Measuring cognitive performance has been a hallmark of clinical psychology since its origin (Wood, Garb, & Nezworski, 2007). Knowledge of a client’s level of cognitive functioning, including both strengths and deficits, can help a clinical psychologist with diagnosis and treatment of many presenting problems. Some assessments, such as those for specific learning disorder (formerly known as learning disabilities), intellectual disability (also known as intellectual developmental disorder, and formerly known as mental retardation), or giftedness, focus on cognitive issues from the start. In other assessments, such as those focusing on mood disorders or disruptive behavior, cognitive tests can provide important contextual information.

This chapter focuses on three types of tests, each related to cognitive functioning in some way but each with a distinct purpose. Intelligence tests measure a client’s intellectual abilities. Achievement tests, in contrast, measure what a client has accomplished with those intellectual abilities. Neuropsychological tests focus on issues of cognitive or brain dysfunction, including the effects of brain injuries and illnesses.
INTELLIGENCE TESTING

Classic Theories of Intelligence

The specific intelligence tests used by contemporary clinical psychologists are rooted in contrasting theories of intelligence. Each of these theories puts forth a different answer to the essential question: What is intelligence? Actually, clinical psychologists have never reached a consensus regarding the definition of intelligence. Experts in this area of clinical psychology have emphasized many abilities as central to intelligence: speed of mental processing, sensory capacity, abstract thinking, imagination, adaptability, capacity to learn through experience, memory, reasoning, and inhibition of instinct, to name a few (as summarized by Sternberg, 2000; Sternberg & Grigorenko, 2008; Wasserman & Tulsky, 2005). In the debate about defining intelligence, perhaps no specific issue has received as much attention as the singular versus plural nature of intelligence (Wasserman, 2018). In other words, is intelligence one thing or many things?

Charles Spearman: Intelligence Is One Thing

In the early 1900s, Charles Spearman proposed a theory: Intelligence is a singular characteristic. Spearman labeled this characteristic “g” for general intelligence and argued that it represented a person’s global, overall intellectual ability. His theory was based on research in which he measured many different, specific capabilities of his participants, including academic abilities and sensory-discrimination tasks. The primary finding was a strong correlation between this wide range of abilities, suggesting that a single factor underlies them all. Spearman acknowledged that more specific abilities (“s”) existed, but he argued that they played a relatively minor role in intelligence. Essentially, according to Spearman, intelligence was one thing (Brody, 2000; Revelle, 2015).

Louis Thurstone: Intelligence Is Many Things

Louis Thurstone was among the first and the strongest opponents to Spearman’s singular theory of intelligence. According to Thurstone, intelligence should not be understood as a single, unified ability but as numerous distinct abilities that have little relationship to one another. Among his other contributions in the first half of the 1900s, Thurstone was a pioneer of the statistical procedure called multiple factor analysis, which enabled him to identify underlying factors in a large data set. When he performed these statistical analyses on examinations of various intellectual abilities, he could have found one dominant factor underlying all abilities. Instead, he found several independent factors. These factors were given labels such as verbal comprehension, numerical ability, spatial reasoning, and memory. The specific names and number of the factors are less important,
however, than Thurstone’s fundamental point: Intelligence is not one thing; it is many things (Brody, 2000). Thus, according to Thurstone, if you know how capable someone is regarding, say, mathematics, you cannot predict with confidence how capable that person is regarding, say, verbal skills.

Eventually, Spearman and Thurstone each acknowledged the validity of the other’s arguments and came to somewhat of a compromise. They settled on a hierarchical model of intelligence in which specific abilities (“s”) existed and were important, but they were all at least somewhat related to one another and to a global, overall, general intelligence (“g”) (Willis, Dumont, & Kaufman, 2015a).

**More Contemporary Theories of Intelligence**

Other theories of intelligence have emerged since the debates between Spearman and Thurstone, most of which provide a fresh perspective on the singular versus plural issue (Brody, 2000; Davidson & Downing, 2000; Nisbett et al., 2012). For example, in the latter half of the 1900s, James Cattell proposed two separate intelligences: fluid intelligence—the ability to reason when faced with novel problems—and crystallized intelligence—the body of knowledge one has accumulated as a result of life experiences. With exactly two very broad categories of intelligence, Cattell’s theory falls somewhere between Spearman’s theory of a singular intelligence and Thurstone’s theory of many intelligences. More recently, new hierarchical models of intelligence have emerged as well, including John Carroll’s (2005) three-stratum theory of intelligence, in which intelligence operates at three levels: a single “g” at the top, eight broad factors immediately beneath “g,” and more than 60 highly specific abilities beneath these broad factors. This model, like most contemporary models of intelligence, not only acknowledges “g” but also recognizes that more specific abilities exist as well (Wasserman & Tulsky, 2005). As we see in this chapter, most contemporary intelligence tests mirror this view by producing a single overall score—one number to represent how intelligent a person is—in addition to a number of other scores representing more specific abilities. The Wechsler and Stanford-Binet intelligence tests are primary among these, and we will turn our attention to them soon. But it is worth mentioning that there are other intelligence tests besides the Wechsler and Stanford-Binet tests that are also widely used among clinical psychologists. Those tests include the Woodcock-Johnson tests, such as the Woodcock-Johnson IV Tests of Early Cognitive and Academic Development (for children aged 2.5–7) and the Woodcock-Johnson IV Tests of Cognitive Abilities (generally for clients aged 5–95), as well as the Kaufman Assessment Battery for Children-II (for children aged 3–18) (Drozdick, Singer, et al., 2018; Morrison, Singer, & Raiford, 2018; Schrank & Wendling, 2018a, 2018b). Each of these tests is paired with an achievement test created and typically co-normed by the same company (Woodcock-Johnson and Kaufman, respectively).
Metaphorically Speaking

If You’ve Watched Multi-Sport Athletes, You Understand the Challenges of Defining and Assessing Intelligence

Michael Jordan is undoubtedly one of the greatest basketball players of all time. His five MVP awards and six NBA championships offer indisputable evidence. But if we broaden the question—is Michael Jordan a great athlete?—we can begin to appreciate some of the challenges in defining ability, whether athletic or intellectual.

In 1993, after 9 extremely successful years in the NBA, Jordan decided (to the surprise of many) to leave basketball and focus on baseball. He signed on with the Birmingham Barons, a Double-A minor league team. His prowess did not easily transfer from the basketball court to the baseball diamond. Instead of the familiar highlights of acrobatic dunks and picturesque jump shots, sports fans saw Jordan flailing at curve balls and misjudging fly balls. His hitting and fielding improved slightly but not enough for him to advance any closer to the major leagues.

Should we expect that a person who excels at one sport will similarly excel at others? If so, perhaps each of us possesses a general athletic ability that influences our performance at any sport—a lot like Spearman’s concept of a general intelligence (“g”). If not, perhaps each of us has a number of specific athletic abilities that are unrelated to one another—a lot like Thurstone’s theory of specific (“s”), multiple intelligences.

Which theory does the example of Michael Jordan support? On the one hand, the contrast between Jordan’s basketball excellence and baseball mediocrity supports the idea that our abilities are independent. Indeed, in the history of athletics, there have been very few people (e.g., Jim Thorpe, Bo Jackson, Danny Ainge, Deion Sanders) who have reached the highest level of multiple sports. On the other hand, there are quite a few athletes who have reached the highest level of one sport and had some success (perhaps at the college or minor league level) at another sport as well, including Russell Wilson (football and baseball), Tim Tebow (football and baseball), Jim Brown (football and lacrosse), Stephen Curry (basketball and golf), and Jackie Robinson (baseball, basketball, football, and track). Michael Jordan may belong on that list, too. Despite focusing exclusively on basketball for many years, Jordan was able, with minimal preparation, to step onto a minor league baseball team and at least hold his own. He may not have excelled, but he did achieve some success at a level of baseball—a minor league team just two steps away from the majors—that far exceeds the talents of the vast majority of players. This would suggest that he possesses some fundamental, general athletic ability that applies, to some extent, to any sport. Perhaps Michael Jordan’s experiences best support the hierarchical models of ability, which acknowledge a general, across-the-board ability (of athletics or intelligence) as well as more specific talents (basketball vs. baseball, math vs. verbal skills) that are influenced by the general ability but may be somewhat independent as well.

In your personal experience, does athletic ability seem to be one thing or many things? Consider the athletic abilities of people you have known (or perhaps played with or against). Are the best at one sport typically among the best at other sports? What about intellectual ability? Among your friends and family, do people seem to have a singular, general intelligence that influences all their abilities, or many distinct intelligences that operate independently?
Wechsler Intelligence Tests

Since David Wechsler’s earliest attempts to measure intelligence in the early 1900s, the tests that bear his name have risen to prominence among clinical psychologists (Coalson, Raiford, Saklofske, & Weiss, 2010; Drozdick, Raiford, et al., 2018; G. Goldstein, 2008; Lichtenberger & Kaufman, 2009; Willis et al., 2015b). There are three separate Wechsler intelligence tests, each the most widely used among psychologists for its respective age range (Wright et al., 2017). The names and current editions of the Wechsler intelligence tests are as follows:

- Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV)
- Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V)
- Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition (WPPSI-IV)

Table 9.1 provides a brief description for each of these Wechsler intelligence tests, as well as the other intelligence, achievement, and neuropsychological tests covered in detail in this chapter.

<table>
<thead>
<tr>
<th>Table 9.1</th>
<th>At-a-Glance Information About the Tests Detailed in This Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Most Recent Edition</td>
</tr>
<tr>
<td><strong>Intelligence Tests</strong></td>
<td></td>
</tr>
<tr>
<td>Wechsler Adult Intelligence Scale (WAIS)</td>
<td>IV</td>
</tr>
<tr>
<td>Wechsler Intelligence Scale for Children (WISC)</td>
<td>V</td>
</tr>
<tr>
<td>Wechsler Preschool and Primary Scale of Intelligence (WPPSI)</td>
<td>IV</td>
</tr>
<tr>
<td>Stanford-Binet Intelligence Scales</td>
<td>V</td>
</tr>
<tr>
<td>Universal Nonverbal Intelligence Test (UNIT)</td>
<td>2</td>
</tr>
<tr>
<td><strong>Achievement Tests</strong></td>
<td></td>
</tr>
<tr>
<td>Wechsler Individual Achievement Test (WIAT)</td>
<td>III</td>
</tr>
<tr>
<td><strong>Neuropsychological Tests</strong></td>
<td></td>
</tr>
<tr>
<td>Halstead-Reitan Neuropsychological Test Battery (HRB)</td>
<td>I</td>
</tr>
<tr>
<td>Bender Visual-Motor Gestalt Test</td>
<td>II</td>
</tr>
<tr>
<td>NEPSY</td>
<td>II</td>
</tr>
</tbody>
</table>
Collectively, the three Wechsler intelligence tests cover virtually the entire life span. They vary slightly from one another, as necessitated by the demands of measuring intelligence at different ages. And, in practical terms, they are indeed separate tests rather than slight variants of one another. (They are purchased as separate kits with separate manuals, answer sheets, materials, etc.) However, the three Wechsler intelligence tests share many fundamental characteristics:

- They yield a single **full-scale intelligence score**, four or five **index scores**, and about a dozen (give or take a few, depending on optional subtests chosen) specific **subtest scores**. Together, this collection of scores indicates that the Wechsler tests employ a hierarchical model of intelligence (as discussed earlier), in which the full-scale intelligence score reflects a general, global level of intelligence (“g”) and the index/factor scores and subtest scores represent increasingly specific areas of ability (“s”). These scores allow clinical psychologists to focus broadly or narrowly when making interpretations regarding intellectual ability.

- They are administered one-on-one and face-to-face. In other words, the Wechsler tests cannot be administered to a group of examinees at the same time, nor are they entirely pencil-and-paper tests (e.g., multiple choice, true/false, essay) that examinees simply administer to themselves. Administration of the Wechsler intelligence tests is a structured interpersonal interaction requiring extensive training, typically received during graduate programs in clinical psychology (Raiford, Coalson, Saklofske, & Weiss, 2010).

- Each subtest is brief (lasting about 2–10 minutes) and consists of items that increase in difficulty as the subtest progresses. Most often, the subtests are designed such that examinees continue until they fail a predetermined number of consecutive items (or “max out,” to state it informally). The nature of the tasks performed varies widely across the dozen or so subtests but includes both verbal and nonverbal tasks. (More detailed descriptions of the subtests follow.)

- Although each of the three Wechsler tests has a small number of unique subtests, most subtests appear in all three tests as either core or supplemental subtests. These common subtests are described in detail in Table 9.2.

- Originally, the Wechsler tests were designed with two categories of subtests: verbal and performance. (“Performance” was essentially equivalent to “nonverbal.”) In more recent years, statistical tests including factor analyses have concluded that the subtests don’t, in fact, cluster together into those two groups. Instead, they cluster together in four or five, rather than two, factors. Thus, the original verbal/performance split has been replaced by four or five factors, each receiving contributions from several subtests (Groth-Marnat, 2009; L. G. Weiss, Saklofske, Coalson, & Raiford, 2010). Specifically, the WAIS has four factors (listed below), and the WISC and WPPSI each have five factors (essentially, they share three with the WAIS, but the Perceptual Reasoning Index of the WAIS is divided into Visual Spatial Index and Fluid Reasoning Index).
  - **Verbal Comprehension Index**—a measure of verbal concept formation and verbal reasoning
  - **Perceptual Reasoning Index**—a measure of fluid reasoning, spatial processing, and visual-motor integration
• **Working Memory Index**—a measure of the capacity to store, transform, and recall incoming information and data in short-term memory

• **Processing Speed Index**—a measure of the ability to process simple or rote information rapidly and accurately

  - They feature large, carefully collected sets of normative data. That is, the manual for each Wechsler test includes norms collected from about 2,000 people. These normative groups closely match recent U.S. Census data in terms of gender, age, race/ethnicity, and geographic region, among other variables (Zhu & Weiss, 2005). So when an examinee takes a Wechsler intelligence test, the examinee’s performance is compared with the performance of a large, same-age sample of individuals representing a wide-scale national population.

  - The full-scale and index scores generated by the Wechsler tests are “IQ” scores, meaning that they reflect an intelligence “quotient.” This quotient is the result of a division problem in which the examinees’ raw scores are compared with age-based expectations. For all Wechsler tests, mean IQ scores (full-scale and index) are 100, with a standard deviation of 15. For each subtest, a score of 10 is average, with a standard deviation of 3.

  - The Wechsler intelligence tests share a general approach to interpretation of scores. Assessors are instructed to first consider the full-scale IQ score. Next, they move on to interpret each index score in relation to the others and then the pattern of subtest scores. Finally, they note more detailed aspects of the testing, such as observable patterns of behavior that may have contributed to scores or inconsistent performance within a single subtest. Essentially, these successive steps allow the assessor to grasp the “big picture” of the intellectual profile before delving into increasingly detailed levels of analysis (Groth-Marnat, 2009).

  - The Wechsler intelligence tests are all backed by very impressive psychometric data. That is, a large number of empirical studies suggest that these tests have the characteristics that clinical psychologists should seek in any test: strong reliability and validity. They measure what they intend to measure, and they do so consistently (Canivez & Watkins, 2010; Groth-Marnat, 2009; Lichtenberger & Kaufman, 2004; Zhu & Weiss, 2005).

  - The Wechsler intelligence tests—and most other intelligence tests, for that matter—are used for a wide range of clinical applications, including evaluations that focus on issues of intellectual disability (intellectual developmental disorder), developmental delays, giftedness, educational and vocational planning, school placement and qualification, and other targeted assessment questions (Zhu & Weiss, 2005). They can also be used to provide general intelligence information in broader contexts, including a comprehensive assessment of a client whose presenting problems are more neuropsychological (e.g., Alzheimer’s disease), emotional (e.g., mood disorders), or behavioral (e.g., attention-deficit/hyperactivity disorder [ADHD]).

  - The Wechsler intelligence tests were among the first to become available on a digital platform as an alternative to the traditional pencil-and-paper format. Known as the “Q-interactive” system, this new method of administering and scoring Wechsler tests allows psychologists to use tablets (e.g., iPads) and promise to reduce both the overall length of the tests and the number of scoring errors that can happen when a person rather than a machine does the computations (S. W. Clark, Gulin, Heller, & Vrana, 2017).
However, it also brings a number of challenges, including the need for psychologists to be sufficiently tech-savvy, clients’ potential discomfort with the tablets, and financial expense for psychologists and graduate training programs (S. W. Clark et al., 2017; M. C. Roberts & Campbell, 2017). One study of graduate students trained to use the Q-interactive system found mixed reviews, with most preferring to learn the tests initially in pencil-and-paper format and only slightly more than half preferring to eventually switch to the Q-interactive format (Noland, 2017).

- Generally, the Wechsler intelligence tests have a number of notable strengths: They have impressive reliability and validity; they feature comprehensive and recent normative data; they cover an extremely wide age range; they provide full-scale, index, and subtest scores that have great clinical utility; and, at this point in their history, they are very familiar to most clinical psychologists. On the other hand, these tests have received criticism for some limitations: Some subtests may be culturally loaded or biased, the connection between the tests and day-to-day life (ecological validity) may be limited, and scoring can be complex or subjective on some subtests (Groth-Marnat, 2009; Zhu & Weiss, 2005). On the whole, however, the Wechsler tests are held in high esteem by clinical psychologists.

### Table 9.2 Subtests Common to Wechsler Tests of Intelligence

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Example Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary</td>
<td>Orally explain the meaning of a word</td>
</tr>
<tr>
<td></td>
<td>What does “consistent” mean?</td>
</tr>
<tr>
<td></td>
<td>What is an “intersection”?</td>
</tr>
<tr>
<td>Similarities</td>
<td>Orally explain how two things or concepts are alike</td>
</tr>
<tr>
<td></td>
<td>How are a door and a window alike?</td>
</tr>
<tr>
<td></td>
<td>How are success and failure alike?</td>
</tr>
<tr>
<td>Information</td>
<td>Orally answer questions focusing on specific items of general knowledge</td>
</tr>
<tr>
<td></td>
<td>On what continent is Spain?</td>
</tr>
<tr>
<td></td>
<td>How many cents is a quarter worth?</td>
</tr>
<tr>
<td>Comprehension</td>
<td>Orally answer questions about general social principles and social situations</td>
</tr>
<tr>
<td></td>
<td>Why is it important for people to show identification before being allowed to vote?</td>
</tr>
<tr>
<td></td>
<td>What are some advantages of using only the minimal amount of water necessary in our homes?</td>
</tr>
<tr>
<td>Block Design</td>
<td>Re-create a specific pattern or design of colored blocks</td>
</tr>
</tbody>
</table>
Table 9.1 Various Forms of the Porteus Maze Test

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Picture Completion</td>
<td>View a picture of a simple object or scene and identify the important part that is missing</td>
</tr>
<tr>
<td>Matrix Reasoning</td>
<td>View an incomplete matrix and select the missing portion from multiple choices provided</td>
</tr>
<tr>
<td>Coding</td>
<td>Using pencil and paper, repeatedly copy simple shapes/symbols in appropriate spaces according to a key provided</td>
</tr>
<tr>
<td>Symbol Search</td>
<td>Scan a group of visual shapes/symbols to determine if target shape(s)/symbol(s) appear in group</td>
</tr>
</tbody>
</table>

**Stanford-Binet Intelligence Scales—Fifth Edition**

The Stanford-Binet intelligence tests have a rich history in clinical psychology. The first editions of the Stanford-Binet intelligence test dominated the field in the early 1900s until the Wechsler tests began to provide competition. Although the Wechsler tests have taken a leading role in recent decades, the Stanford-Binet remains highly respected and offers an approach to assessing intelligence that is both similar to and different from that of the Wechsler tests (G. Goldstein, 2008).

The most recent revision of the Stanford-Binet, the **Stanford-Binet Intelligence Scales—Fifth Edition (SB5)**, is like the Wechsler tests in many ways. It is administered face-to-face and one-on-one. It employs a hierarchical model of intelligence and therefore yields a singular measure of full-scale IQ (or “g”), five factor scores, and many more specific subtest scores. It features the same means (100) and standard deviations (15) as the Wechsler intelligence tests for its full-scale and factor scores. Its psychometric data, including reliability and validity, are similarly strong.

The SB5 differs from the Wechsler tests in some important ways, however. Rather than three separate tests for three different age ranges, the SB5 covers the entire life span.
span (ages 2–85+) as a single test. The normative sample is like those of the Wechsler tests in that it matches recent U.S. Census data on important variables, but it additionally includes normative data from individuals with specific relevant diagnoses, including learning problems, intellectual disability (intellectual developmental disorder), and ADHD. Its subtests include extensions at the low and high ends—in other words, a greater number of very easy and very difficult items—as an effort to more accurately assess people at the extremes, including those who may have intellectual disability or be gifted (Kamphaus & Kroncke, 2004).

Perhaps the most important difference between the SB5 and the Wechsler tests involves their specific factors and subtests. Whereas the Wechsler tests feature four or five factors, each of which yields an index score, the SB5 features exactly five, described briefly here:

- **Fluid Reasoning**—the ability to solve novel problems
- **Knowledge**—general information accumulated over time via personal experiences, including education, home, and environment
- **Quantitative Reasoning**—the ability to solve numerical problems
- **Visual-Spatial Processing**—the ability to analyze visually presented information, including relationships between objects, spatial orientation, assembling pieces to make a whole, and detecting visual patterns
- **Working Memory**—the ability to hold and transform information in short-term memory

Each of these five factors is measured both verbally and nonverbally, a deliberate innovation that is new to the SB5 and to IQ tests more generally. Each of these 10 areas (five factors, each measured both verbally and nonverbally) is assessed by one to three specific types of items (Kamphaus & Kroncke, 2004).

Overall, current editions of the Wechsler tests may have become more widely used than the current edition of the Stanford-Binet, but the SB5 continues to hold a similar position of respect and, in some settings, popularity as it has for the past century.

**Additional Tests of Intelligence: Addressing Cultural Fairness**

Especially in recent years, one of the primary criticisms of the Wechsler tests, the SB5, and other renowned IQ tests has centered on issues of cultural fairness. Specifically, these tests have been described as featuring numerous subtests, especially those relying on verbal skills, that place individuals from minority cultural groups at a disadvantage. In other words, to the extent that an intelligence test is based on culture-specific concepts, it may unfairly assess the intelligence of people of other cultures (Ortiz, Piazza, Ochoa, & Dynda, 2018). Certainly, the authors of prominent tests, including the Wechsler tests and the SB5, have made significant efforts to make recent editions of their tests less culturally biased or loaded. But entirely new tests have also emerged, with the explicit purpose of measuring IQ in a more culturally fair way.
A leading example of such a test is the Universal Nonverbal Intelligence Test-2 (UNIT-2; Bracken & McCallum, 2009, 2015; McCallum & Bracken, 2005, 2018). Originally published in 1996 and revised in 2015, the UNIT-2 is a completely language-free test of intelligence. It requires no speaking or shared understanding of language between the person administering the test and the person taking it. Like the Wechsler and Stanford-Binet intelligence tests, the UNIT-2 is administered one-on-one and face-to-face, but rather than using verbal instructions, the examiner presents instructions via eight specific hand gestures taught in the test manual and demonstrated in an accompanying video. Additionally, the responses of the examinee all consist of either pointing with fingers or minor manipulation of objects with hands or fingers.

The UNIT-2 is appropriate for clients aged 5 to 21 years and was normed on 1,800 people who match recent U.S. Census data in terms of age, sex, race, parent education, community size, geographic region, and ethnicity. A small body of psychometric data on the UNIT and UNIT-2 has been published and suggests generally acceptable reliability and validity (Bracken & McCallum, 2015).

The UNIT-2 consists of six subtests organized into a two-tiered model of intelligence. The two tiers are identified as Memory and Reasoning. The three subtests contributing to the memory tier are

- Object Memory, in which the examinee views a visual assortment of common objects for 5 seconds and then views a larger array and identifies the objects from the first array;
- Spatial Memory, in which the examinee recalls the placement of colored chips on a three-by-three or four-by-four grid; and
- Symbolic Memory, in which the examinee recalls and re-creates sequences of visually presented symbols.

The three subtests contributing to the reasoning tier are

- Cube Design, in which the examinee arranges colored blocks in a specific three-dimensional design;
- Mazes, in which the examinee completes traditional maze puzzles; and
- Analogic Reasoning, in which the examinee solves analogy problems that are presented visually rather than verbally.

The UNIT-2 is not without its shortcomings. It assesses a more limited range of abilities than more traditional IQ tests; it is appropriate only for young clients (no version for preschool children or adults older than 21 has yet been developed); and its psychometric data, although encouraging, is limited in quantity. Additionally, as a relative newcomer to the intelligence test field, it is nowhere near as well established or popular as the standard Wechsler or Stanford-Binet tests, so the structure and format of the UNIT-2 remain unfamiliar to many clinical psychologists. (And the Wechsler family of tests includes their own version of a nonverbal intelligence test: the Wechsler Nonverbal Scale of Ability. It includes only minimal spoken instructions at the outset and mostly pictorial directions for clients throughout the subtests [Naglieri & Otero, 2018].) Such limitations are common...
to nonverbal or “culture-free” tests more generally (Ortiz & Dynda, 2005). However, the development and increasing acceptance of the UNIT-2 and similar tests represent a significant step forward in the culturally sensitive practice of intelligence assessment. Considering the large numbers of non-English-speaking people in the United States, clinical psychologists will increasingly require intelligence tests that are applicable and fair to clients of a broad cultural and linguistic range (Pieterse & Miller, 2010).

**ACHIEVEMENT TESTING**

**Achievement Versus Intelligence**

Clinical psychologists assess clients’ intelligence, but with separate tests they also assess clients’ achievement. What’s the difference between intelligence and achievement? As we have discussed, intelligence refers to a person’s cognitive capacity. In short, intelligence is what a person can accomplish intellectually. In contrast, achievement is what a person has accomplished, especially in the kinds of subjects that people learn in school, such as reading, spelling, writing, or math. Achievement tests typically produce age- or grade-equivalency scores as well as standard scores (D. Schneider & Mather, 2015).

Prior to DSM-5, the comparison of intelligence and achievement was a key factor in determining learning disabilities. In DSM-5, the terminology changed (“learning disability” is now called **specific learning disorder**), and the definition has changed as well, such that the primary comparison is between the person’s achievement (as measured by achievement tests and performance at school or work) and expected levels of achievement for people of the same age (American Psychiatric Association, 2013; Hogan & Tsushima, 2016).

When clinical psychologists conduct tests of intelligence or achievement (or any other kind of assessment), they typically explain their findings in a written psychological report. Psychological reports are the formal means by which they communicate their findings to others. Depending on the case, these “others” can represent a wide range of individuals: clients, parents of child clients, teachers, lawyers, judges, physicians, or other mental health professionals, just to name a few. It is important that clinical psychologists not only write intelligible, accurate reports that address the reason for referral but that they also “know their audience” in order to customize their reports in terms of readability, selection of relevant information, and recommendations (Blais & Smith, 2008; Goldfinger & Pomerantz, 2010).

Achievement tests come in many varieties. Some measure a single area of achievement in detail, such as the KeyMath or the Gray Oral Reading achievement tests. Others are more comprehensive, featuring a wide range of subtests. Examples of these include the Woodcock-Johnson Tests of Achievement (WJ-ACH), the Kaufman Test of Educational Achievement (KTEA), and the Wechsler Individual Achievement Test (WIAT),...
Considering Culture

Defining Intelligence Around the World

What is intelligence? It depends on the cultural values of those we ask. When we direct the question to people outside traditional Western culture, the answer sometimes features characteristics that are quite different from definitions that Spearman, Thurstone, Cattell, and Carroll have proposed (as summarized by Sternberg, 2000; Sternberg & Grigorenko, 2008).

- In some societies in Africa (e.g., Zambia, Mali, and Kenya), intelligence consists largely of skills that preserve harmony in interpersonal relationships, both between and within groups. Interviews with some residents of Zambia, for example, suggest that cooperation, deference, respect for elders, and acceptance of social responsibilities characterize intelligent people.

- In Zimbabwe, the word for intelligence—ngware—literally translates into a prudent and cautious approach to life and especially to social relationships.

- In some Asian cultures, the definition of intelligence also involves heavy doses of social responsibility and benevolence. More specifically, Taoist conceptions of intelligence highlight humility, independent (rather than conventional) standards of judgment, and thorough knowledge of self.

- The emphasis on social duties as central to intelligence appears in some Hispanic cultures as well. In fact, in a study of parents of schoolchildren in San Jose, California, in the 1990s, parents of Hispanic descent rated social competence as more closely related to intelligence than did parents of European descent (Okagaki & Sternberg, 1993).

- The Western emphasis on speed of mental processing is not shared by all ethnic groups. In fact, some ethnic groups may value depth of thought more highly than speed of thought and may look unfavorably or doubtfully on work performed very quickly.

This variety of defining characteristics of intelligence raises a number of important questions. Is the definition of intelligence completely dependent on cultural context, or are some aspects of intelligence universal? To what extent should intelligence tests reflect the alternate definitions of intelligence held around the world or around the United States? Where should we draw the line between personality traits and intelligence? And, as Spearman and Thurstone debated many years ago, is intelligence one thing or many things?
Wechsler Individual Achievement Test—Third Edition

The Wechsler Individual Achievement Test—Third Edition (WIAT-III) is a comprehensive achievement test for clients aged 4 to 50 years. Like the Wechsler intelligence tests, the WIAT-III is administered face-to-face and one-on-one.

The WIAT-III measures achievement in four broad areas: reading, math, written language, and oral language. Each of these broad areas is assessed by two to four subtests. The Oral Language Composite derives from two subtests: Listening Comprehension (paying attention to orally presented information and answering questions about it) and Oral Expression (using speech to repeat spoken material, create stories about presented pictures, provide directions, etc.). The Reading Composite derives from scores on such subtests as Word Reading (reading isolated words), Pseudoword Decoding (using phonetic skills to sound out nonsense words, such as *plore* or *tharch*), and Reading Comprehension (reading sentences or passages and answering questions about their content). The Mathematics Composite derives subtests such as Numerical Operations (written math problems) and Math Problem Solving (word problems, numerical patterns, statistics and probability questions, etc.). The Written Language Composite derives subtests such as Spelling (increasingly difficult words) and Sentence Composition, Essay Composition (constructing sentences, paragraphs, or essays as instructed; Lichtenberger & Breaux, 2010).

The WIAT-III yields standard scores on the same scale as most intelligence tests: a mean of 100 and a standard deviation of 15. It also yields age and grade equivalencies for each subtest. The test was standardized on 3,000 people who were chosen to match recent U.S. Census data in terms of sex, age, race/ethnicity, geographic region, and parent education level. Many of the individuals in the standardization sample also took the age-appropriate Wechsler intelligence scale, so the WIAT-III is “linked” to the Wechsler IQ tests, which enhances the validity of comparisons between these two types of tests. Overall, the reliability and validity data supporting the WIAT-III are quite strong (Lichtenberger & Breaux, 2010).

Neuropsychological testing represents a specialized area of assessment within clinical psychology, typically practiced by clinical psychologists whose training includes extra training in neuropsychology during graduate school courses, the predoctoral internship, and the postdoctoral internship (Hill & Westervelt, 2015; Spencer & Adams, 2016). The intent of neuropsychological tests is to measure cognitive functioning or impairment of the brain and its specific components or structures. Medical procedures such as computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) scans can show that parts of the brain look abnormal, but neuropsychological
tests can indicate how the brain is actually functioning. Such tests are especially useful for targeted assessment of problems that might result from a head injury, prolonged alcohol or drug use, or a degenerative brain illness. They can also be used to make a prognosis for improvement, plan rehabilitation, determine eligibility for accommodations at school or work, and establish a baseline of neuropsychological abilities to be used as a comparison at a later time (Dawson & Jacquin, 2015b; Golden, 2008; Hebben & Milberg, 2009; McCaffrey, Lynch, & Westervelt, 2011).

Some neuropsychological testing procedures are lengthy, comprehensive batteries that include a broad array of subtests. The patterns of scores on these subtests can go a long way toward pinpointing specific cognitive weaknesses. Other neuropsychological tests are much briefer and are typically used as screens for neuropsychological impairment rather than as full-fledged neuropsychological assessment tools. Over time, the field has evolved from a “fixed-battery phase” in which psychologists typically use the same standard set of tests for most clients to a more “flexible-battery phase” in which psychologists create a more customized battery by picking and choosing tests after considering the specific referral questions and areas of concern (Hale, Wilcox, & Reddy, 2016).

Psychologists who use neuropsychological tests have also become more increasingly aware of neuropsychological assessment anxiety, an experience in which clients’ performance on these tests is influenced by the anxiety in the moment. One review of dozens of studies on the subject found that about two thirds of those studies identified a connection between performance and anxiety, with some areas of performance (e.g., attention, verbal memory, working memory) being especially vulnerable to decrease (Dorenkamp & Vik, 2018).

Let’s consider several examples of full neuropsychological batteries and briefer neuropsychological measures.

**Full Neuropsychological Batteries**

The Halstead-Reitan Neuropsychological Battery (HRB) is a battery of eight standardized neuropsychological tests. It is suitable for clients of age 15 years and older, but alternate versions are available for younger clients. The HRB (also known as HRNB) is administered only as a whole battery; its components are not to be administered separately. Thus, it is a thorough (and rather lengthy) neuropsychological battery. Essentially, its primary purpose is to identify people with brain damage and, to the extent possible, provide detailed information or hypotheses about any brain damage identified, including specific cognitive impairments or physiological regions of the brain that may be deficient. The findings of the HRB can help in diagnosis and treatment of problems related to brain malfunction (Broshek & Barth, 2000).

Some of the eight tests in the HRB involve sight, whereas others involve hearing, touch, motor skills, and pencil-and-paper tasks. An example of one of the tests of the HRB is the Trail Making Test, which resembles the familiar “dot-to-dot” puzzles that children complete, but this test is timed, contains both numbers and letters, and produces a rather haphazard line instead of an identifiable figure or shape. A second example is the Category Test, in which a client sees a pattern of shapes and designs on a screen and presses the number key (1–4) that is suggested by the pattern. Clients hear a bell when correct and a buzzer when wrong, and their successes and failures are noted...
by the examiner. The Finger Tapping (or Finger Oscillation) Test is a third example; in this test, clients tap a single typewriter key as rapidly as they can with the index finger for 10-second intervals. The number of taps they can produce, averaged across multiple attempts with each hand, estimates motor speed (G. Goldstein & Sanders, 2004; Reitan & Wolfson, 2004).

The examiner using the HRB compares a client’s test scores with published norms to assess overall performance and also with each other to determine the client’s own relative strengths and weaknesses. Interpretation of results can include detailed inferences about specific neuropsychological pathologies and the localization of problems in cognitive functioning (e.g., a particular lobe or hemisphere of the brain).

In the decades since it was introduced, the HRB has been evaluated more than any other comprehensive neuropsychological test. These studies suggest strongly that the HRB and each of its tests have been established as reliable and valid. Additional strengths of the HRB include its comprehensiveness and clinical usefulness. Drawbacks of the HRB include its length (and corresponding expense), inflexibility (as a fixed battery), and a limited overlap with real-life, day-to-day tasks (Broshek & Barth, 2000; Hebben & Milberg, 2009).

The HRB may be among the most commonly administered comprehensive neuropsychological batteries (Guilmette & Faust, 1991; Rabin, Paolillo, & Barr, 2016), but another long-standing, respected and popular alternative is the Luria-Nebraska Neuropsychological Battery (LNNB). The LNNB is similar to the HRB in that it is a wide-ranging test of neuropsychological functioning. It consists of 12 scales, with a similar range to that of the HRB. A primary difference is the LNNB’s emphasis on qualitative data in addition to quantitative data. In other words, to a greater extent than the HRB, the LNNB relies on qualitative written comments from the examiner about the testing process. These comments describe what the examiner observed about the client, such as problems comprehending the test (e.g., confusion, poor attention), how or why the client is missing items (e.g., slow movement, sight or hearing problems, speech flaws), or unusual behaviors (e.g., inappropriate emotional reactions, hyperactivity, distraction). Another difference between the HRB and the LNNB is that the LNNB tends to be slightly briefer. As for the HRB, a strong body of psychometric data supports the LNNB’s reliability and validity (Golden, 2004; Golden, Freshwater, & Vayalakkara, 2000; Hebben & Milberg, 2009).

The NEPSY-II is one of the few batteries of neuropsychological tests designed specifically for children. Specifically, it is suited for use in people between 3 and 16 years old. It is based on the same general principles put forth by the creator of the Luria-Nebraska test (Alexander Luria). It includes 32 separate subtests across 6 different categories: attention and executive functioning, language, memory and learning, sensorimotor, social perception, and visuospatial processing. Psychologists have flexibility in determining exactly how to use the various components of the test (Matthews & Davis, 2018).

**Brief Neuropsychological Measures**

The Bender-Gestalt test, originally published in 1938 and currently available as the Bender Visual-Motor Gestalt Test—Second Edition (Bender-Gestalt-II), is the most commonly used neuropsychological screen among clinical psychologists.
The test is a straightforward copying task: The client is given a pencil, blank paper, and nine simple geometric designs (primarily made of combinations of circles, dots, lines, angles, and basic shapes) and is asked to copy each design as accurately as possible. It measures visuocognitive abilities, which are also commonly known as perceptual-motor or visual-spatial skills (Lacks, 2000). The current edition is quite similar to the original Bender-Gestalt, but it offers memory tasks and additional stimuli.

This test is remarkably brief—on average, it takes only 6 minutes to administer (Lacks, 2000). It is appropriate for clients of any age above 3 years. For these reasons, clinical psychologists frequently include it as a quick “check” for neuropsychological problems. In other words, if it is already established that the client being evaluated has or is strongly suspected to have a neuropsychological problem, it is likely that the clinical psychologist will select a more comprehensive battery, such as the HRB or the LNNB, rather than the Bender-Gestalt. However, when the evaluation is for another purpose and the clinical psychologist simply wants a rapid appraisal of overall neuropsychological functioning, the Bender-Gestalt can be a good choice. Its results cannot specify locations of brain damage, but poor performance on the Bender-Gestalt can suggest brain damage in a diffuse way. As such, it can alert the clinical psychologist to the general presence of neuropsychological problems, and more thorough testing can subsequently be conducted.

Poor performance on the Bender-Gestalt is indicated by a variety of errors that clients may make in copying the figures, including figures in which details are missing or notably inexact; figures that collide with each other on the page; inability to accurately “close” shapes such as circles or squares; disproportionate size of a figure or part of a figure; and angles in copied figures that do not match the angles in the originals (Lacks, 1999).

Another common neurological screen among clinical psychologists is the Rey-Osterrieth Complex Figure Test, originally published in 1941. This test is also a brief pencil-and-paper drawing task, but it involves only a single, more complex figure. The Rey-Osterrieth also features the use of pencils of different colors at various points in the test; this way, the examiner can trace the client’s sequential approach to this complex copying task. It also includes a memory component, in which clients are asked 3 to 60 minutes after copying the form to reproduce it again from memory (Helmes, 2000; Lacks, 2000).

The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, 1998) is a neuropsychological screen that focuses on a broader range of abilities than does either the Bender-Gestalt or the Rey-Osterrieth. The RBANS tests not only visuomotor abilities but also verbal skills, attention, and visual memory. It takes 20 to 30 minutes to complete and includes 12 subtests in 5 categories: immediate and delayed memory, visuospatial/constructional, language, attention, and delayed memory. The subtests involve such tasks as learning a list of 10 words presented orally, naming pictures of various objects, recalling an orally presented list of numbers, recalling a story told 20 minutes earlier, and (like the Bender-Gestalt and Rey-Osterrieth) copying of visual figures. The RBANS yields a total score as well as scores for each of the 12 subtests and 5 indices into which they are organized (Groth-Marnat, 2009; Groth-Marnat & Wright, 2016). A revision, called the RBANS Update, was published in 2012 and enables the test to be administered to children as young as 12 years old (Hale et al., 2016; Randolph, 2012).
Although it is used for a variety of other purposes as well, the Wechsler Memory Scale—Fourth Edition (WMS-IV; Wechsler, 2009a, 2009b) is frequently utilized for neuropsychological purposes. The WMS-IV is a memory test often used to assess individuals aged 16 to 90 who are suspected of having memory problems due to brain injury, dementia, substance abuse, or other factors. It assesses both visual and auditory memory across its seven subtests. It also assesses both immediate and delayed recall. One of the WMS-IV subtests, Logical Memory, involves the client hearing a story read aloud and then trying to recall the story both immediately and after a 20- to 30-minute delay. Another subtest, Visual Reproduction, involves presenting abstract visual figures to the client and asking him or her to reproduce them after they have been removed. For adults over 65 years of age, the WMS-IV offers a version that is shorter in duration, therefore minimizing the impact of fatigue on test results (Drozdick, Holdnack, & Hilsabeck, 2011; Drozdick, Raiford, et al., 2018; Holdnack & Drozdick, 2010).

A final note about neuropsychological testing: As with any other kind of assessment, it is essential for clinical psychologists to develop and use cultural competence when selecting, administering, scoring, and interpreting neuropsychological tests. Many neuropsychological instruments were standardized (either originally or currently) on European Americans, and their form or structure can emphasize European American values. When working with diverse clients, it is important to caution against overpathologizing and to consider (or consider developing) culture-specific norms whenever possible (Horton, 2008).

CHAPTER SUMMARY

Clinical psychologists use intelligence tests, achievement tests, and neuropsychological tests to assess various intellectual capacities of their clients. The most widely accepted intelligence tests endorse a hierarchical model of intelligence, as indicated by the fact that they yield a single overall intelligence score as well as more specific index, factor, or subtest scores. The Wechsler intelligence tests, including the WAIS-IV for adults, the WISC-V for school-age children, and the WPPSI-IV for preschoolers, rank among the most commonly used and psychometrically sound measures of intelligence. Each Wechsler test features about a dozen subtests grouped into four or five factors, each of which contributes to the full-scale IQ. The Stanford-Binet intelligence test, which covers the entire age range in a single version, is also frequently used, especially among clinical psychologists who may be testing clients at the extremes of the intelligence range. Concerns about the cultural fairness of prominent intelligence tests has fostered the development of less culturally dependent tests such as the UNIT-2, which does not depend on linguistic compatibility between examiner and client. Whereas intelligence tests measure abilities, achievement tests such as the WIAT-III measure accomplishments, particularly in core academic areas such as math and reading. Achievement tests are commonly used when assessing for specific learning disorder. Neuropsychological tests are intended to measure cognitive function and dysfunction and, in some cases, to localize impairment to a particular region of the brain. Some neuropsychological measures are lengthy and comprehensive, such as the HRB, LNNB, and NEPSY-II; others, such as the Bender-Gestalt-II, Rey-Osterrieth Complex Figure Test, and RBANS, are briefer screens for neuropsychological problems.
KEY TERMS AND NAMES

achievement 12
achievement tests 1
Bender Visual-Motor Gestalt Test—Second Edition
(Bender-Gestalt-II) 16
John Carroll 3
James Cattell 3
crystallized intelligence 3
cultural fairness 3
fluid intelligence 3
fluid reasoning 000
full-scale intelligence score 6
“g” 2
Halstead-Reitan Neuropsychological Battery (HRB) 15
hierarchical model of intelligence 3
index scores 6
intelligence 2
intelligence tests 1
knowledge 000
Luria-Nebraska Neuropsychological Battery (LNNB) 16
NEPSY-II 16
neuropsychological tests 1
normative data 7
Perceptual Reasoning Index 6
Processing Speed Index 7
quantitative reasoning 10
Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) 20
Rey-Osterrieth Complex Figure Test 17
“s” 2
Charles Spearman 2
specific learning disorder 12
Stanford-Binet Intelligence Scales—Fifth Edition (SB5) 9
subtest scores 6
three-stratum theory of intelligence 3
Louis Thurstone 2
Universal Nonverbal Intelligence Test-2 (UNIT-2) 000
Verbal Comprehension Index 6
visual-spatial processing 10
David Wechsler 5
Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV) 000
Wechsler Individual Achievement Test—Third Edition (WIAT-III) 000
Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V) 000
Wechsler Memory Scale—Fourth Edition (WMS-IV) 18
Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition (WPPSI-IV) 000
working memory 10
Working Memory Index 7

CRITICAL THINKING QUESTIONS

1. In your opinion, what aspects of intelligence do contemporary intelligence tests overemphasize?

2. In your opinion, what aspects of intelligence do contemporary intelligence tests overlook?

3. To what extent do you agree with the idea that intelligence is a single ability ("g"), as opposed to separate unrelated abilities ("s")?

4. Which subtests of the Wechsler intelligence tests seem most culturally fair? Which seem most potentially culturally unfair?

5. Why is it important for clinical psychologists who do not specialize in neuropsychology to be competent in the use of brief neuropsychological screens?

LOOKING TOWARD GRADUATE PROGRAMS

The following table is a list of index entries in the Insider’s Guide to Graduate Programs in Clinical and Counseling Psychology (Sayette & Norcross, 2018) relevant to the topics in this chapter (see page XX in
Chapter 1 for more information about the *Insider’s Guide*). Keep in mind a few things as you consider these listings: They are current as of the 2018/2019 edition of the *Insider’s Guide* (which typically revises every two years); they may have changed as faculty members retire, take new jobs, or develop new interests; and a graduate program’s own website may have more extensive or up-to-date information.

<table>
<thead>
<tr>
<th>Research Areas</th>
<th>Specialty Clinical and Practica Sites</th>
<th>Program Concentrations and Tracks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer’s Disease</td>
<td>Assessment/Testing</td>
<td>Assessment/Testing</td>
</tr>
<tr>
<td>Assessment/Diagnosis</td>
<td>Neuroimaging</td>
<td>Neuropsychology</td>
</tr>
<tr>
<td>Brain Injury/Head Injury</td>
<td>Neuropsychology</td>
<td>Neuroscience</td>
</tr>
<tr>
<td>Neuroimaging/Functional Neuroimaging</td>
<td>Traumatic Brain Injury</td>
<td></td>
</tr>
<tr>
<td>Neuropsychology</td>
<td>Miscellaneous—Alzheimer’s Dementia</td>
<td></td>
</tr>
<tr>
<td>Miscellaneous and Other—Clinical Neuroscience</td>
<td>Miscellaneous—Gifted Evaluation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Miscellaneous—Neurological Disorders</td>
<td></td>
</tr>
</tbody>
</table>

**KEY JOURNALS**

Links available at edge.sagepub/pomerantz5e

*Assessment*

http://asm.sagepub.com/

*Neuropsychology*


*The Clinical Neuropsychologist*

http://www.tandfonline.com/loi/ntcn20#.V08kr_krLX4

*Psychological Assessment*


*Intelligence*

http://www.journals.elsevier.com/intelligence

**STUDENT STUDY SITE RESOURCES**

Visit the study site at edge.sagepub/pomerantz5e for these additional learning tools:

- Self-quizzes
- eFlashcards
- Culture expert interviews
- Full-text SAGE journal articles
- Additional web resources
- Mock assessment data