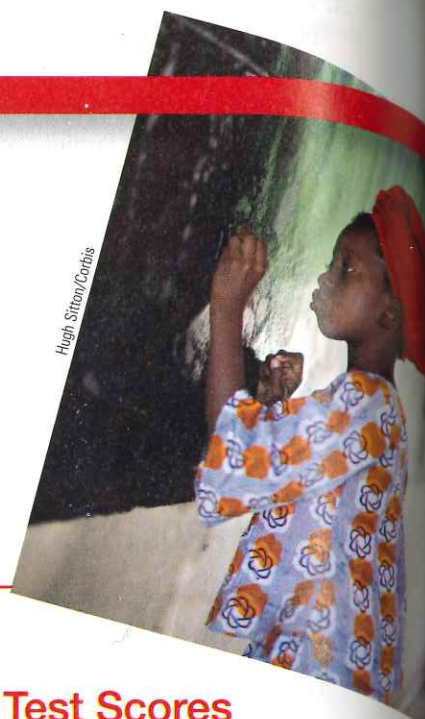


Module 64

Group Differences and the Question of Bias

Module Learning Objectives

- 64-1** Describe how and why the genders differ in mental ability scores.
- 64-2** Describe how and why racial and ethnic groups differ in mental ability scores.
- 64-3** Discuss whether intelligence tests are inappropriately biased.



Group Differences in Intelligence Test Scores

If there were no group differences in aptitude scores, psychologists could politely debate hereditary and environmental influences in their ivory towers. But there are group differences. What are they? And what shall we make of them?

Gender Similarities and Differences

- 64-1** How and why do the genders differ in mental ability scores?

In science, as in everyday life, differences, not similarities, excite interest. Compared with the anatomical and physiological similarities between men and women, our differences are minor. In that 1932 testing of all Scottish 11-year-olds, for example, girls' average intelligence score was 100.6 and boys' was 100.5 (Deary et al., 2003). So far as *g* is concerned, boys and girls, men and women, are the same species.

Yet, most people find differences more newsworthy. Girls are better spellers, more verbally fluent, better at locating objects, better at detecting emotions, and more sensitive to touch, taste, and color (Halpern et al., 2007). Boys outperform girls in tests of spatial ability and complex math problems, though in math computation and overall math performance, boys and girls hardly differ (Else-Quest et al., 2010; Hyde & Mertz, 2009; Lindberg et al., 2010). Males' mental ability scores also vary more than females'. Thus, boys worldwide outnumber girls at both the low extreme and the high extreme (Machin & Pekkarinen, 2008; Strand et al., 2006). Boys, for example, are more often found in special education classes. And among 12- to 14-year-olds scoring extremely high (700 or higher) on the SAT® exam math section, boys outnumber girls 4 to 1 (Wai et al., 2010).

The most reliable male edge appears in spatial ability tests like the one shown in **FIGURE 64.1**. The solution requires speedily rotating three-dimensional objects in one's mind (Collins & Kimura, 1997; Halpern, 2000). Today, such skills help when fitting suitcases into a car trunk, playing chess, or doing certain types of geometry problems. From an evolutionary perspective, those same skills would have helped our ancestral fathers track prey and make their way home (Geary, 1995, 1996; Halpern et al., 2007). The survival of our ancestral mothers may have benefited more from a keen memory for the location of edible plants—a legacy that lives today in women's superior memory for objects and their location.

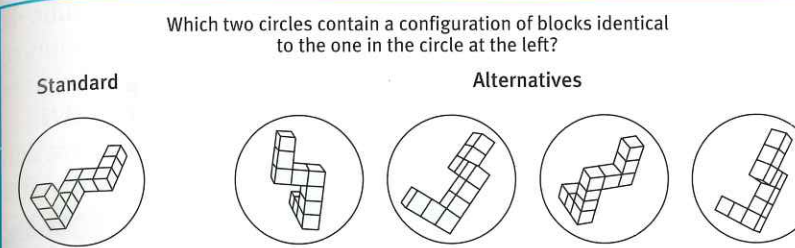


Figure 64.1

The mental rotation test This is a test of spatial abilities. (From Vandenberg & Kuse, 1978.) See inverted answer below.

ANSWER: The first and fourth alternatives.

But experience also matters. One experiment found that playing action video games boosts spatial abilities (Feng et al., 2007). And you probably won't be surprised to know that among entering American collegians, six times as many men (23 percent) as women (4 percent) report playing video/computer games six or more hours a week (Pryor et al., 2010).

Evolutionary psychologist Steven Pinker (2005) argues that biological as well as social influences appear to affect gender differences in life priorities (women's greater interest in people versus men's in money and things), in risk-taking (with men more reckless), and in math reasoning and spatial abilities. Such differences are, he notes, observed across cultures, stable over time, influenced by prenatal hormones, and observed in genetic boys raised as girls. Culturally influenced preferences also help explain women selecting people- rather than math-intensive vocations (Ceci & Williams, 2010, 2011).

Other critics urge us to remember that social expectations and divergent opportunities shape boys' and girls' interests and abilities (Crawford et al., 1995; Eccles et al., 1990). Gender-equal cultures, such as Sweden and Iceland, exhibit little of the gender math gap found in gender-unequal cultures, such as Turkey and Korea (Guiso et al., 2008).



Racial and Ethnic Similarities and Differences

- 64-2** How and why do racial and ethnic groups differ in mental ability scores?

Fueling the group-differences debate are two other disturbing but agreed-upon facts:

- Racial groups differ in their average intelligence test scores.
- High-scoring people (and groups) are more likely to attain high levels of education and income.

There are many group differences in average intelligence test scores. New Zealanders of European descent outscore native Maori New Zealanders. Israeli Jews outscore Israeli Arabs. Most Japanese outscore most Burakumin, a stigmatized Japanese minority. Those who can hear outscore those born deaf (Braden, 1994; Steele, 1990; Zeidner, 1990). And White Americans have outscored Black Americans. This Black-White difference has diminished somewhat in recent years, especially among children (Dickens & Flynn, 2006; Nisbett, 2009). Such *group* differences provide little basis for judging individuals. Worldwide, women outlive men by 4 years, but knowing only that you are male or female won't tell us much about how long you will live.

We have seen that heredity contributes to *individual* differences in intelligence. But group differences in a heritable trait may be entirely environmental. Consider one of nature's experiments: Allow some children to grow up hearing their culture's dominant language, while others, born deaf, do not. Then give both groups an intelligence test rooted in

Nature or nurture? At this 2005 Google Inc.-sponsored computer coding competition, programmers competed for cash prizes and possible jobs. What do you think accounted for the fact that only one of the 100 finalists was female?



© Larry Williams/CORBIS

the dominant language, and (no surprise) those with expertise in that language will score highest. Although individual performance differences may be substantially genetic, the group difference is not (FIGURE 64.2).

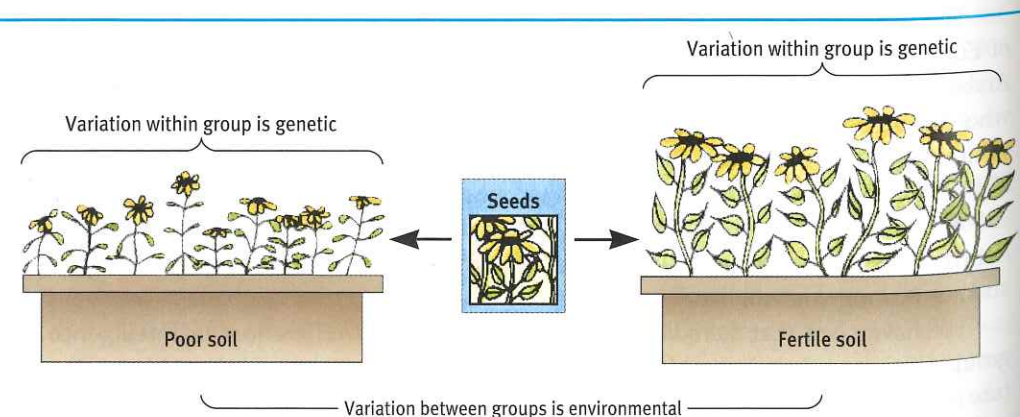
Might the racial gap be similarly environmental? Consider:

Genetics research reveals that under the skin, the races are remarkably alike. The average genetic difference between two Icelandic villagers or between two Kenyans greatly exceeds the group difference between Icelanders and Kenyans (Cavalli-Sforza et al., 1994; Rosenberg et al., 2002). Moreover, looks can deceive. Light-skinned Europeans and dark-skinned Africans are genetically closer than are dark-skinned Africans and dark-skinned Aboriginal Australians.

Race is not a neatly defined biological category. Some scholars argue that there is a reality to race, noting that there are genetic markers for race (the continent of one's ancestry), that medical risks (such as skin cancer or high blood pressure) vary by race, and that most people self-identify with a given race (Hunt & Carlson, 2007). Behavioral traits may also vary by race. "No runner of Asian or European descent—a majority of the world's population—has broken 10 seconds in the 100-meter dash, but dozens of runners of West African descent have done so," observed psychologist David Rowe (2005). Many social scientists, though, see race primarily as a social construction without well-defined physical boundaries, as each race blends seamlessly into the race of its geographical neighbors (Helms et al., 2005; Smedley & Smedley, 2005). People with varying ancestry may categorize themselves in the same race. Moreover, with increasingly mixed ancestries, more and more people defy neat racial categorization and self-identify as multiracial (Pauker et al., 2009).

The intelligence test performance of today's better-fed, better-educated, and more test-prepared population exceeds that of the 1930s population—by a greater margin than the intelligence test score of the average White today exceeds that of the average Black. One research review noted that the average IQ test performance of today's sub-Saharan Africans is the same as Brit-

Figure 64.2
Group differences and environmental impact Even if the variation between members within a group reflects genetic differences, the average difference between groups may be wholly due to the environment. Imagine that seeds from the same mixture are sown in different soils. Although height differences *within* each window box will be genetic, the height difference *between* the two groups will be environmental. (From Lewontin, 1976.)



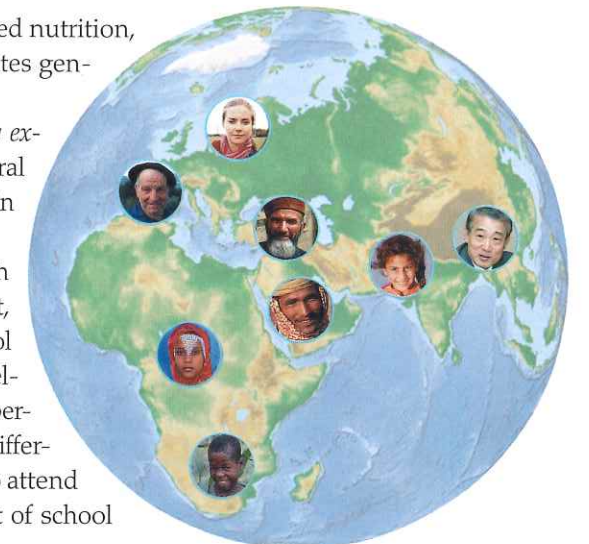
ish adults in 1948, with the possibility of similar gains to come, given improved nutrition, economic development, and education (Wicherts et al., 2010). No one attributes generational group differences to genetics.

When Blacks and Whites have or receive the same pertinent knowledge, they exhibit similar information-processing skill. "The data support the view that cultural differences in the provision of information may account for racial differences in IQ," report researchers Joseph Fagan and Cynthia Holland (2007).

Schools and culture matter. Countries whose economies create a large wealth gap between rich and poor tend also to have a large rich/poor IQ gap (Nisbett, 2009). Moreover, educational policies such as kindergarten attendance, school discipline, and instructional time per year predict national differences in intelligence and knowledge tests (Rindermann & Ceci, 2009). Asian students outperform North American students on math achievement and aptitude tests. This difference may reflect conscientiousness more than competence. Asian students also attend school 30 percent more days per year and spend much more time in and out of school studying math (Geary et al., 1996; Larson & Verma, 1999; Stevenson, 1992).

In different eras, different ethnic groups have experienced golden ages—periods of remarkable achievement. Twenty-five-hundred years ago, it was the Greeks and the Egyptians, then the Romans; in the eighth and ninth centuries, genius seemed to reside in the Arab world; 500 years ago it was the Aztec Indians and the peoples of Northern Europe. Today, people marvel at Asians' technological genius and Jews' cultural success. In today's United States, Jews are 2 percent of the population, 21 percent of Ivy League student bodies, 37 percent of Academy Award-winning directors, and 51 percent of Pulitzer Prize winners for nonfiction; worldwide, they have been 27 percent of Nobel physics laureates and 54 percent of world chess champions (Brooks, 2010). Cultures rise and fall over centuries; genes do not. That fact makes it difficult to attribute a natural superiority to any race.

Moreover, consider the striking results of a national study that looked back over the mental test performances of White and Black young adults after graduation from college. From eighth grade through the early high school years, the average aptitude score of the White students increased, while that of the Black students decreased—creating a gap that reached its widest point at about the time that high school students like you take college admissions tests. But during college, the Black students' scores increased "more than four times as much" as those of their White counterparts, thus greatly decreasing the aptitude gap. "It is not surprising," concluded researcher Joel Myerson and his colleagues (1998), "that as Black and White students complete more grades in high school environments that differ in quality, the gap in cognitive test scores widens. At the college level, however, where Black and White students are exposed to educational environments of comparable quality . . . many Blacks are able to make remarkable gains, closing the gap in test scores."



Nature's own morphing Nature draws no sharp boundaries between races, which blend gradually one into the next around the Earth. Thanks to the human urge to classify, however, people socially define themselves in racial categories, which become catchall labels for physical features, social identity, and nationality.

© Rob Howard/Corbis;
© David Turnley/Corbis;
© Barbara Bannister, Gallo Images/Corbis;
© Dave Bartruff/Corbis;
© Haruyoshi Yamaguchi/Corbis;
© Richard T. Nowitz/Corbis;
© Owen Franken/Corbis;
© John-Francis Bourke/zefa/Corbis

"Do not obtain your slaves from Britain, because they are so stupid and so utterly incapable of being taught." -CICERO, 106-43 B.C.E.

The Question of Bias

64-3 Are intelligence tests inappropriately biased?

If one assumes that race is a meaningful concept, the debate over race differences in intelligence divides into three camps, note Earl Hunt and Jerry Carlson (2007):

- There are genetically disposed race differences in intelligence.
- There are socially influenced race differences in intelligence.
- There are race differences in test scores, but the tests are inappropriate or biased.

Are intelligence tests biased? The answer depends on which of two very different definitions of *bias* we use.

Two Meanings of Bias

We consider a test biased if it detects not only innate differences in intelligence but also performance differences caused by cultural experiences. This in fact happened to Eastern European immigrants in the early 1900s. Lacking the experience to answer questions about their new culture, many were classified as feeble-minded.

In this popular sense, intelligence tests are biased. They measure your developed abilities, which reflect, in part, your education and experiences. You may have read examples of intelligence test items that make middle-class assumptions (for example, that a cup goes with a saucer). Do such items bias the test against those who do not use saucers? Could such questions explain racial differences in test performance? If so, are tests a vehicle for discrimination, consigning potentially capable children, some of whom may have a different native language, to dead-end classes and jobs? And could creating culture-neutral questions—such as by assessing people's ability to learn novel words, sayings, and analogies—enable culture-fair aptitude tests (Fagan & Holland, 2007, 2009)?

Defenders of the existing aptitude tests note that racial group differences persist on nonverbal items, such as counting digits backward (Jensen, 1983, 1998). Moreover, they add, blaming the test for a group's lower scores is like blaming a messenger for bad news. Why blame the tests for exposing unequal experiences and opportunities? If, because of malnutrition, people were to suffer stunted growth, would you blame the measuring stick that reveals it? If unequal past experiences predict unequal future achievements, a valid aptitude test will detect such inequalities.

The second meaning of *bias*—its *scientific* meaning—is different. It hinges on a test's validity—on whether it predicts future behavior only for some groups of test-takers. For example, if the SAT® exam accurately predicted the college achievement of women but not that of men, then the test would be biased. In this statistical meaning of the term, the near-consensus among psychologists (as summarized by the U.S. National Research Council's Committee on Ability Testing and the American Psychological Association's Task Force on Intelligence) is that the major U.S. aptitude tests are *not* biased (Hunt & Carlson, 2007; Neisser et al., 1996; Wigdor & Garner, 1982). The tests' predictive validity is roughly the same for women and men, for Blacks and Whites, and for rich and poor. If an intelligence test score of 95 predicts slightly below-average grades, that rough prediction usually applies equally to all.

Test-Takers' Expectations

Throughout this text, we have seen that our expectations and attitudes can influence our perceptions and behaviors, and we find this effect in intelligence testing. When Steven Spencer and his colleagues (1997) gave a difficult math test to equally capable men and women, women did not do as well—except when they had been led to expect that women usually do as well as men on the test. Otherwise, the women apparently felt apprehensive, which affected their performance. With Claude Steele and Joshua Aronson, Spencer (2002) also observed this self-fulfilling **stereotype threat** with Black students. When reminded of their race just before taking verbal aptitude tests, they performed worse. Follow-up experiments confirm that negatively stereotyped minorities and women may have unrealized academic potential (Nguyen & Ryan, 2008; Walton & Spencer, 2009). If, when taking an exam, you are worried that your type often doesn't do well, your self-doubts and self-monitoring may hijack your working memory and impair your performance (Schmader, 2010). For such reasons, stereotype threat may also impair attention and learning (Inzlicht & Kang, 2010; Rydell et al., 2010).

Critics note that stereotype threat does not fully account for the Black-White aptitude score difference (Sackett et al., 2004, 2008). But it does help explain why Blacks have scored higher when tested by Blacks than when tested by Whites (Danso & Esses, 2001; Inzlicht & Ben-Zeev, 2000). It gives us insight into why women have scored higher on math tests with no male test-takers present, and why women's chess play drops sharply when they *think* they are playing a male opponent (Maass et al., 2008). And it explains “the Obama effect”—

“Political equality is a commitment to universal human rights, and to policies that treat people as individuals rather than representatives of groups; it is not an empirical claim that all groups are indistinguishable.” -STEVEN PINKER (2006)

stereotype threat a self-confirming concern that one will be evaluated based on a negative stereotype.

“Math class is tough!” -“TEEN TALK” TALKING BARBIE DOLL (INTRODUCED JULY 1992, RECALLED OCTOBER 1992)

the finding that African-American adults performed better if taking a verbal aptitude test administered immediately after watching Barack Obama's stereotype-defying nomination acceptance speech or just after his 2008 presidential victory (Marx et al., 2009).

Steele (1995, 2010) concludes that telling students they probably won't succeed (as is sometimes implied by remedial “minority support” programs) functions as a stereotype that can erode performance. Over time, such students may detach their self-esteem from academics and look for recognition elsewhere. Indeed, as African-American boys progress from eighth to twelfth grade, there is a growing disconnect between their grades and their self-esteem and they tend to underachieve (Osborne, 1997). One experiment randomly assigned some African-American seventh-graders to write for 15 minutes about their most important values (Cohen et al., 2006, 2009). That simple exercise in self-affirmation had the apparent effect of boosting their semester grade point average by 0.26 in a first experiment and 0.34 in a replication. Minority students in university programs that challenge them to believe in their potential, or to focus on the idea that intelligence is malleable and not fixed, have likewise produced markedly higher grades and had lower dropout rates (Wilson, 2006).

What, then, can we realistically conclude about aptitude tests and bias? The tests are indeed biased (appropriately so, some would say) in one sense—sensitivity to performance differences caused by cultural experience. But they are not biased in the scientific sense of failing to make valid statistical predictions for different groups.

Bottom line: Are the tests discriminatory? Again, the answer can be *Yes* or *No*. In one sense, *Yes*, their purpose is to discriminate—to distinguish among individuals. In another sense, *No*, their purpose is to reduce discrimination by reducing reliance on subjective criteria for school and job placement—who you know, what school you're from, or whether you are the “right kind of person.” Civil service aptitude tests, for example, were devised to discriminate more fairly and objectively by reducing the political, racial, and ethnic discrimination that preceded their use. Banning aptitude tests would lead those who decide on jobs and admissions to rely more on other considerations, such as personal opinion.

Perhaps, then, our goals for tests of mental abilities should be threefold. First, we should realize the benefits Alfred Binet foresaw—to enable schools to recognize who might profit most from early intervention. Second, we must remain alert to Binet's fear that intelligence test scores may be misinterpreted as literal measures of a person's worth and potential. Third, we must remember that the competence that general intelligence tests sample is important; it helps enable success in some life paths. But it reflects only one aspect of personal competence. Our practical intelligence and emotional intelligence matter, too, as do other forms of creativity, talent, and character. Because there are many ways of being successful, our differences are variations of human adaptability.

Finally, life's great achievements result not only from “can do” abilities but also from “will do” motivation. Competence + Diligence → Accomplishment.

Before You Move On

▶ ASK YOURSELF

How have your expectations influenced your own test performance? What steps could you take to control this influence?

▶ TEST YOURSELF

What is the difference between a test that is biased culturally, and a test that is biased in terms of its validity?

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



BananaStock/Jupiter images

Unstable compassion

Intelligence test scores are only one part of the picture of a whole person. They don't measure the abilities, talent, and commitment of, for example, people who devote their lives to helping others.

“Almost all the joyful things of life are outside the measure of IQ tests.” -MADELEINE L'ENGLE, *A CIRCLE OF QUIET*, 1972

“[Einstein] showed that genius equals brains plus tenacity squared.” -WALTER ISAACSON, “EINSTEIN'S FINAL QUEST,” 2009

Module 64 Review

64-1 How and why do the genders differ in mental ability scores?

- Males and females tend to have the same average intelligence test scores. They differ in some specific abilities.
- Girls are better spellers, more verbally fluent, better at locating objects, better at detecting emotions, and more sensitive to touch, taste, and color.
- Boys outperform girls at spatial ability and related mathematics, though girls outperform boys in math computation. Boys also outnumber girls at the low and high extremes of mental abilities.
- Psychologists debate evolutionary, brain-based, and cultural explanations of such gender differences.

64-2 How and why do racial and ethnic groups differ in mental ability scores?

- Racial and ethnic groups differ in their average intelligence test scores.
- The evidence suggests that environmental differences are largely, perhaps entirely, responsible for these group differences.

Multiple-Choice Questions

1. Which of the following is true of boys compared with girls?
 - a. Boys have a higher average intelligence score.
 - b. Boys are better spellers than girls.
 - c. Boys are better at detecting emotions.
 - d. Boys are more verbally fluent.
 - e. Boys are more likely to have extremely low intelligence scores.
2. Which of the following provides the best evidence that race is more of a social construct than a biological category?
 - a. People of varying ancestry may categorize themselves in the same race.
 - b. The races arose in different continents.
 - c. Behavior traits (like running speed) are associated with race.
 - d. Skin cancer rates vary by race.
 - e. The incidence of high blood pressure varies by race.

64-3 Are intelligence tests inappropriately biased?

- Aptitude tests aim to predict how well a test-taker will perform in a given situation. So they are necessarily “biased” in the sense that they are sensitive to performance differences caused by cultural experience.
- By “inappropriately biased,” psychologists mean that a test predicts less accurately for one group than for another. In this sense, most experts consider the major aptitude tests unbiased.
- *Stereotype threat*, a self-confirming concern that one will be evaluated based on a negative stereotype, affects performance on all kinds of tests.

3. According to most experts, intelligence tests are not biased because
 - a. the average scores for various racial and ethnic groups do not differ by much.
 - b. the tests do a pretty good job of predicting what they are supposed to predict.
 - c. cultural background has little influence on test scores.
 - d. scores on the test are not very stable even when you don’t consider race.
 - e. scores are increasing for almost all groups because of the Flynn effect.

Practice FRQs

1. Robert and Maya are having an argument about whether intelligence tests are biased. Robert thinks they are but Maya insists they are not. How can they both be right?
2. Explain three reasons why racial differences in intelligence might be caused by environmental factors. (3 points)

Answer

1 point: Intelligence tests can be considered biased because they, in part, measure experience. Therefore, people from a middle-class, or higher, background are at an advantage.

1 point: Intelligence tests do a good job of predicting the future no matter what demographic group the test-taker is from. In other words, children who do well on intelligence tests are likely to do well in school no matter what their economic or ethnic background is. Likewise, children from all backgrounds who do poorly on the tests do poorly in school.

Unit XI Review

Key Terms and Concepts to Remember

intelligence, p. 607	intelligence quotient (IQ), p. 618	content validity, p. 622
intelligence test, p. 607	achievement test, p. 619	predictive validity, p. 622
general intelligence (g), p. 608	aptitude test, p. 619	cohort, p. 625
factor analysis, p. 608	Wechsler Adult Intelligence Scale (WAIS), p. 620	crystallized intelligence, p. 626
savant syndrome, p. 609	standardization, p. 621	fluid intelligence, p. 626
grit, p. 610	normal curve, p. 621	intellectual disability, p. 629
emotional intelligence, p. 612	reliability, p. 622	Down syndrome, p. 629
mental age, p. 618	validity, p. 622	heritability, p. 632
Stanford-Binet, p. 618		stereotype threat, p. 642

Key Contributors to Remember

Charles Spearman, p. 608	Robert Sternberg, p. 611	Louis Terman, p. 618
L. L. Thurstone, p. 608	Francis Galton, p. 617	David Wechsler, p. 620
Howard Gardner, p. 609	Alfred Binet, p. 618	Carol Dweck, p. 635