

CHAPTER 13: TESTING AND INDIVIDUAL DIFFERENCES

IF YOU LEARN ONLY FIVE THINGS IN THIS CHAPTER . . .

1. Binet created the first intelligence test and developed the concept of mental age, but Terman's revision, the Stanford-Binet, created a way to compute an IQ score.
2. Aptitude tests predict future success and achievement tests assess what individuals already know.
3. When designing tests, psychometricians focus on standardization, reliability, validity, and culture fairness.
4. Normal distributions are bell-shaped curves in which most scores fall near the average and the percentage of scores between standard deviations is fixed by a formula.
5. Reliability refers to a test being repeatable and validity refers to a test being accurate.

INTRODUCTION

Debate continues in psychology over the meaning of **intelligence**, but a generally accepted definition is the mental capacity to solve problems and adapt to the environment. **Francis Galton**, who originally coined the phrase *nature versus nurture*, researched the hereditary basis of intelligence by studying twins. He developed the field of psychometrics by applying key statistical concepts including correlation and percentile rank to studies on human intelligence and other factors.

The first intelligence measure, the **Binet-Simon scale** created by **Alfred Binet**, was designed for French schools to identify students who would benefit from additional support. Binet also

developed the concept of **mental age**, which indicated an individual was capable of reasoning at the level of a typical person at that chronological age. **Lewis Terman** revised and expanded Binet's test for use in the United States and renamed it the **Stanford-Binet Intelligence Scale**. The new test included a formula for determining intelligence quotient (IQ), developed by William Stern, and allowed for testing of adults. IQ was determined by dividing mental age by chronological age and multiplying the result by 100, providing a method for comparing individuals. **David Wechsler** suggested IQ tests were limited because they measured intelligence only verbally; he devised a test that included both **verbal** and **performance** components to compute a total IQ score. Wechsler developed separate intelligence tests for adults (WAIS) and children (WISC) and used a scoring system based on normal distribution.

THEORIES OF INTELLIGENCE

Charles Spearman presented a two-factor theory of intelligence separating general and specific mental abilities. Utilizing the statistical method of **factor analysis**, which identifies groups of associated ideas by combining like items, Spearman discovered that most cognitive skills are related to a single essential trait he called general mental ability or **g factor** intelligence. The **g factor** or **general intelligence** is the ability of individuals to solve complex problems, and the **s factor** or **specific mental abilities** is what he called the ability of an individual to utilize math or verbal skills.

Robert Sternberg proposed the **triarchic theory of intelligence**, in which intelligence was divided into three aspects: **practical**, the ability to adapt to changes in the environment; **analytical**, the reasoning and problem solving that is evaluated by most intelligence tests; and **creative**, the capacity to create new ideas and unique solutions to problems.

Howard Gardner's multiple intelligences theory states that intelligence is not fixed and that all individuals possess at least eight different types of intelligences. The eight intelligences are linguistic, logical-mathematical, visual-spatial, musical, bodily-kinesthetic, intrapersonal, interpersonal, and naturalistic.

DESIGNING INTELLIGENCE TESTS— PSYCHOMETRICS

The field of **psychometrics** is engaged in the design and analysis of quantitative tests for measuring psychological variables such as intelligence and personality traits.

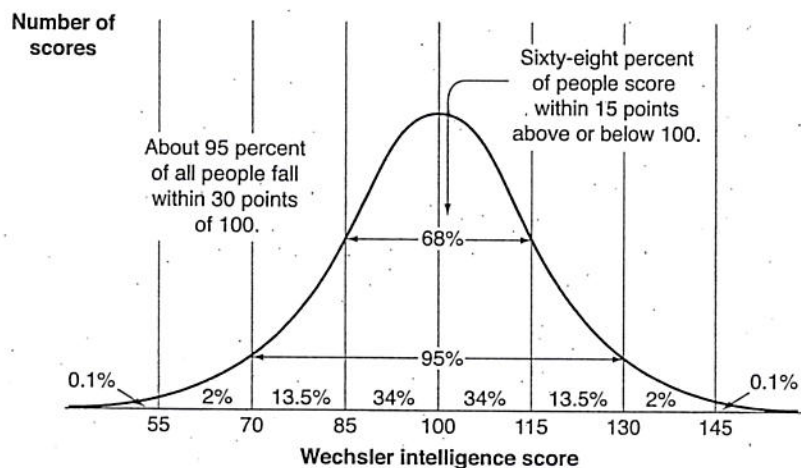
AP EXPERT TIP

To remember the three parts of Sternberg's intelligence and love theories, utilize the acronyms P.A.C. (practical, analytical, and creative intelligence) and P.I.C. (passion, intimacy, and commitment in love).

Aptitude tests are used to predict future success and assess the ability to acquire new skills.

Achievement tests evaluate how well a person has mastered a subject.

Three major aspects of psychometrics in terms of test construction are standardization, reliability, and validity. **Standardization** involves utilizing scores from a representative sample to determine how well an individual did on the test relative to other test takers, and creating similar testing conditions for all individuals taking the exam to prevent any potentially confounding variables. Intelligence tests originally indicated a ratio between a person's mental age and chronological age, but modern tests are based on a normal distribution created through standardization. **Normal distributions** of scores form symmetrical bell-shaped curves in which the **mean**, **median**, and **mode** are equal and located in the center of the distribution and the percentages of scores falling between **standard deviations** are fixed by a formula. In a normal distribution, 68 percent of scores fall within one standard deviation, 95 percent within two standard deviations, and 98 percent within three standard deviations of the mean in either direction. Because percentages in normal distribution are fixed, it is possible to compute the **percentile rank**, or the percentage of scores in a distribution below the score you are considering. For example, an individual with a score of 68, which is greater than 90 percent of scores of all people taking the test, is in the 90th percentile.



Reliability is the degree to which a psychological test such as an IQ test is consistent or dependable. To determine if a test is reliable, psychologists utilize three main methods. **Test-retest** reliability is computed by having the same individuals take the same test at two different times.

Alternate form reliability involves testing the same individuals twice but giving a different version on the retake date. **Split-half** reliability involves checking for consistency between the scores on

two halves of the same test, which can mean comparing the first half to the second half of the test or even questions to odd.

Validity is the degree to which a particular psychological test is accurate and inferences drawn from the results are correct. **Content validity** evaluates how well a test measures the total meaning of the concept and if it is reasonably representative of the material it is evaluating. **Construct validity** refers to whether a test is really evaluating an abstract psychological or theoretical idea. Constructs are difficult to measure and define operationally and include ideas such as extraversion or intelligence. **Criterion or predictive validity** refers to how well test results relate to another measure of what you are evaluating or how well they predict success in the future.

EXTREMES IN INTELLIGENCE

The average IQ score is 100 and scores below 70 signify **mental retardation**. The term *mental retardation* has received much criticism and will likely be replaced with the term *cognitively disabled*. Generally, anyone with an IQ below 70 can be considered mentally retarded; however, psychologists do not use IQ exclusively to determine mental retardation and now utilize additional social factors when making a diagnosis. IQ scores between 50 and 70 are considered mild mental retardation, which makes up the largest percentage of individuals with cognitive deficits. These individuals are capable of achieving grade 6 education levels and living independently. IQ scores between 35 and 49 are considered moderate, scores between 20 and 34 are considered severe, and scores below 20 are classified as profoundly mentally retarded. Known causes of mental retardation include **Down syndrome**, in which individuals are born with all or part of an extra chromosome, or other genetic problems such as **phenylketonuria (PKU)** and **fragile X syndrome**. Mental retardation can also have external causes related to problems during pregnancy such as malnutrition, exposure to toxins, fetal alcohol syndrome, and injuries during birth.

Psychologists also investigate individuals with extremely high IQ scores who are often referred to as **gifted**. One of the most well-known studies of gifted individuals was **Lewis Terman's** longitudinal study of a group of gifted students whom he called "Terman's Termites." Individuals from the study scored higher than average in terms of family income, physical and mental health, and reported happiness. Terman's research disproved a popular misconception that gifted individuals were not well adjusted or successful socially.

IMPACT OF SOCIOCULTURAL FACTORS ON INTELLIGENCE

Although intelligence tests are widely used in the United States and other Western cultures, they are not utilized worldwide. In addition, tests standardized on Western samples do not accurately measure intelligence in other parts of the world. When creating intelligence tests for other cultures, it is important that the tests are designed with the values and experiences of the people in that

culture in mind. Differences related to ethnicity, culture, and gender that exist within Western societies can affect performance on intelligence tests, and critics argue that **culture-fair tests** must include examples relevant to the experiences of the individuals taking the test. The reason for the disparities in scores among groups is a subject of debate in which most psychologists believe that both genetics and environment influence an individual's IQ. A **heritable** or genetic component to intelligence has been established with twin and adoption studies, but research also indicates that socioeconomic status has an impact on IQ.

The **Flynn effect** is the finding that IQ scores have been steadily improving across generations. The reason for the Flynn effect has not been identified, but it is generally considered to have an environmental basis such as better education and nutrition, because 100 years is not long enough to create an evolutionary change. Another potential cause for the gap in intelligence scores among groups may be stereotype threat. **Stereotype threat** results when individuals are reminded of a negative stereotype about a group they belong to prior to a test, which results in their scoring lower. Psychologists today work to create culture-fair tests that eliminate racial, ethnic, socioeconomic, and gender biases.