

**60-5** To what extent is intelligence related to brain anatomy?

- Some studies have found a positive correlation between intelligence score and brain size and activity, especially in the frontal and parietal lobes.
- Ample gray matter and white matter enable efficient communication between brain circuits.

**Multiple-Choice Questions**

1. According to Robert Sternberg, what kind of intelligence is assessed by traditional intelligence tests?
  - a. Linguistic
  - b. Practical
  - c. Creative
  - d. Spatial
  - e. Analytical
2. According to Charles Spearman and others, which of the following underlies specific mental abilities and is measured by every task on an intelligence test?
  - a. Savant syndrome
  - b. General intelligence (*g*)
  - c. Factor analysis
  - d. Intelligence
  - e. Emotional intelligence
3. Of the following, which term best describes the condition in which a person with limited mental ability excels at a specific skill such as computation?
  - a. Savant syndrome
  - b. *g* factor
  - c. Creative intelligence
  - d. Emotional intelligence
  - e. Street smarts
4. Which of the following is not a component of emotional intelligence?
  - a. Understanding emotions
  - b. Perceiving emotions
  - c. Using emotions
  - d. Managing emotions
  - e. Inventing emotions

**Practice FRQs**

1. Give a summary, a strength, and a weakness of Charles Spearman's idea of general intelligence.
2. Name and describe Robert Sternberg's three intelligences.

**(3 points)****Answer**

**1 point:** General intelligence is basic intelligence that predicts our abilities in varied academic areas.

**1 point:** A strength of this idea is that different abilities, such as verbal and spatial, tend to correlate.

**1 point:** A weakness of this idea is that human abilities are too diverse to be explained by a single general intelligence factor.

**60-6** To what extent is intelligence related to neural processing speed?

- People who score high on intelligence tests tend also to have agile brains and score high in speed of perception and speed of neural processing.
- The direction of correlation has not been determined, and some third factor may influence both intelligence and processing speed.

# Module 61

## Assessing Intelligence

**Module Learning Objectives**

- 61-1** Discuss the history of intelligence testing.
- 61-2** Distinguish between aptitude and achievement tests.
- 61-3** Explain the meaning of standardization, and describe the normal curve.
- 61-4** Explain the meanings of reliability and validity.

**H**ow do we assess intelligence? And what makes a test credible? Answering these questions begins with a look at why psychologists created tests of mental abilities and how they have used those tests.

**The Origins of Intelligence Testing**

- 61-1** When and why were intelligence tests created?

Some societies concern themselves with promoting the collective welfare of the family, community, and society. Other societies emphasize individual opportunity. Plato, a pioneer of the individualist tradition, wrote more than 2000 years ago in *The Republic* that “no two persons are born exactly alike; but each differs from the other in natural endowments, one being suited for one occupation and the other for another.” As heirs to Plato's individualism, people in Western societies have pondered how and why individuals differ in mental ability.

Western attempts to assess such differences began in earnest over a century ago. The English scientist Francis Galton (1822–1911) had a fascination with measuring human traits. When his cousin Charles Darwin proposed that nature selects successful traits through the survival of the fittest, Galton wondered if it might be possible to measure “natural ability” and to encourage those of high ability to mate with one another. At the 1884 London Exposition, more than 10,000 visitors received his assessment of their “intellectual strengths” based on such things as reaction time, sensory acuity, muscular power, and body proportions. But alas, on these measures, well-regarded adults and students did not outscore others. Nor did the measures correlate with one another.

Although Galton's quest for a simple intelligence measure failed, he gave us some statistical techniques that we still use (as well as the phrase “nature and nurture”). And his persistent belief in the inheritance of genius—reflected in his book, *Hereditary Genius*—illustrates an important lesson from both the history of intelligence research and the history of science: Although science itself strives for objectivity, individual scientists are affected by their own assumptions and attitudes.

Marnabuke Sr. John/Alamy







**Alfred Binet** “Some recent philosophers have given their moral approval to the deplorable verdict that an individual’s intelligence is a fixed quantity, one which cannot be augmented. We must protest and act against this brutal pessimism” (Binet, 1909, p. 141).

“The IQ test was invented to predict academic performance, nothing else. If we wanted something that would predict life success, we’d have to invent another test completely.” —SOCIAL PSYCHOLOGIST ROBERT ZAJONC (1984b)

**mental age** a measure of intelligence test performance devised by Binet; the chronological age that most typically corresponds to a given level of performance. Thus, a child who does as well as the average 8-year-old is said to have a mental age of 8.

**Stanford-Binet** the widely used American revision (by Terman at Stanford University) of Binet’s original intelligence test.

**intelligence quotient (IQ)** defined originally as the ratio of mental age (*ma*) to chronological age (*ca*) multiplied by 100 (thus,  $IQ = ma/ca \times 100$ ). On contemporary intelligence tests, the average performance for a given age is assigned a score of 100, with scores assigned to relative performance above or below average.

## Alfred Binet: Predicting School Achievement

The modern intelligence-testing movement began at the turn of the twentieth century, when France passed a law requiring that all children attend school. Some children, including many newcomers to Paris, seemed incapable of benefiting from the regular school curriculum and in need of special classes. But how could the schools objectively identify children with special needs?

The French government hesitated to trust teachers’ subjective judgments of children’s learning potential. Academic slowness might merely reflect inadequate prior education. Also, teachers might prejudge children on the basis of their social backgrounds. To minimize bias, France’s minister of public education in 1904 commissioned Alfred Binet (1857–1911) and others to study the problem.

Binet and his collaborator, Théodore Simon, began by assuming that all children follow the same course of intellectual development but that some develop more rapidly. On tests, therefore, a “dull” child should perform as does a typical younger child, and a “bright” child as does a typical older child. Thus, their goal became measuring each child’s **mental age**, the level of performance typically associated with a certain chronological age. The average 9-year-old, then, has a mental age of 9. Children with below-average mental ages, such as 9-year-olds who perform at the level of typical 7-year-olds, would struggle with age-appropriate schoolwork.

To measure mental age, Binet and Simon theorized that mental aptitude, like athletic aptitude, is a general capacity that shows up in various ways. After testing a variety of reasoning and problem-solving questions on Binet’s two daughters, and then on “bright” and “backward” Parisian schoolchildren, Binet and Simon identified items that would predict how well French children would handle their schoolwork.

Note that Binet and Simon made no assumptions concerning *why* a particular child was slow, average, or precocious. Binet personally leaned toward an environmental explanation. To raise the capacities of low-scoring children, he recommended “mental orthopedics” that would help develop their attention span and self-discipline. He believed his intelligence test did not measure inborn intelligence as a meter stick measures height. Rather, it had a single practical purpose: to identify French schoolchildren needing special attention. Binet hoped his test would be used to improve children’s education, but he also feared it would be used to label children and limit their opportunities (Gould, 1981).

## Lewis Terman: The Innate IQ

Binet’s fears were realized soon after his death in 1911, when others adapted his tests for use as a numerical measure of inherited intelligence. This began when Stanford University professor Lewis Terman (1877–1956) found that the Paris-developed questions and age norms worked poorly with California schoolchildren. Adapting some of Binet’s original items, adding others, and establishing new age norms, Terman extended the upper end of the test’s range from teenagers to “superior adults.” He also gave his revision the name it retains today—the **Stanford-Binet**. For Terman, intelligence tests revealed the intelligence with which a person was born.

From such tests, German psychologist William Stern derived the famous **intelligence quotient**, or **IQ**. The IQ is simply a person’s mental age divided by chronological age and multiplied by 100 to get rid of the decimal point:

$$IQ = \frac{\text{mental age}}{\text{chronological age}} \times 100$$

Thus, an average child, whose mental and chronological ages are the same, has an IQ of 100. But an 8-year-old who answers questions as would a typical 10-year-old has an IQ of 125.

The original IQ formula worked fairly well for children but not for adults. (Should a 40-year-old who does as well on the test as an average 20-year-old be assigned an IQ of only 50?) Most current intelligence tests, including the Stanford-Binet, no longer compute an IQ in this manner (though the term *IQ* still lingers as a shorthand expression for “intelligence test score”). Instead, they represent the test-taker’s performance *relative to the average performance of others the same age*. This average performance is arbitrarily assigned a score of 100, and about two-thirds of all test-takers fall between 85 and 115.

Terman promoted the widespread use of intelligence testing. His motive was to “take account of the inequalities of children in original endowment” by assessing their “vocational fitness.” In sympathy with Francis Galton’s *eugenics*—a much-criticized nineteenth-century movement that proposed measuring human traits and using the results to encourage only smart and fit people to reproduce—Terman (1916, pp. 91–92) envisioned that the use of intelligence tests would “ultimately result in curtailing the reproduction of feeble-mindedness and in the elimination of an enormous amount of crime, pauperism, and industrial inefficiency” (p. 7).

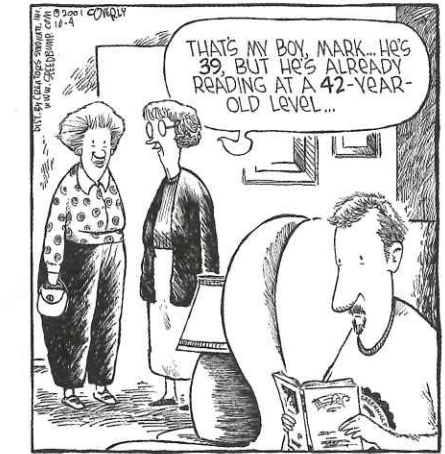
With Terman’s help, the U.S. government developed new tests to evaluate both newly arriving immigrants and World War I army recruits—the world’s first mass administration of an intelligence test. To some psychologists, the results indicated the inferiority of people not sharing their Anglo-Saxon heritage. Such findings were part of the cultural climate that led to a 1924 immigration law that reduced Southern and Eastern European immigration quotas to less than one-fifth of those for Northern and Western Europe.

Binet probably would have been horrified that his test had been adapted and used to draw such conclusions. Indeed, such sweeping judgments became an embarrassment to most of those who championed testing. Even Terman came to appreciate that test scores reflected not only people’s innate mental abilities but also their education, native language, and familiarity with the culture assumed by the test. Abuses of the early intelligence tests serve to remind us that science can be value-laden. Behind a screen of scientific objectivity, ideology sometimes lurks.

## Modern Tests of Mental Abilities

### 61-2 What’s the difference between achievement and aptitude tests?

By this point in your life, you’ve faced dozens of ability tests: school tests of basic reading and math skills, course exams, intelligence tests, and driver’s license exams, to name just a few. Psychologists classify such tests as either **achievement tests**, intended to *measure* what you have learned, or **aptitude tests**, intended to *predict* your ability to learn a new skill. Exams covering what you have learned in this course (like the AP® Exam) are achievement tests. A college entrance exam, which seeks to predict your ability to do college work, is an aptitude test—a “thinly disguised intelligence test,” says Howard Gardner (1999a). Indeed, total scores on the U.S. SAT® correlated +.82 with general intelligence scores in a national sample of 14- to 21-year-olds (Frey & Detterman, 2004; **FIGURE 61.1** on the next page).



Mrs. Randolph takes mother’s pride too far.

### AP® Exam Tip

David Myers did not use the terms *nature* or *nurture* as he described the contributions of Binet and Terman, but he dropped many hints that should allow you to figure out who leaned toward the nature position and who leaned toward the nurture position. Can you do it?

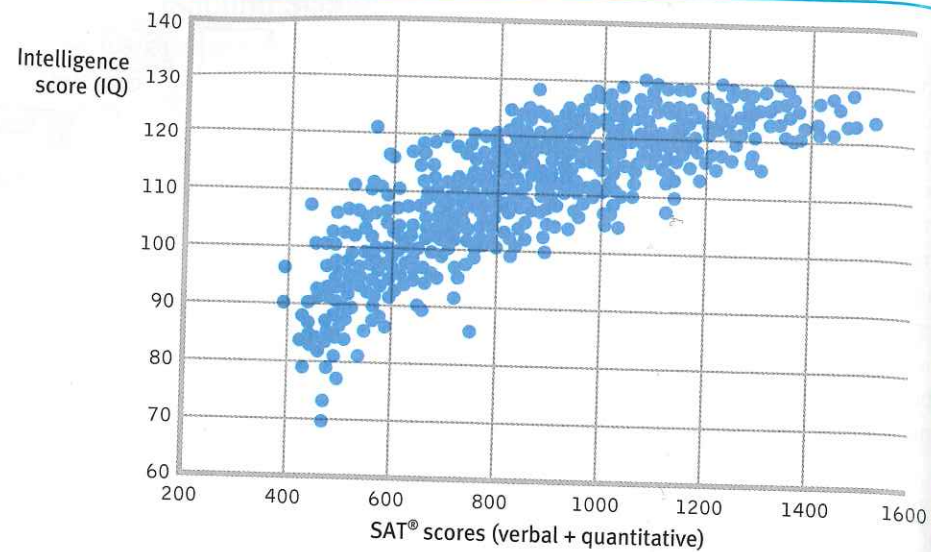
**achievement test** a test designed to assess what a person has learned.

**aptitude test** a test designed to predict a person’s future performance; aptitude is the capacity to learn.



Figure 61.1

**Close cousins: Aptitude and intelligence scores** A scatterplot shows the close correlation between intelligence scores and verbal and quantitative SAT® scores. (From Frey and Detterman, 2004.)



**Wechsler Adult Intelligence Scale (WAIS)** the WAIS is the most widely used intelligence test; contains verbal and performance (nonverbal) subtests.



**Matching patterns** Block design puzzles test visual abstract processing ability. Wechsler's individually administered intelligence test comes in forms suited for adults and children.

Psychologist David Wechsler created what is now the most widely used individual intelligence test, the **Wechsler Adult Intelligence Scale (WAIS)**, with a version for school-age children (the *Wechsler Intelligence Scale for Children [WISC]*), and another for preschool children. The latest (2008) edition of the WAIS consists of 15 subtests, including these:

- **Similarities**—Reasoning the commonality of two objects or concepts, such as “In what way are wool and cotton alike?”
- **Vocabulary**—Naming pictured objects, or defining words (“What is a guitar?”)
- **Block design**—Visual abstract processing, such as “Using the four blocks, make one just like this.”
- **Letter-number sequencing**—On hearing a series of numbers and letters, repeat the numbers in ascending order, and then the letters in alphabetical order: “R-2-C-1-M-3.”



It yields not only an overall intelligence score, as does the Stanford-Binet, but also separate scores for verbal comprehension, perceptual organization, working memory, and processing speed. Striking differences among these scores can provide clues to cognitive strengths or weaknesses that teachers or therapists can build upon. For example, a low verbal comprehension score combined with high scores on other subtests could indicate a reading or language disability. Other comparisons can help a psychologist or psychiatrist establish a rehabilitation plan for a stroke patient. Such uses are possible, of course, only when we can trust the test results.

## Principles of Test Construction

### 61-3 What are standardization and the normal curve?

To be widely accepted, psychological tests must meet three criteria: They must be *standardized*, *reliable*, and *valid*. The Stanford-Binet and Wechsler tests meet these requirements.

### Standardization

The number of questions you answer correctly on an intelligence test would tell us almost nothing. To evaluate your performance, we need a basis for comparing it with others' performance. To enable meaningful comparisons, test-makers first give the test to a representative sample of people. When you later take the test following the same procedures, your score

Number of scores

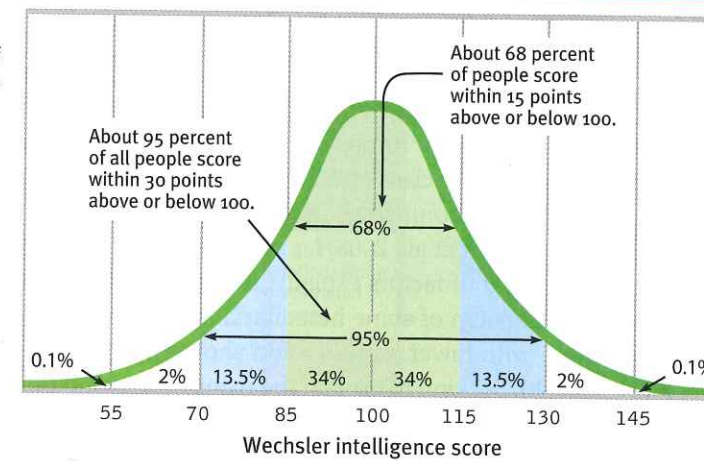


Figure 61.2

**The normal curve** Scores on aptitude tests tend to form a normal, or bell-shaped, curve around an average score. For the Wechsler scale, for example, the average score is 100.

can be compared with the sample's scores to determine your position relative to others. This process of defining meaningful scores relative to a pretested group is called **standardization**.

Group members' scores typically are distributed in a bell-shaped pattern that forms the **normal curve** shown in **FIGURE 61.2**. No matter what we measure—height, weight, or mental aptitude—people's scores tend to form this roughly symmetrical shape. On an intelligence test, we call the midpoint, the average score, 100. Moving out from the average toward either extreme, we find fewer and fewer people. For both the Stanford-Binet and Wechsler tests, a person's score indicates whether that person's performance fell above or below the average. As Figure 61.2 shows, a performance higher than all but 2 percent of all scores earns an intelligence score of 130. A performance lower than 98 percent of all scores earns an intelligence score of 70.

To keep the average score near 100, the Stanford-Binet and Wechsler scales are periodically restandardized. If you took the WAIS Fourth Edition recently, your performance was compared with a standardization sample who took the test during 2007, not to David Wechsler's initial 1930s sample. If you compared the performance of the most recent standardization sample with that of the 1930s sample, do you suppose you would find rising or declining test performance? Amazingly—given that college entrance aptitude scores were dropping during the 1960s and 1970s—intelligence test performance was improving. This worldwide phenomenon is called the *Flynn effect*, in honor of New Zealand researcher James Flynn (1987, 2009b, 2010), who first calculated its magnitude. As **FIGURE 61.3** indicates, the average person's intelligence test score in 1920 was—by today's standard—only a 76! Such rising performance has been observed in 29 countries, from Canada to rural Australia (Ceci & Kanaya, 2010). Although the gains have recently reversed in Scandinavia, the historic increase is now widely accepted as an important phenomenon (Lynn, 2009; Teasdale & Owen, 2005, 2008).

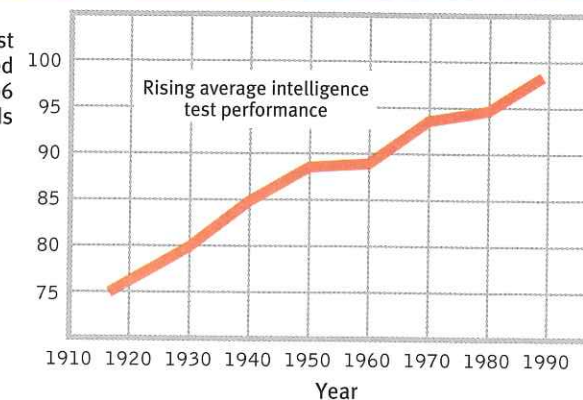
**standardization** defining uniform testing procedures and meaningful scores by comparison with the performance of a pretested group.

**normal curve** the symmetrical, bell-shaped curve that describes the distribution of many physical and psychological attributes. Most scores fall near the average, and fewer and fewer scores lie near the extremes.

### AP® Exam Tip

Can you remember why the intelligence test scores in Figure 61.2 are marked off in 15-point intervals? Do the 68 percent and 95 percent areas seem familiar? They should—you've seen this graph before. It's Figure 7.3 from the module on statistical reasoning. Intelligence tests are being used to illustrate that 68 percent of a population will be within one standard deviation of the mean for normally distributed data. Ninety-five percent will be within two standard deviations.

Intelligence test scores, based on 1996 standards



Lewis J. Merriam/Science Source

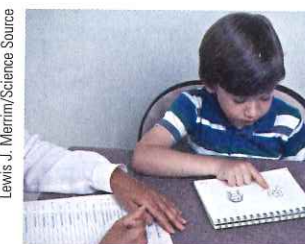


Figure 61.3

**Getting smarter?** In every country studied, intelligence test performance rose during the twentieth century, as shown here with U.S. Wechsler and Stanford-Binet test performance between 1918 and 1989. In Britain, test scores have risen 27 points since 1942. (From Hogan, 1995.) Very recent data indicate this trend may have leveled off or may even be reversing.



The Flynn effect's cause has been a mystery. Did it result from greater test sophistication? (But the gains began before testing was widespread and have even been observed among preschoolers.) Better nutrition? As the nutrition explanation would predict, people have gotten not only smarter but taller. But in post-war Britain, notes Flynn (2009a), the lower-class children gained the most from improved nutrition but the intelligence performance gains were greater among upper-class children. Or did the Flynn effect stem from more education? More stimulating environments? Less childhood disease? Smaller families and more parental investment (Sundet et al., 2008)?

Regardless of what combination of factors explains the rise in intelligence test scores, the phenomenon counters one concern of some hereditarians—that the higher twentieth-century birthrates among those with lower scores would shove human intelligence scores downward (Lynn & Harvey, 2008). Seeking to explain the rising scores, and mindful of global mixing, one scholar has even speculated about the influence of a genetic phenomenon comparable with “hybrid vigor,” which occurs in agriculture when cross-breeding produces corn or livestock superior to the parent plants or animals (Mingroni, 2004, 2007).

**reliability** the extent to which a test yields consistent results, as assessed by the consistency of scores on two halves of the test, on alternate forms of the test, or on retesting.

**validity** the extent to which a test measures or predicts what it is supposed to. (See also *content validity* and *predictive validity*.)

**content validity** the extent to which a test samples the behavior that is of interest.

**predictive validity** the success with which a test predicts the behavior it is designed to predict; it is assessed by computing the correlation between test scores and the criterion behavior. (Also called *criterion-related validity*.)

## Reliability

### 61-4 What are reliability and validity?

Knowing where you stand in comparison to a standardization group still won't tell us much about your intelligence unless the test has **reliability**—unless it yields dependably consistent scores. To check a test's reliability, researchers retest people. They may use the same test or they may split the test in half to see whether odd-question scores and even-question scores agree. If the two scores generally agree, or *correlate*, the test is reliable. The higher the correlation between the *test-retest* or the *split-half* scores, the higher the test's reliability. The tests we have considered so far—the Stanford-Binet, the WAIS, and the WISC—all have reliabilities of about +.9, which is very high. When retested, people's scores generally match their first score closely.

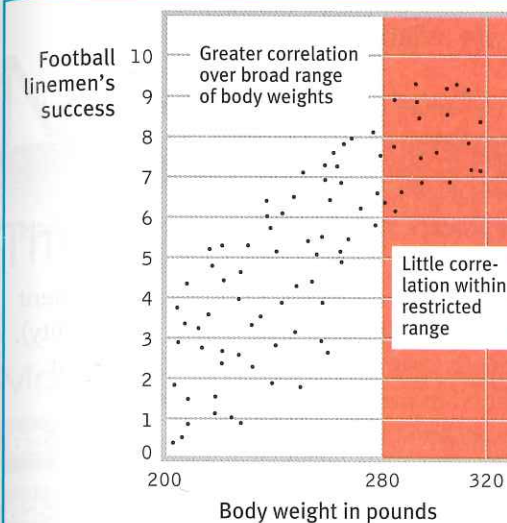
## Validity

High reliability does not ensure a test's **validity**—the extent to which the test actually measures or predicts what it promises. If you use an inaccurate tape measure to measure people's heights, your height report would have high reliability (consistency) but low validity. It is enough for some tests that they have **content validity**, meaning the test taps the pertinent behavior, or *criterion*. The road test for a driver's license has content validity because it samples the tasks a driver routinely faces. Course exams have content validity if they assess one's mastery of a representative sample of course material. But we expect intelligence tests to have **predictive validity**: They should predict the criterion of future performance, and to some extent they do.

Are general aptitude tests as predictive as they are reliable? As critics are fond of noting, the answer is plainly *No*. The predictive power of aptitude tests is fairly strong in the early school years, but later it weakens. Academic aptitude test scores are reasonably good predictors of achievement for children ages 6 to 12, where the correlation between intelligence score and school performance is about +.6 (Jensen, 1980). Intelligence scores correlate even more closely with scores on *achievement tests*: +.81 in one comparison of 70,000 English children's intelligence scores at age 11 with their academic achievement in national exams at age 16 (Deary et al., 2007, 2009). The SAT® exam, used in the United States as a college entrance exam, is less successful in predicting first-year college grades. (The correlation, which is less than +.5, is, however, a bit higher when adjusting for high scorers electing tougher courses [Berry & Sackett, 2009; Willingham et al., 1990].) By the time we get to the Graduate Record Examination® (GRE®; an aptitude test similar to the SAT® exam but for those applying to graduate school), the correlation with graduate school performance is an even more modest but still significant +.4 (Kuncel & Hezlett, 2007).

### AP® Exam Tip

Be careful! The terms *reliability* and *validity* have more precise meanings to psychologists than they do to the general public.



**Figure 61.4**  
**Diminishing predictive power** Let's imagine a correlation between football linemen's body weight and their success on the field. Note how insignificant the relationship becomes when we narrow the range of weight to 280 to 320 pounds. As the range of data under consideration narrows, its predictive power diminishes.

Why does the predictive power of aptitude scores diminish as students move up the educational ladder? Consider a parallel situation: Among all American and Canadian football linemen, body weight correlates with success. A 300-pound player tends to overwhelm a 200-pound opponent. But within the narrow 280- to 320-pound range typically found at the professional level, the correlation between weight and success becomes negligible (**FIGURE 61.4**). The narrower the *range* of weights, the lower the predictive power of body weight becomes. If an elite university takes only those students who have very high aptitude scores, those scores cannot possibly predict much. This will be true even if the test has excellent predictive validity with a more diverse sample of students. So, when we validate a test using a wide range of people but then use it with a restricted range of people, it loses much of its predictive validity.

## Before You Move On

### ▶ ASK YOURSELF

Are you working to the potential reflected in your standardized test scores? What, other than your aptitude, is affecting your school performance?

### ▶ TEST YOURSELF

What was the purpose of Binet's pioneering intelligence test?

Answers to the Test Yourself questions can be found in Appendix E at the end of the book.

## Module 61 Review

### 61-1 When and why were intelligence tests created?

- In the late 1800s, Francis Galton, who believed that genius was inherited, attempted but failed to construct a simple intelligence test.
- In France in 1904, Alfred Binet, who tended toward an environmental explanation of intelligence differences, started the modern intelligence-testing movement by developing questions to measure children's *mental age* and thus predict progress in the school system.

- During the early twentieth century, Lewis Terman of Stanford University revised Binet's work for use in the United States.
- Terman believed intelligence is inherited, and he thought his *Stanford-Binet* could help guide people toward appropriate opportunities.
- During this period, intelligence tests were sometimes used to “document” scientists' assumptions about the innate inferiority of certain ethnic and immigrant groups.



**61-2** What's the difference between achievement and aptitude tests?

- *Achievement tests* are designed to assess what you have learned.
- *Aptitude tests* are designed to predict what you can learn.
- The *WAIS (Wechsler Adult Intelligence Scale)*, an aptitude test, is the most widely used intelligence test for adults.

**61-3** What are standardization and the normal curve?

- *Standardization* establishes a basis for meaningful score comparisons by giving a test to a representative sample of future test-takers.
- The distribution of test scores often forms a *normal* (bell-shaped) *curve* around the central average score, with fewer and fewer scores at the extremes.

**Multiple-Choice Questions**

1. A test-developer defines uniform testing procedures and meaningful scores by comparison with the performance of a pretested group. Which of the following best describes this process?
  - a. Reliability testing
  - b. Validation
  - c. Content validation
  - d. Standardization
  - e. Predictive validity
2. Which of the following best describes the extent to which a test yields consistent results upon retesting?
  - a. Content validity
  - b. Validity
  - c. Reliability
  - d. Predictive validity
  - e. Normal curve
3. Which of the following can be used to demonstrate that only about 2 percent of the population scores are at least two standard deviations above the mean on an intelligence test?
  - a. Reliability test
  - b. Aptitude test
  - c. Predictive validity test
  - d. Test-retest procedure
  - e. Normal curve

**Practice FRQs**

1. What are the fundamental differences between achievement and aptitude tests?
2. Name and briefly describe the three essential principals of test construction.

**(3 points)****Answer**

**1 point:** Achievement tests are designed to assess what a person has learned.

**1 point:** An aptitude test is designed to predict a person's future performance.

**61-4** What are reliability and validity?

- *Reliability* is the extent to which a test yields consistent results (on two halves of the test, or when people are retested).
- *Validity* is the extent to which a test measures or predicts what it is supposed to.
  - A test has *content validity* if it samples the pertinent behavior (as a driving test measures driving ability).
  - It has *predictive validity* if it predicts a behavior it was designed to predict. (Aptitude tests have predictive ability if they can predict future achievements.)

# Module 62

## The Dynamics of Intelligence

### Module Learning Objectives

- 62-1** Describe the stability of intelligence scores over the life span.
- 62-2** Describe the traits of those at the low and high intelligence extremes.

We now can address some age-old questions about the dynamics of human intelligence—about its stability over the life span, and about the extremes of intelligence.

### Stability or Change?

- 62-1** How stable are intelligence scores over the life span?

If we retested people periodically throughout their lives, would their intelligence scores be stable? Let's first explore the stability of mental abilities in later life.

### Aging and Intelligence

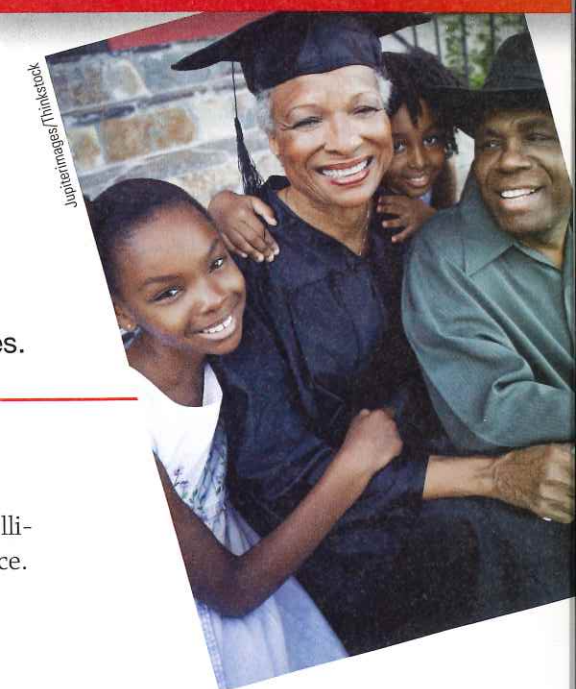
What happens to our broader intellectual muscles as we age? Do they gradually decline, as does our body strength (even if relative intellectual and muscular strength in later life is predictable from childhood)? Or do they remain constant? The quest for answers to these questions illustrates psychology's self-correcting process. This research developed in phases.

#### PHASE I: CROSS-SECTIONAL EVIDENCE FOR INTELLECTUAL DECLINE

In *cross-sectional studies*, researchers at one point in time test and compare people of various ages. In such studies, researchers have consistently found that older adults give fewer correct answers on intelligence tests than do younger adults. WAIS-creator, David Wechsler (1972) therefore concluded that "the decline of mental ability with age is part of the general [aging] process of the organism as a whole." For a long time, this rather dismal view went unchallenged. Many corporations established mandatory retirement policies, assuming the companies would benefit by replacing aging workers with younger, presumably more capable, employees. As "everyone knows," you can't teach an old dog new tricks.

#### PHASE II: LONGITUDINAL EVIDENCE FOR INTELLECTUAL STABILITY

After colleges in the 1920s began giving intelligence tests to entering students, several psychologists saw their chance to study intelligence *longitudinally*. They retested the same **cohort**—the same group of people—over a period of years (Schaie & Geiwitz, 1982). What they found was a surprise: Until late in life, intelligence remained stable (**FIGURE 62.1** on the next page). On some tests, it even increased.



**cohort** a group of people from a given time period.