

Psychological Testing and Assessment

All fields of human endeavor use measurement in some form, and each field has its own set of measuring tools and measuring units. For example, 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*. As a student of psychological measurement, you need a working familiarity with some of the commonly used units of measure in psychology and a knowledge of some of the many measuring tools employed. In the pages that follow, you will gain that knowledge as well as an acquaintance with the history of measurement in psychology and an understanding of its theoretical basis.

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 and 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. There were tests to measure not only intelligence but also personality, aspects of brain functioning, performance at work, and many other aspects of psychological and social functioning.

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 enterprise. *Testing* was the term used to refer to everything from the administration of a test (as in “Testing in progress”) to the interpretation of a test score (“The testing indicated that . . .”). During World War I, the process of testing aptly described the group screening of thousands of military recruits. We suspect that it was then that *testing* gained a powerful foothold in the vocabulary of professionals and laypeople. The use of *testing* to denote everything from test administration to test interpretation can be found in postwar textbooks (such as Chapman, 1921; Hull, 1922; Spearman, 1927) as well as in various 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 U.S. 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 Assessment Staff, 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.

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 semantic distinction between *psychological testing* and *psychological assessment* is blurred in everyday conversation. Somewhat surprisingly, the distinction between the two terms remains blurred even in edition after edition of some published “psychological testing” textbooks. Yet the distinction is important. Society at large is best served by a clear definition of and differentiation between these two terms as well as related terms such as *psychological test user*

and *psychological assessor*. Clear distinctions between such terms may also play a role in avoiding the turf wars now brewing between psychology professionals and members of other professions seeking to use various psychological tests. In many psychological evaluation contexts, it requires greater education, training, and skill to conduct an assessment than to simply administer a test.

We define **psychological assessment** as the gathering and integration of psychology-related data for the purpose of making a psychological evaluation that is 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. Some of the differences between these two processes are further discussed in Table 1–1.

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

Table 1–1
Testing in Contrast to Assessment

In contrast to the process of administering, scoring, and interpreting psychological tests (psychological testing), psychological assessment may be conceived as a problem-solving process that can take many different forms. How psychological 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.

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. It seems useful to distinguish the differences between testing and assessment in terms of the objective, process, and outcome of an evaluation and also in terms of the role and skill of the evaluator. Keep in mind that, although these are useful distinctions to consider, exceptions can always be found.

Testing	Assessment
<i>Objective</i>	
Typically, to obtain some gauge, usually numerical in nature, with regard to an ability or attribute.	Typically, to answer a referral question, solve a problem, or arrive at a decision through the use of tools of evaluation.
<i>Process</i>	
Testing may be individual or group in nature. After test administration, the 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 typically individualized. In contrast to testing, assessment more typically focuses on <i>how</i> an individual processes rather than simply the results of that processing.
<i>Role of Evaluator</i>	
The tester is not key to the process; practically speaking, one tester may be substituted for another tester without appreciably affecting the evaluation.	The assessor is key to the process of selecting tests and/or other tools of evaluation as well as in drawing conclusions from the entire evaluation.
<i>Skill of Evaluator</i>	
Testing typically requires technician-like skills in terms of administering and scoring a test as well as in interpreting a test result.	Assessment typically requires an educated selection of tools of evaluation, skill in evaluation, and thoughtful organization and integration of data.
<i>Outcome</i>	
Typically, testing yields a test score or series of test scores.	Typically, assessment entails a logical problem-solving approach that brings to bear many sources of data designed to shed light on a referral question.

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 in order to clarify aspects of the reason for referral. The assessor prepares for the assessment by selecting the tools of assessment to be used. For example, if the assessment is in a corporate or military setting and the referral question concerns the assessee’s leadership ability, the assessor may wish to employ a measure (or two) of leadership. Typically, it the assessor’s own past experience, education, and training that play a key

role in the specific tests or other tools to be employed in the assessment. Sometimes an institution in which the assessment is taking place has prescribed guidelines for which instruments can and cannot be used. In most every assessment situation, particularly situations that are relatively novel to the assessor, the tool selection process may be informed by some research in preparation for the assessment. For example, in the assessment of leadership, the tool selection procedure might be informed by publications dealing with general approaches to leadership measurement (Foti & Hauenstein, 2007), psychological studies of leaders (Kouzes & Posner, 2007), or cultural issues in leadership (Byrne & Bradley, 2007).

Subsequent to the selection of the instruments or procedures to be employed, the formal assessment will begin. After the assessment, the assessor writes a report of the findings that is designed to answer the referral question. More 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. Other assessors view the process of assessment as more of a collaboration between the assessor and the assessee. For example, in one approach to assessment, referred to (logically enough) as **collaborative psychological assessment**, the assessor and assessee may work as "partners" from initial contact through final feedback (Fischer, 1978, 2004, 2006). 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 a collaborative approach to assessment called **therapeutic psychological assessment**. Here, therapeutic self-discovery and new understandings are encouraged throughout the assessment process.

Another approach to assessment that seems to have picked up momentum in recent years, most notably in educational settings, is referred to as *dynamic assessment*. While the term *dynamic* may at first glance suggest to some a psychodynamic or psychoanalytic approach to assessment, as used in this context it refers to the interactive, changing, or varying nature of the assessment. In general, **dynamic assessment** refers to an interactive approach to psychological assessment that usually follows a model of (1) evaluation,

(2) intervention of some sort, and (3) evaluation. Dynamic assessment is most typically employed in educational settings, although it may be employed in correctional, corporate, neuropsychological, clinical, and most any other setting as well.

Intervention between evaluations, sometimes even between individual questions posed or tasks given, might take many different forms, depending upon the purpose of the dynamic assessment (Haywood & Lidz, 2007). For example, an assessor may intervene in the course of an evaluation of an assessee's abilities with increasingly more explicit feedback or hints. The purpose of the intervention may be to provide assistance with mastering the task at hand. Progress in mastering the same or similar tasks is then measured. In essence, dynamic assessment provides a means for evaluating how the assessee processes or benefits from some type of intervention (feedback, hints, instruction, therapy, etc.) during the course of evaluation. In some educational contexts, dynamic assessment may be viewed as a way of measuring not just learning but so-called learning potential, or "learning how to learn" skills. We'll revisit the topic of dynamic assessment in greater detail in Chapter 11, which deals with assessment in educational settings. For now, let's move on and define some other basic terminology in the world of testing and assessment.

JUST THINK . . .

Besides tests, what other tools of psychological assessment come to mind? (No peeking!)

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). Whereas a medical test might involve analysis of a sample of blood, tissue, or the like, a psychological test almost always involves 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 with respect to 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. This is so because what is deemed important in measuring “personality” for one test developer might be entirely different for another test developer; different test developers employ different definitions of “personality.” Additionally, different test developers come to the test development process with different theoretical orientations. For example, items on a psychoanalytically oriented personality test may have little resemblance to those on a behaviorally oriented personality test, yet both are personality tests. A psychoanalytically oriented personality test might be chosen for use by a psychoanalytically oriented assessor, and an existentially oriented personality test might be chosen for use by an existentially oriented assessor.

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: PC- or Apple/Mac-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 an active and knowledgeable test administrator. The test administration may involve demonstration of various kinds of tasks on the part of the assessee as well as trained observation of an assessee’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 more than a dozen different methods that can be used to formally derive cut scores (Dwyer, 1996). In Chapter 7, we present a sampling of the ways that a cut score may be derived.

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. There is also 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 the testtakers themselves, and others are designed to be scored by trained examiners. Still other tests may be scored and fully interpreted within seconds by computer. Some tests, such as most tests of intelligence, come with test manuals that are 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. Variants of these words include the adjective *psychometric* (which refers to measurement that is psychological in nature) and the nouns **psychometrist** and **psychometrician** (both referring to psychological test users). One speaks of the *psychometric soundness* of a test when referring to how consistently and how accurately a psychological test measures what it purports to measure. Assessment professionals also speak of the psychometric *utility* of a particular test or assessment method. In this context, **utility** refers to the usefulness or practical value that a test or assessment technique has for a particular purpose. We elaborate on the subject of utility in Chapter 7.

Throughout this book and consistent with common practice, we sometimes use the word *test* (as well as related terms such as *test score*) in the broadest and most generic sense when discussing general principles applicable to various measurement



Figure 1-1
Emotion Engendered by Categorical Cutoffs

According to research by Victoria Husted Medvec and her colleagues (Medvec et al., 1995; Medvec & Savitsky, 1997), people who just make some categorical cutoff may feel better about their accomplishment than those who make the cutoff by a substantial margin. But those who just miss the cutoff may feel worse than those who miss it by a substantial margin. Evidence consistent with this view was presented in research with Olympic athletes. Bronze medalists were—somewhat paradoxically—happier with the outcome than silver medalists. Bronze medalists might say to themselves “At least I won a medal” and be happy about it. By contrast, silver medalists might feel frustrated about having gone for the gold and missed winning it.

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 other, more specific terms (such as *situational performance measures*). In other words, the term *test* may be used in shorthand fashion throughout this book to apply to the widest possible array of measurement procedures. With that disclaimer duly noted, we now return to our discussion of tools of assessment. Next up, please meet one tool of measurement that, as they say, “needs no introduction.”

The Interview

In everyday conversation, the word *interview* conjures images of face-to-face talk. But the interview as a tool of psychological assessment typically involves more than talk. If the interview is conducted face-to-face, then the interviewer is probably taking note of not only the content of what is said but also the way it is being said. More specifically,

the interviewer is taking note of both verbal and nonverbal behavior. Nonverbal behavior may include the interviewee's "body language," movements and facial expressions in response to the interviewer, the extent of eye contact, and apparent willingness to cooperate. The interviewer may also take note of the way that the interviewee is dressed. Here, variables such as neat versus sloppy and appropriate versus inappropriate may be noted.

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?

Because of a potential wealth of nonverbal information to be gained, interviews are ideally conducted face-to-face. However, face-to-face contact is not always possible and interviews may be conducted in other formats, such as by telephone. In an interview conducted by telephone, the interviewer may still be able to gain information beyond the responses to questions by being sensitive to

variables such as changes in the interviewee's voice pitch or the extent to which particular questions precipitate long pauses or signs of emotion in response. Of course, interviews need not involve verbalized speech, as when they are conducted in sign language. Interviews may be conducted by various electronic means, as would be the case with online interviews, e-mail interviews, and interviews conducted by means of text messaging. 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, length, and nature. Interviews may be used by psychologists in various specialty areas to help make diagnostic, treatment, selection, or other decisions. So, for example, school psychologists may use an interview to help make a decision about the appropriateness of various educational interventions or class placements. A court-appointed psychologist may use an interview to help guide the court in determining whether a defendant was insane at the time of a commission of a crime. A specialist in head injury may use an interview to help shed light on questions related to the extent of damage to the brain that was caused by the injury. A psychologist studying consumer behavior may use an interview to learn about the market for various products and services as well as how best to advertise and promote them.

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, especially in the field of human resources, a specialized interview called a **panel interview** may be employed. Here, more than one interviewer participates in the personnel assessment. A presumed advantage of this approach, which has also been

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 relates to its utility; the cost of using multiple interviewers may not be justified, especially when the return on this investment is questionable (Dixon et al., 2002).

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?

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 inter-

view on television, radio, or the Internet! Regardless of the medium through which it is conducted, an interview is a reciprocal affair in that the interviewee reacts to the interviewer and the interviewer reacts to the interviewee. The quality, if not the quantity, of useful information produced by an interview depends in no small part on the skills of



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?

the interviewer. Interviewers differ in many ways: 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 two celebrities might affect responses of interviewees. Would you characterize them as good or bad interviewers? Why?

The Portfolio

Students and professionals in many different fields of endeavor ranging from art to architecture keep files of their work products. These work products—whether retained on paper, canvas, film, video, audio, or some other medium—constitute what is called a **portfolio**. As samples of one’s ability and accomplishment, a portfolio may be used as a tool of evaluation. Employers of commercial artists, for example, will make hiring decisions based, in part, on the impressiveness of an applicant’s portfolio of sample drawings. As another example, consider the employers of on-air radio talent. They, too, will make hiring decisions that are based partly upon their judgments of audio samples of the candidate’s previous work.

The appeal of portfolio assessment as a tool of evaluation extends to many other fields, including education. Some have argued, for example, that the best evaluation of a student’s writing skills can be accomplished not by the administration of a test but by

JUST THINK . . .

How might portfolio assessment be used as a tool of evaluation for one aspiring to hold public office?

JUST THINK . . .

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

asking the student to compile a selection of writing samples. Also in the field of education, portfolio assessment has been employed as a tool in the hiring of instructors. An instructor's portfolio may consist of various documents such as lesson plans, published writings, and visual aids developed expressly for teaching certain subjects. All of these materials can be extremely useful to those who must make hiring decisions.

Case History Data

Case history data refers to records, transcripts, and other accounts in written, pictorial, or other form that preserve archival information, official and informal accounts, and other data and items relevant to an assessee. Case history data may include files or excerpts from files maintained at institutions and agencies such as schools, hospitals, employers, religious institutions, and criminal justice agencies. Other examples of case history data are letters and written correspondence, 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 is a 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 on the events and circumstances that may have contributed to any changes in adjustment. Case history data can be of critical value in neuropsychological evaluations, where it often provides information about neuropsychological functioning prior to the occurrence of a trauma or other event that results in a deficit. School psychologists rely on case history data for insight into a student's current academic or behavioral standing. Case history data is also useful in making judgments concerning future class placements.

JUST THINK . . .

What are the strengths of the case study as a tool of assessment? What are the weaknesses of the case study as a tool of assessment?

Case history data can be of critical value in neuropsychological evaluations, where it often provides information about neuropsychological functioning prior to the occurrence of a trauma or other event that results in a deficit. School psychologists rely on case history data for insight into a student's current academic or behavioral

standing. Case history data is also useful in making judgments concerning future class placements. Another use of the term *case history*, one synonymous with *case study*, concerns the assembly of case history data into an illustrative account. For example, a case study might shed light on how one individual's personality and a particular set of environmental conditions combined to produce a successful world leader. A case study of an individual who attempted to assassinate a high-ranking political figure could 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 situation, observe his or her behavior in that situation. Such "down-home" wisdom underlies at least one approach to evaluation. **Behavioral observation**, as it is employed by assessment 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 is often used as a diagnostic aid in various settings such as inpatient facilities, behavioral research laboratories, and classrooms. In addition to diagnosis, behavioral observation may be used for selection purposes, as in corporate

settings. Here behavioral observation may be used as a tool to help identify people who demonstrate the abilities required to perform a particular task or job. Sometimes researchers venture outside of the confines of clinics, classrooms, workplaces, and research laboratories in order to observe behavior of humans in a natural setting—that is, the setting in which the behavior would typically be expected to occur. This variety of behavioral observation is referred to as **naturalistic observation**. As an example, one team of researchers studying the socializing behavior of autistic children with same-aged peers opted for natural settings rather than a controlled, laboratory environment (Bellini et al., 2007).

Behavioral observation as an aid to designing therapeutic intervention has proven to be 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 individuals and design interventions accordingly. In a school situation, for example, naturalistic observation on the playground of a culturally different child suspected of having 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.

In practice, behavioral observation tends to be used infrequently outside of research facilities, prisons, inpatient clinics, and other types of facilities where the observers have ready access to assessees. This is so more for economic reasons than anything else. For private practitioners, it is typically not economically feasible to spend hours out of the consulting room observing clients. Still, there are some mental health professionals, such as those in the field of assisted living, who find great value in behavioral observation of patients outside of their institutional environment. For them, it may be necessary to accompany a patient outside of the institution’s walls to learn if that patient is capable of independently performing activities of daily living. In this context, a tool that relies heavily on behavioral observation, such as the Test of Grocery Shopping Skills (see Figure 1–3), may be extremely useful.

Role-Play Tests

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.)

Role play is useful in evaluating various skills. So, for example, grocery shopping skills (Figure 1–3) could conceivably be evaluated through role play, and a trip to the supermarket could be saved. Of course, role play may not be as useful as the real thing in all situations. Still, role play is used quite extensively, especially in situations where it is too time-consuming, too expensive, or simply too inconvenient to assess in a “real” situation. Astronauts in training may be required to role play many situations “as if” in outer space. The “as if” scenario for training purposes in this case will result in cost savings of many millions of dollars; the cost of actually putting such trainees in the real situation would be . . . well, astronomical.

JUST THINK . . .

What are the strengths and weaknesses of behavioral observation, including naturalistic observation, as a tool of assessment?

JUST THINK . . .

What are the strengths of role play as a tool of assessment? What are the weaknesses of role play as a tool of assessment?



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.

Individuals being evaluated in a corporate, industrial, organizational, or military context for managerial or leadership ability are routinely placed in role-play situations. They may be asked, for example, 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. The same test is then administered again subsequent to completion of treatment.

Computers as Tools

We have already made reference to the role computers play in contemporary assessment in the context of generating simulations. They may also help in the measurement of variables that in the past were quite difficult to quantify (see Figure 1-4). But perhaps the more obvious role as a tool of assessment is their role in test administration, scoring, and interpretation.

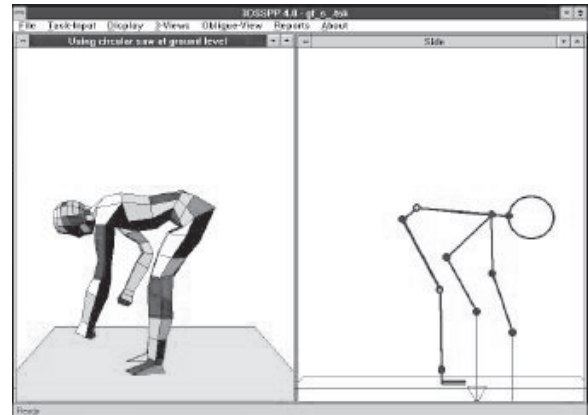


Figure 1-4
A Method of Quantifying Back Stress

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.

As test administrators, computers do much more than replace the “equipment” that was so widely used in the past (a No. 2 pencil). Computers can serve as test administrators (online or off) and as highly efficient test scorers. Within seconds they can derive not only test scores but patterns of test scores. Scoring may be done on-site (**local processing**) or conducted at some central location (**central processing**). If processing occurs at a central location, test-related data may be sent to and returned from this central facility by means of phone lines (**teleprocessing**), by mail, or courier. Whether processed locally or centrally, the account of performance spewed out can range from a mere listing of score or scores (i.e., a **simple scoring report**) to the more detailed **extended scoring report**, which includes statistical analyses of the testtaker’s performance. A step up from a scoring report is the **interpretive report**, which is distinguished by its inclusion of numerical or narrative interpretive statements in the report. Some interpretive reports contain relatively little interpretation and are limited to calling the test user’s attention to certain scores that need to be focused on. At the high end of interpretive reports is what is sometimes referred to as a **consultative report**. This type of report, usually written in language appropriate for communication between assessment professionals, may provide expert opinion concerning analysis of the data. Yet another type of computerized scoring report is designed to integrate data from sources other than the test itself into the interpretive report. Such an **integrative report** will employ previously collected data (such as medication records or behavioral observation data) into the test report.

The acronym **CAPA** refers to the term *computer assisted psychological assessment*. By the way, here the word *assisted* typically refers to the assistance computers provide to the test user, not the testtaker. Another acronym you may come across is **CAT**, this for *computer adaptive testing*. The *adaptive* in this term is a reference to the computer's ability to tailor the test to the testtaker's ability or testtaking pattern. So, for example, on a computerized test of academic abilities, the computer might be programmed to switch from testing math skills to English skills after three consecutive failures on math items.

JUST THINK . . .

What are the pros and cons of the various types of CAPA processing?

CAPA opened a world of possibilities for test developers, enabling them to develop psychometrically sound tests using mathematical procedures and calculations so complicated that they may have taken weeks or months to

use in a bygone era. It opened a new world to test users, enabling the construction of tailor-made tests with built-in scoring and interpretive capabilities previously unheard of. For many test users, CAPA represents a great advance over the past, when they had to personally administer tests and possibly even place the responses in some other form prior to analysis (manually using a scoring template or other device) before beginning the often laborious tasks of scoring and interpreting the resulting data. Still, every rose has its thorns; some of the pros and cons of CAPA are presented in Table 1–2.

Other Tools

The next time you have occasion to play 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 video's potential applications to assessment is endless.

In addition to video, many other commonplace items that you may not readily associate with psychological assessment may be pressed into service for just that purpose. For example, psychologists may use many of the tools traditionally associated with medical health, such as thermometers to measure body temperature and

JUST THINK . . .

In general, when is assessment using video a good idea? What are the drawbacks, if any, to using video in assessment?

gauges to measure blood pressure. Biofeedback equipment is sometimes used to obtain measures of bodily reactions (such as muscular tension) to various sorts of stimuli. And then there are some less common instruments, such as the penile plethysmograph. This instrument, designed to measure male sexual arousal, may be 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 fundamental questions regarding the who, what, why, how, and where of testing and assessment.

Table 1–2
CAPA: Some Pros and Cons

Pros	Cons
CAPA saves professional time in test administration, scoring, and interpretation.	Professionals must still spend significant time reading software and hardware documentation and even ancillary books on the test and its interpretation.
CAPA results in minimal scoring errors resulting from human error or lapses of attention or judgment.	With CAPA, the possibility of software or hardware error is ever present, from difficult-to-pinpoint sources such as software glitches or hardware malfunction.
CAPA assures standardized test administration to all testtakers with little, if any, variation in test administration procedures.	CAPA 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.).
CAPA yields standardized interpretation of findings due to elimination of unreliability traceable to differing points of view in professional judgment.	CAPA's standardized interpretation of findings based on a set, unitary perspective may not be optimal; interpretation could profit from alternative viewpoints.
Computers' capacity to combine data according to rules is more accurate than that of humans.	Computers lack the flexibility of humans to recognize the exception to a rule in the context of the "big picture."
Nonprofessional assistants can be used in the test administration process, and the test can typically be administered to groups of testtakers in one sitting.	Use of nonprofessionals leaves diminished, if any, opportunity for the professional to observe the assessee's testtaking behavior and note any unusual extra-test conditions that may have affected responses.
Professional groups such as APA develop guidelines and standards for use of CAPA products.	Profit-driven nonprofessionals may also create and distribute tests with little regard for professional guidelines and standards.
Paper-and-pencil tests may be converted to CAPA products with consequential advantages, such as a shorter time between the administration of the test and its scoring and interpretation.	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.
Security of CAPA products can be maintained not only by traditional means (such as locked filing cabinets) but by high-tech electronic products (such as firewalls).	Security of CAPA 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.
Computers can automatically tailor test content and length based on responses of testtakers.	Not all testtakers take the same test or have the same test-taking experience.

Who, What, Why, How, and Where?

Who are the parties in the assessment enterprise? In *what* types of settings are assessments conducted? *Why* is assessment conducted? *How* are assessments 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. Additionally, we may consider society at large as a party to the assessment enterprise.

The test developer Test developers and publishers create tests or other methods of assessment. The American Psychological Association (APA) has estimated that more than 20,000 new psychological tests are developed each year. 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.¹

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 Psychological tests and assessment methodologies are used by a wide range of professionals, including clinicians, counselors, school psychologists, human resources personnel, consumer psychologists, experimental psychologists, social psychologists, . . . ; the list goes on. To provide readers with a sense of what it means to be a user of psychological tests, we introduce a new feature to this edition of our textbook: *Meet an Assessment Professional*. Here, psychological test users with varying perspectives on the assessment enterprise will share some of their thoughts and experiences in their own words. We'll present an excerpt of one assessment professional's writings in each chapter, and the complete version of the essay will be available online.

The *Standards*, as well other published guidelines from specialty professional organizations, have had much to say in terms of identifying just who is a qualified test user and who should have access to (and be permitted to purchase) psychological tests and related tools of psychological assessment. Still, controversy exists about which professionals with what type of training should have access to which tests. Members of various professions, with little or no psychological training, have sought the right to obtain and use psychological tests. In many countries, no ethical or legal regulation of psychological test use exists (Leach & Oakland, 2007).

Another *Who?* question of emerging importance involves the presence of third parties in a room in which assessment is being conducted. A discussion of issues related to the presence of third parties during psychological assessment took place before a standing-room-only audience at a recent annual meeting of the American Psychological Association. Attendees were treated to scholarly presentations (Krauss, 2007; McCaffrey, 2007; Otto, 2007) and a lively debate concerning these issues. Although the focus of this discussion, as well as much of the published literature, was on neuropsychological and forensic assessments, the issues raised also apply to other psychological assessment situations—particularly those involving the assessee's intelligence or related cognitive abilities (see this chapter's *Close-up*).

So who are (or should be) test users? Should occupational therapists, for example, be allowed to administer psychological tests? What about employers and human resource executives with no formal training in psychology? Should supervisors of psychological assessors be allowed in the room while an assessment is taking place? These are con-

1. For an intriguing glimpse at biographical information on a sampling of test developers, navigate to the "Test Developer Profiles" section of our Web site: www.mhhe.com/cohentesting7.

MEET AN ASSESSMENT PROFESSIONAL

Meet Dr. Barbara C. Pavlo

“What tools of assessment do I use?” Perhaps a better question would be, “What tools of assessment *don’t* I use?” I probably use interviews the most (sometimes structured, more often semi-structured) with written tests such as Beck’s test (for assessment of depression and anxiety) a close second. But depending on the assessment objective, I can also use various other tools such as case history document analysis, behavioral observation, figure drawing analysis, and evaluations that employ role play. Each tool of assessment can have a place in providing important pieces to a puzzle. Great artisans, craft people, and others who have earned respect and admiration for doing what they do have mastered the art of using the tools available to them to best advantage. Why should it be different for psychologists?

How do I use assessment data? I use it in the development, implementation, and fine-tuning of interventions. It helps greatly in terms of decisions concerning where to focus therapeutic efforts. It can be extremely useful in enlightening patients with little insight into their own condition. With some patients, a test can serve as a kind of “icebreaker” that opens floodgates of memories that had previously been stored and shelved neatly away. Most people who seek psychotherapy are eager to learn more about themselves, and valid tests, skillfully administered and interpreted, can put them on the fast track to doing just that. Moreover, when used in a so-called dynamic (test-intervention-retest) manner, tests can provide feedback to clients regarding their progress. And by the way, it’s very nice (and very useful) for therapists to get that feedback too. . . .



Barbara C. Pavlo, Psy.D., Independent Practice, West Hills, California

Many students aspiring to become psychologists go through school with visions of seeing themselves conducting psychotherapy. Relatively few students go through school with visions of seeing themselves administering psychological tests. This is unfortunate given the potentially great role psychological tests can play not only in clinical diagnosis and intervention, but in research. Take it from one who also gave little thought to assessment when she was an aspiring psychologist: This stuff is important, you need to know it, and the better you know it, the better your chances for success in whatever area of psychology you choose to work in.

Read more of what Dr. Pavlo had to say—her complete essay—at www.mhhe.com/cohentesting7.

roversial *Who?* questions that knowledgeable assessment professionals still debate. Fortunately, there is another *Who?* question that stimulates far less controversy: the one regarding who the testtaker or assessee is.

The testtaker Having all taken tests, we all have had firsthand experience in the role of testtaker. However, putting ourselves in the position of test users, it is important to develop an appreciation for the many varied ways that testtakers can approach an

Should Observers Be Parties to the Assessment Process?

The assessor and the assessee are two parties in any assessment. The third party in an assessment may be an observer who is there for any number of reasons. The third-party observer may be a supervisor of the assessor, a friend or relative of the assessee, a representative of the institution in which the assessment is being conducted, a translator, an attorney, or someone else. But do third parties have a legitimate place in the room during an assessment for any reason? According to Robert J. McCaffrey (2007), the answer to this question is clear and unambiguous: “No, third parties should not be allowed to be present during an assessment.”

McCaffrey and others cite research to support the view that a social influence process takes place through the mere presence of a third party (Yantz & McCaffrey, 2005). This social influence may be sufficient to affect the assessee’s performance, particularly on tasks involving memory, attention, and other cognitive functions (Gavett et al., 2005). The effect of third-party observation on an assessee’s performance may even take place in the absence of the physical presence of the third party. So, for example, third-party observation by means of such devices as a video camera or a one-way mirror may also affect an assessee’s performance (Constantinou et al., 2005).

The social influence effect that occurs has been referred to in the testing and assessment literature as **social facilitation**, probably because the presence of third parties was initially associated with increments in performance (Aiello & Douthitt, 2001). However, in the light of additional research suggesting that an audience may also have the effect of inhibiting performance, a more inclusive term—such as *social facilitation and inhibition*—would probably be more accurate (McCaffrey et al., 2005).

Proponents of third-party access to psychological assessment argue that it is necessary for purposes such as clinical training; there is no substitute for having a supervisor right there, in the room, to correct any test administration errors that an assessor-in-training might make during the course of an assessment. Other arguments in favor of third-party access may cite the need for translators or for an attorney to ensure that an assessee’s rights are respected.



Some state statutes specifically provide for the presence of third-party observers under certain conditions, although most states still have not addressed this issue either by legislation or judicial action (Duff & Fisher, 2005). One polling of a small sample ($n = 27$) of forensic experts concluded that a majority ($n = 14$) was in favor of allowing third-party observers; the remaining respondents were either against it or unclassifiable as favoring either position (Witt, 2003).

Advocates of the strict enforcement of a policy that prohibits third-party observers during psychological assessment argue that alternatives to such observation either exist or must be developed. For example, instead of allowing supervisors in the room during an assessment, better training procedures—including greater reliance on role play prior to actual test administrations—should be instituted. McCaffrey (2005) has cautioned that certain assessment data gathered in the presence of a third-party observer may be deemed unreliable in a court of law and thus inadmissible. He further advised that, any time there is a third party to an assessment, that fact should be clearly noted on the assessment report along with the possible consequences of the third party’s presence.

assessment. On the appointed day of a 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
- The extent to which they understand and agree with the rationale for the assessment
- 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
- 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 with people who knew him or her. For example, using psychological autopsies, Townsend (2007) explored the question of whether suicide terrorists were indeed suicidal from a classical psychological perspective; she concluded that they were not. Other researchers have provided fascinating post-mortem psychological evaluations of people from various walks of life in many different cultures (Bhatia et al., 2006; Chan et al., 2007; Dattilio, 2006; Fortune et al., 2007; Giner et al., 2007; Goldstein et al., 2008; Heller et al., 2007; McGirr et al., 2007; Owens et al., 2008; Palacio et al., 2007; Phillips et al., 2007; Pouliot & De Leo, 2006; Sanchez, 2006; Thoresen et al., 2006; Zonda, 2006).

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. Proponents of palmistry, podoscopy, astrology, and phrenology, among other pursuits,

have argued that the best means of understanding and predicting human behavior was through the study of the palms of the hands, 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) described as society's demand for "some way of organizing or systematizing the many-faceted complexity of individual differences."

Society at large exerts its influence as a party to the assessment enterprise in many ways. As society evolves and as the need to measure different psychological variables emerges, test developers respond by devising new tests. Through elected representatives to the legislature, laws are enacted that govern aspects of test development, test administration, and test interpretation. Similarly, by means of court decisions, society at large exerts its influence on various aspects of the testing and assessment enterprise.

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 parties to a greater or lesser extent in the assessment enterprise.

Having introduced you to some of the parties involved in the *Who?* of psychological testing and assessment, let's move on to tackle some of the *What?* and *Why?* questions.

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 have special needs. In addition to **school ability tests**, another type of test commonly given in schools is an **achievement test**, which evaluates accomplishment or the degree of learning that has taken place. Some of the achievement tests you have taken in school were constructed by your teacher. Other achievement tests were constructed for more widespread use by educators working with 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 11).

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. Similarly, the term **diagnostic test** refers to a tool of assessment used to help narrow down and identify areas of deficit to be targeted for intervention. In educational settings, diagnostic tests of reading, mathematics, and other academic subjects may be administered to assess the need for educational intervention as well as to establish or rule out 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 tools of evaluation we have discussed in the context of educational settings (such as achievement tests, diagnostic tests, and informal evaluations) are also administered in various other settings. And some of the types of tests we discuss in the context of the settings described next are also administered in educational settings. So please keep in mind that the tools of evaluation and measurement techniques that we discuss in one context may well be used in other contexts. Our objective at this early stage in our survey of the field is simply to introduce a sampling (not a comprehensive list) of the types of tests used in different settings.

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 14 and elsewhere, we will look at the nature, uses, and benefits of assessment in both clinical and counseling settings.

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, 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?" Having mentioned retirement, let's hasten to introduce you to another type of setting in which psychological tests are used extensively.

Geriatric settings In the United States, more than 12 million adults are currently in the age range of 75 to 84; this 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, which is a 33-fold increase in the number of people of that age since 1900. 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. At issue in many such assessments is the extent to which assessees are enjoying as good a **quality of life** as possible. The definition of *quality of life* has varied as a function of perspective in different studies. In some research, for example, *quality of life* is defined

from the perspective of an observer; in other research, it is defined from the perspective of assessees themselves and refers to one's own self-report regarding lifestyle-related variables. However defined, what is typically assessed in such research includes evaluation with respect to variables such as perceived stress, loneliness, sources of satisfaction, personal values, quality of living conditions, and quality of friendships and other social support.

JUST THINK . . .

Tests are used in geriatric, counseling, and other settings to help improve quality of life. But are there some things a psychological test just can't measure?

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. They help corporations predict the public's receptivity to a new product, a new brand, or a new advertising or marketing campaign. They help "diagnose" the needs of existing and older brands and products and identify how they might be revitalized or made more appealing in the eyes of the consumer.

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 present themselves to the public as attorneys until they pass their state's bar examination. Psychologists, too, must pass an examination before adopting the official title of "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 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, whether it is 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 discipline 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. For example, measurement tools may be used to quantify postpartum depression in women who have recently given birth (Chabrol et al., 2002). These same measurement tools might be used to gauge the degree of improvement that has taken place as a result of counseling, medication, or any other intervention. Measurement tools may be used to compare one group of research subjects to another on some targeted variable. In a study that employed behavioral observation, for example, children with attention deficit hyperactivity disorder (ADHD) were compared to nonreferred children on the variable of television watching (Acevedo-Polakovich et al., 2007).

Another general line of research in health psychology focuses on aspects of personality, behavior, or lifestyle as they relate to good physical health and longevity versus impaired physical health and death. 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. Other investigators explored college athletes’ motivation to use alcohol using a test called the Drinking Motives Measure (DMM). Consistent with prior research, these investigators concluded that athletes involved in intercollegiate sports may be particularly susceptible to using alcohol and other drugs as a coping mechanism in the face of elevated stressors (Martens et al., 2003). The researchers viewed the DMM as effective in predicting alcohol consumption and believed it might therefore have an application in prevention or intervention programs.

Many forms of exploratory research in health psychology, as well as other specialty areas, rely heavily on interviews and/or group discussion for information gathering. For example, relying primarily on group discussion, research in Russia explored the

use of alcohol during pregnancy and its relation to fetal alcohol syndrome (Balachova et al., 2007).

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 meaningful answers are to be forthcoming.

How Are Assessments Conducted?

If a need exists to measure a particular variable, a way to measure that variable will be devised. As Figure 1–5 just begins to illustrate, the ways in which measurements can be taken are limited only by imagination. Keep in mind that this figure illustrates only a small sample of the many methods used in psychological testing and assessment. The photos are not designed to illustrate the most typical kinds of assessment procedures. Rather, their purpose is to call attention to the wide range of measuring tools that have been created for varied uses.

Regardless of the specific test or measurement procedure employed, there will most likely be some common ground in terms of how the assessor prepares for the assessment, how the assessment is administered, how the scores or results of the assessment are used, and how the entire record of the assessment is stored. This is so because of published guidelines for test use promulgated in the *Standards* and related publications. Responsible test users have obligations before, during, and after a test or any measurement procedure is administered. For purposes of illustration, consider the administration of a paper-and-pencil test. 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 According to principles of professional ethics promulgated by the National Association of School Psychologists (NASP), school psychologists have another pretest obligation: selecting and using tests that are most appropriate for the individual student being tested.

Test users have the responsibility of ensuring that the room in which the test will be conducted is suitable and conducive to the testing. 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. Of course, creating an ideal testing environment is not always something every examiner can do (see Figure 1–6).

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 this context, **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

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 or booklet on which the testtaker's responses are entered.

about the nature of the test and why it is important for examinees to do their best may also be helpful. In other instances—for example, 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.

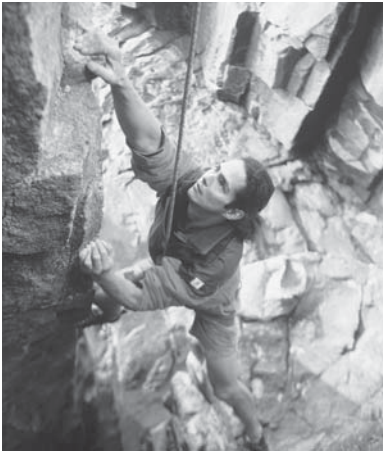
After a test administration, test users have many obligations as well. These obligations range from safeguarding the test protocols to conveying the test results in a clearly understandable fashion. In addition, there are other obligations such as those related to scoring the test. If a test is to be scored by people, scorers need to be in agreement about scoring criteria. Interpreting the test results and seeing to it that the test data are used in accordance with established procedures and ethical guidelines constitute further obligations of test users. If there were third parties present during testing or if anything else that might be considered out of the ordinary happened during testing, it is the test user's responsibility to make a note of such events on the report of the testing.

Assessment of people with disabilities People with disabilities are assessed for exactly the same reasons that people with no disabilities are assessed: to obtain employment, to earn a professional credential, to be screened for psychopathology, and so forth. A number of laws have been enacted that affect the conditions under which tests are administered to people with disabling conditions. For example, one law mandates the development and implementation of “alternate assessment” programs for children who, as a result of a disability, could not otherwise participate in state- and districtwide assessments. The law left defining *alternate assessment* up to the individual states or their local school districts. It is the responsibility of the states (or school districts) to define who requires alternate assessment, how such assessments are to be conducted, and how meaningful inferences are to be drawn from the data derived from such assessments.

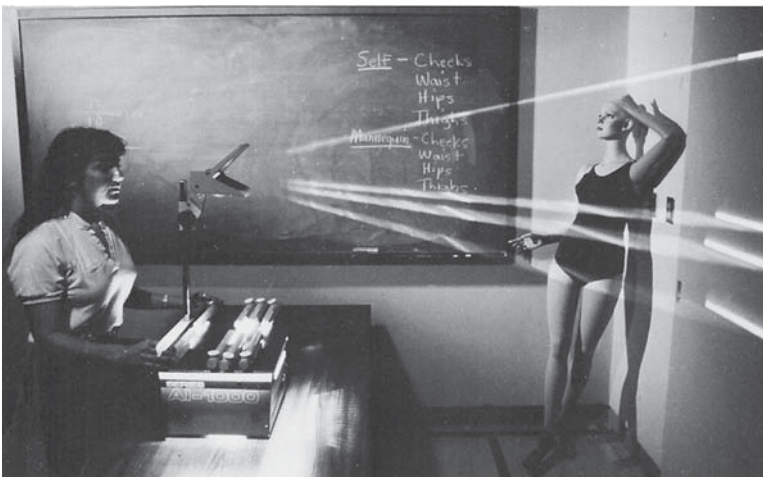
In general, alternate assessment is typically accomplished by means of some *accommodation* made to the assessee. The verb *to accommodate* may be defined as “to adapt, adjust, or make suitable.” In the context of psychological testing and assessment, **accommodation** may be defined as *the adaptation of a test, procedure, or situation, or the substitution of one test for another, to make the assessment more suitable for an assessee with exceptional needs.*

At first blush, the process of accommodating students, employees, or other testtakers with special needs might seem straightforward. For example, the individual 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. An individual with ADHD might have an extended evaluation time, with frequent breaks during periods of evaluation. Although this may all seem simple at first, it can actually become quite complicated.

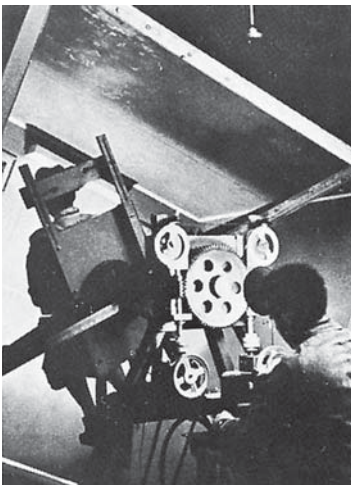
Consider, for example, the case of a student with a visual impairment who is scheduled to be given a written, multiple-choice test using an alternate procedure. There are several possible alternate procedures. For instance, the test could be translated into Braille and administered in that form, or it 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.



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.



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. Assesseees 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 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. Assesseees were asked questions designed to evaluate their dependence on or independence of visual cues.

Figure 1-5
The Wide World of Measurement



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 involves 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.



The Stresseraser is a handheld and self-administered biofeedback device designed to facilitate change in bodily relaxation. Vagus nerve functioning is monitored from the pulse of the index finger and fed back to the user through images on a screen. Users may then alter breathing and mental focus to affect sympathetic-parasympathetic functioning so as to obtain the therapeutic benefits associated with high heart-rate variability. The unit has the advantage of portability; it can be used to facilitate relaxation in a variety of settings.



Figure 1-6
Less-Than-Optimal Testing Conditions

In 1917, new Army recruits sat on the floor as they were administered the first group tests of intelligence—not ideal testing conditions by current standards.

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 are test scores affected by 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? Taking a step back from such complex issues, how do we define *alternate assessment*?

Given the complexities involved, we propose the following definition of this rather 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 assessment tool for another. Currently, there are many different ways to accommodate people with disabilities in an assessment situation (see this chapter's *Everyday Psychometrics*).

Everyday Accommodations

It has been estimated that as many as one in seven Americans has a disability that interferes with activities of daily living. In recent years, society has acknowledged more than ever before the special needs of citizens challenged by physical and/or mental disabilities. The effects of this ever-increasing acknowledgment are visibly evident: special access ramps alongside flights of stairs, captioned television programming for the hearing-impaired, and large-print newspapers, books, and magazines for the visually impaired. In general, there has been a trend toward altering environments to make individuals with handicapping conditions feel less challenged.

Depending on the nature of a testtaker's disability and other factors, modifications—referred to as *accommodations*—may need to be made in a psychological test (or measurement procedure) in order for an evaluation to proceed. Accommodation may take many different forms. One general type of accommodation involves *the form of the test as presented to the testtaker*. For example, a written test may be modified for presentation to a visually impaired testtaker by being set in larger type. Another general type of accommodation concerns *the way responses to the test are obtained*. For example, a speech-impaired individual may be accommodated by being allowed to write out responses in an examination that would otherwise be administered orally. Students with learning disabilities may be accommodated by being permitted to read test questions aloud (Fuchs et al., 2000).

Modification of the physical environment in which a test is conducted is yet another general type of accommodation. For example, a test that is usually group-administered at a central location may on occasion be administered individually to a disabled person in his or her home. *Modifications of the interpersonal environment in which a test is conducted* is another possibility (see Figure 1).

Which of many different types of accommodation should be employed? An answer to this question is typically approached by consideration of at least four variables:

1. the capabilities of the assessee;
2. the purpose of the assessment;
3. the meaning attached to test scores; and
4. the capabilities of the assessor.

The Capabilities of the Assessee

Which of several alternate means of assessment is best tailored to the needs and capabilities of the assessee? Case

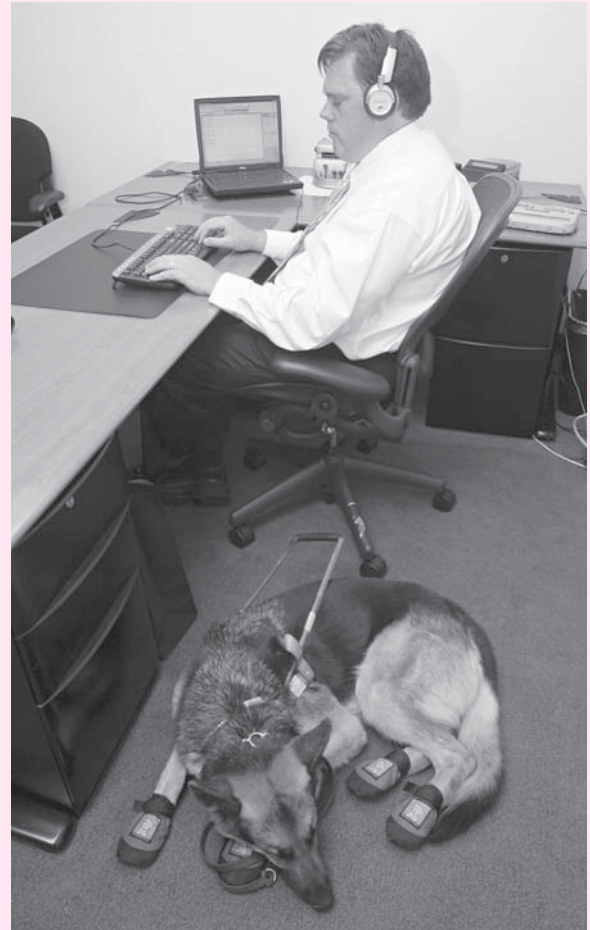


Figure 1
Modification of the Interpersonal Environment

An individual testtaker who requires the aid of a helper or working dog may require the presence of a third party (or animal) if a particular test is to be administered. In some cases, because of the nature of the testtaker's disability and the demands of a particular test, a more suitable test might have to be substituted for the test usually given if a meaningful evaluation is to be conducted.

history data, records of prior assessments, and interviews with friends, family, teachers, and others who know the assessee all can provide a wealth of useful information concerning which of several alternate means of assessment is most suitable.

(continued)

Everyday Accommodations (*continued*)**The Purpose of the Assessment**

Accommodation is appropriate under some circumstances and inappropriate under others. In general, one looks to the purpose of the assessment and the consequences of the accommodation in order to judge the appropriateness of modifying a test to accommodate a person with a disability. For example, modifying a written driving test—or a road test—so a blind person could be tested for a driver’s license is clearly inappropriate. For their own as well as the public’s safety, the blind are prohibited from driving automobiles. On the other hand, changing the form of most other written tests so that a blind person could take them is another matter entirely. In general, accommodation is simply a way of being true to a social policy that promotes and guarantees equal opportunity and treatment for all citizens.

The Meaning Attached to Test Scores

What happens to the meaning of a score on a test when that test has not been administered in the manner that it was designed to be? More often than not, when test administration instructions are modified (some would say “compromised”), the meaning of scores on that test becomes questionable at best. Test users are left to their own devices with regard to making interpretations from such data. Professional judgment, expertise, and, quite frankly, guesswork can all enter into the process of drawing inferences from scores on modified tests. Of course, a precise record of just how a test was modified for accommodation purposes should be made on the test report.

The Capabilities of the Assessor

Although most persons charged with the responsibility of assessment would like to think that they can administer an assessment professionally to most anyone, this is actually not the case. It is important to acknowledge that some assessors may experience a level of discomfort in the presence of people with particular disabilities, and this discomfort may affect their evaluation. It is also important to acknowledge that some assessors may require additional training prior to conducting certain assessments, including supervised experience with members of certain populations. Alternatively, the assessor may refer such assessment assignments to another assessor who has had more training and experience with members of a particular population.

A burgeoning scholarly literature has focused on various aspects of accommodation, including issues related to general policies (Burns, 1998; Nehring, 2007; Shriner, 2000; Simpson et al., 1999), method of test administration (Calhoon et al., 2000; Danford & Steinfeld, 1999), score comparability (Elliott et al., 2001; Johnson, 2000; Pomplun & Omar, 2000, 2001), documentation (Schulte et al., 2000), and the motivation of testtakers to request accommodation (Baldrige & Veiga, 2006). Before a decision about accommodation is made for any individual testtaker, due consideration must be given to issues regarding the meaning of scores derived from modified instruments and the validity of the inferences that can be made from the data derived.

Having considered some of the *who*, *what*, *how*, and *why* of assessment, it remains for us to raise the question of *where* to go for more 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, others provide detailed information regarding technical aspects, and still others provide critical reviews complete with discussion of the pros and cons of usage.

Test catalogues Perhaps one of the most readily accessible sources of information is a catalogue distributed by the publisher of the test. Because most test publishers make



Figure 1–7
Oscar Krisen Buros (1906–1978)

Buros is best remembered as the creator of the Mental Measurements Yearbook (MMY), a kind of Consumer Reports for tests and a much-needed source of “psychometric policing” (Peterson, 1997, p. 718). His work lives on at The Buros Institute of Mental Measurements, located at the University of Nebraska in Lincoln. In addition to the MMY, which is updated periodically, the institute publishes a variety of other test-related materials.

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 might require. Moreover, the catalogue’s objective 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 **test manual**, which is usually available from the test publisher. However, for security purposes, the test publisher will typically require documentation of professional training before filling an order for a test manual. Besides purchasing a manual from the publisher, the chances are good that a collection of popular test manuals is maintained somewhere within your university (the library or counseling center). If the test manual you seek is not available there, ask your instructor about how best to obtain a reference copy. In surveying various test manuals, you are likely to see that they vary not only in the details of how they were developed and deemed psychometrically sound but also in the candor with which they describe their own limitations.

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* 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 *17th Annual Mental Measurements Yearbook* published in 2007 (though the 18th cannot be far behind). The Buros Institute also disseminates a series of publications called *Tests in Print* that contains a listing of all commercially available English-language tests in print. This volume, which is also updated periodically, provides detailed information for each test listed, including test publisher, test author, test purpose, intended test population, and test administration time.

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 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; Turchik et al., 2007). Although 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 U.S. Department of Education and operated out of the University of Maryland, the ERIC Web site at www.eric.ed.gov 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 that 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 concerning 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 on CD-ROM (updated twice a year). PsycLAW is a free database, available to everyone, that contains discussions of selected topics involving psychology and law. It can be accessed at www.apa.org/psyclaw. For more information on any of these databases, visit APA's Web site at www.apa.org.

The world's largest private measurement institution is Educational Testing Service (ETS). This company, based in Princeton, New Jersey, maintains a staff of some 2,500 people, including about 1,000 measurement professionals and education specialists. These are the folks who bring you the Scholastic Aptitude Test (SAT) and the Graduate Record Exam (GRE), among many other tests. Descriptions of these and the many other tests developed by this company may be found at their Web site, www.ets.org.

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*

Table 1–3
Sources of Information About Tests: Some Pros and Cons

Information Source	Pros	Cons
Test catalogues available from the publisher of the test as well as affiliated distributors of the test	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.	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.
Test manual	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.	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. A test manual itself may be difficult to obtain by students, as its distribution may be restricted to qualified professionals.
Reference volumes such as the <i>Mental Measurements Yearbook</i> , available in bound book form or online	Much like a <i>Consumer Reports</i> for tests, contains descriptions and critical reviews of a 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.	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.
Journal articles	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.	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.
Online databases	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.	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."

Unpublished Experimental Mental Measures (Goldman & Mitchell, 2007) and *Tests in Microfiche*, administered by ETS. Yet another source of information—this one about sources of information—is Table 1–3.

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.

Self-Assessment

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

accommodation
 achievement test
 alternate assessment
 behavioral observation
 CAPA

case history data
 central processing
 collaborative psychological
 assessment
 consultative report

cut score
 diagnosis
 diagnostic test
 dynamic assessment
 extended scoring report

format
health psychology
informal evaluation
integrative report
interpretive report
interview
local processing
naturalistic observation
panel interview
portfolio
protocol
psychological assessment
psychological autopsy
psychological test

psychological testing
psychometrician
psychometrics
psychometrist
quality of life
rapport
role play
role-play test
school ability test
score
scoring
scoring report
simple scoring report
social facilitation

teleprocessing
test
test catalogue
test developer
test manual
testtaker
test user
therapeutic psychological
assessment
third parties in psychological
assessment
utility