr>
<td>Reference volumes such as the Mental Measurements Yearbook available in bound book form or online</td>
<td>Much like a Consumer Reports for tests, contains descriptions and critical reviews of test written by third parties who presumably have nothing to gain or lose by praising or criticizing the instrument, its standardization sample, and its psychometric soundness.</td>
<td>Few disadvantages if reviewer is genuinely trying to be objective and is knowledgeable, but as with any review, can provide a misleading picture if this is not the case. Also, for very detailed accounts of the standardization sample and related matters, it is best to consult the test manual itself.</td>
</tr>
<tr>
<td>Journal articles</td>
<td>Up-to-date source of reviews and studies of psychometric soundness. Can provide practical examples of how an instrument is used in research or applied contexts.</td>
<td>As with reference volumes, reviews are valuable to the extent they are informed and, to the extent that is possible, unbiased. Reader should research as many articles as possible when attempting to learn how the instrument is actually used; any one article alone may provide an atypical picture.</td>
</tr>
<tr>
<td>Online databases</td>
<td>Widely known and respected online databases such as the ERIC database are virtual “gold mines” of useful information containing varying amounts of detail. Although some legitimate psychological tests may be available for self-administration and scoring online, the vast majority are not.</td>
<td>Consumer beware! Some sites masquerading as databases for psychological tests are designed more to entertain or to sell something than to inform. These sites frequently offer tests you can take online. As you learn more about tests, you will probably become more critical of the value of these self-administered and self-scored “psychological tests.”</td>
</tr>
</tbody>
</table>
Self-Assessment

Test your understanding of elements of this chapter by seeing if you can explain each of the following terms, expressions, and abbreviations:

- American Board of Assessment Psychology (ABAP)
- American Board of Professional Psychology (ABPP)
- achievement test
- alternate assessment
- American Psychological Association (APA)
- assessment center approach
- behavioral observation
- case history data
- central processing
- collaborative psychological assessment
- computer-assisted psychological assessment (CAPA)
- consultative report
- cut score
- descriptive report
- diagnosis
- diagnostic test
- dynamic psychological assessment
- extended scoring report
- format
- groupthink
- health psychology
- informal evaluation
- integrative report
- interpretive report
- interview
- local processing
- measurement
- naturalistic observation
- panel interview
- portfolio
- protocol
- psychological assessment
- psychological autopsy
- psychological test
- psychological testing
- psychometrician
- psychometrics
- psychometrist
- psychometry
- PsychINFO
- Public Law 94-142
- Public Law 99-457
- rapport
- role play
- role-play test
- school ability test
- score
- scoring
- scoring report
- screening report
- simple scoring report
- teleprocessing
- test
- test catalogue
- test developer
- test manual
- testtaker
- test user
- therapeutic psychological assessment

Web Watch

Check out the following Web sites for more information about topics in this chapter:

- Standards for Educational and Psychological Testing
  www.apa.org/science/standards.html
- National Council on Measurement in Education
  www.ncme.org
- American Educational Research Association (AERA)
  www.aera.net
- Illinois State Board of Education-Alternative Assessment
  www.isbe.net/assessment/IAA.htm
  www.isbe.net/assessment/default.htm
- IDEA
  www.ed.gov/offices/OSERS/Policy/IDEA/index.html
  www.idea.practices.org/law/index.php
- American Board of Professional Psychology (ABPP)
  www.abpp.org
- American Board of Assessment Psychology (ABAP)
  www.assessmentpsychologyboard.org
- American Academy of School Psychology
  http://espse.ed.psu.edu/spsy/aasp/aasp.ssi
- American Academy of School Psychologists (NASP)
  www.nasponline.org/index2.html
- PsycLAW
  www.apa.org/psyclaw