What We Know about Intelligence from the Weight of Studies

[T]he attack on tests is, to a very considerable and very frightening degree, an attack on truth itself by those who deal with unpleasant and unflattering truths by denying them and by attacking and trying to destroy the evidence for them.

Barbara Lerner (1980)

Intelligence is surely not the only important ability, but without a fair share of intelligence, other abilities and talents usually cannot be fully developed and effectively used … It [intelligence] has been referred to as the "integrative capacity" of the mind.

Arthur Jensen (1981)

The good thing about science is that it's true whether or not you believe in it.

Neil deGrasse Tyson, HBO's *Real Time with Bill Maher*, February 4, 2011

The University of California will no longer consider SAT and ACT scores.

Los Angeles Times, May 15, 2021

Learning Objectives

- How is intelligence defined for most scientific research?
- How does the structure of mental abilities relate to the concept of a general intelligence factor?
- Why do intelligence test scores estimate but not measure intelligence?
- What are four kinds of evidence that intelligence test scores have predictive value?
- Why do myths about intelligence persist?

Introduction

When a computer beats a human champion at games such as chess or Go that require strategy, or a verbal knowledge game such as *Jeopardy*, is the computer smarter than the person? Why can some people memorize exceptionally long strings of random numbers or tell the day of the week for any date in the past, present, or future? What is artistic genius and

is it related to intelligence? These are some of the challenges to defining intelligence for research. It is obvious that no matter how you define it, intelligence must have something to do with the brain, and that is why this book is about neuroscience research.

Among the many myths about intelligence, perhaps the most pernicious is that intelligence is a concept too amorphous and ill-defined for scientific study. In fact, the definitions and measures used for research are sufficiently developed for empirical investigations and have been so for over 100 years. This long research tradition used various kinds of mental ability tests and sophisticated statistical methods known collectively as psychometrics. The new science of intelligence builds on that database and melds it with new technologies of the last two decades or so, especially genetic and neuroimaging methods. These advances, the main focus of this book, are helping to evolve a more neuroscience-oriented approach to intelligence research. The trajectory of this research is similar to that in other scientific fields, which has led from better measurement tools to more sophisticated definitions and understandings of, for example, an "atom" and a "gene." Before we address the brain in subsequent chapters, this chapter reviews the current state of basic research issues regarding the definition of intelligence as a general mental ability, the measurement of intelligence relative to other people, and the validity of intelligence test scores for predicting real-world variables.

1.1 What Is Intelligence? Do You Know It When You See It?

It may seem odd, but let's start our discussion of intelligence with the value of pi, the circumference of a circle divided by its diameter. As you know, the value of pi is always the same: 3.14 … carried out to an infinite, nonrepeating sequence of decimals. For our purpose here, it's just a very long string of numbers in seemingly random order that is always the same. This string of numbers has been used as a simple test of memory. Some people can memorize a longer string of the pi sequence than others. And a few people can memorize a very long string.

Daniel Tammet, a young British man, studied a computer printout of the pi sequence for a month. Then, for a demonstration organized by the BBC, Daniel repeated the sequence from memory publicly while checkers with the computer printout followed along. Daniel stopped over five hours later after correctly repeating 22,514 digits in the sequence. He stopped because he was tired and feared making a mistake (Tammet, 2007).

In addition to his ability to memorize long strings of numbers, Daniel also has a facility to learn difficult languages. The BBC arranged a demonstration of his language ability when they moved him to Iceland to learn the local language with a tutor. Two weeks later, he conversed on Icelandic TV in the native tongue. Do these abilities indicate that Daniel is a genius or, at least, more intelligent than people who do not have these mental abilities?

Daniel has a diagnosis of autism and he may have a brain condition called synesthesia. Synesthesia is a mysterious disorder of sensory perception where numbers, for example, may be perceived as colors, shapes, or even odors. Something about brain wiring seems to be amiss, but it is so rare a condition that research is quite limited. In Daniel's case, he reports that he sees each digit as a different color and shape, and when he recalls the pi sequence, he sees a changing "landscape" of colors and shapes rather than numerical digits. Daniel is also atypical among people with autism because he has a higher-than-average intelligence quotient (IQ) score.

Recalling 22,514 digits of pi from memory is a fascinating achievement no matter how it is accomplished (the official record is an astonishing 70,000 digits – see Chapter 6.2). So is learning to converse in the Icelandic language in two weeks. There are people with extraordinary, specific mental abilities. The term savant is typically used to describe these rare individuals. Sometimes the savant ability is an astonishing memory or the ability to rapidly calculate large numbers mentally or the ability to play any piece of music after only hearing it once or the ability to rapidly create sophisticated artistic drawings or sculptures.

Kim Peek (1951–2009), for example, was able to remember an extraordinary range of facts and figures. He read thousands of books, especially almanacs, and he read each one by quickly scanning page after page. He could then recall this information at will as he demonstrated many times in public forums in response to audience questions: Who was the 10th king of England? When and where was he born? Who were his wives? And so on. Kim's IQ was quite low and he could not care for himself. His father managed all aspects of his life except when he answered questions from memory.

Stephen Wiltshire has a different savant ability. Stephen draws accurate, detailed pictures of city skylines and he does so from memory after a short helicopter tour. He even gets the number of windows in buildings correct. You can buy one of his many city skyline drawings at a gallery in London or online. Alonzo Clemons is a sculptor. He also has a low IQ. His mother claims he was dropped on his head as a baby. Alonzo creates

animal sculptures in precise detail, typically after only a brief look at his subject. The artistry is amazing. Derek Paravicini has a low IQ and cannot care for himself. Blind from birth, Derek is a virtuoso piano player. He amazes audiences by playing any piece of music after hearing it only once, and can play it in any musical style. It is worth noting that Albert Einstein and Isaac Newton did not have any of these memory, drawing, sculpting, or musical abilities.

Savants raise two obvious questions: How do they do it, and why can't I? We don't really know the answer to either question. These individuals also raise a core question about the definition of intelligence. They are important examples of the existence of specific mental abilities. But is extraordinary specific mental ability evidence of intelligence? Most savants are not intelligent. In fact, they typically have low IQ and often cannot care for themselves. Clearly extraordinary but narrow mental ability is not what we usually mean by intelligence.

One more example is Watson, the IBM computer that beat two all-time *Jeopardy* champions. *Jeopardy* is a game where answers are provided and players must deduce the question. The rules were that Watson could not search the web and all information had to be stored inside Watson's 15 petabytes of memory, which was about the size of 10 refrigerators. Here's an example. In the category "Chicks Dig Me," the answer is: "This mystery writer and her archeologist husband dug to find the lost Syrian City of Arkash." This sentence is actually quite complex for a computer to understand, let alone formulate the answer in the form of a question. In case you're still thinking, the answer, in the form of a question is: "Who was Agatha Christie?" Watson answered this faster than the humans, and in the actual match, Watson trounced the two human champions. Does Watson have the same kind of intelligence as humans, or better? Let's look at some definitions to consider if Watson is more like a savant or Albert Einstein.

1.2 Defining Intelligence for Empirical Research

No matter how you define intelligence, you know someone who is not as smart as you are. It would be unusual if you have never called someone an "idiot" or a "moron" or just plain dumb, and meant it literally. And, in all honesty, you know someone who is smarter than you are. Perhaps you refer to such a person in equally pejorative terms such as "nerd" or "egghead," even if in your innermost self you wish you had more "brains." Given their rarity, it is less likely you know a true genius, even if many mothers and fathers say they know at least one.

There are everyday definitions of intelligence that do not lend themselves to scientific inquiry: Intelligence is being smart. Intelligence is what you use when you don't know what to do. Intelligence is the opposite of stupidity (and we all know stupidity when we see it). Intelligence is what we call individual differences in learning, memory, and attention. Researchers, however, have proposed a number of definitions, and mostly they all share a single attribute. Intelligence is a *general* mental ability. Here are two examples:

1. From the American Psychological Association Task Force on Intelligence:

Individuals differ from one another in their ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought. (Neisser et al., 1996)

2. Here's a widely accepted definition among researchers:

[Intelligence is] a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience … It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather it reflects a broader and deeper capability for comprehending our surroundings – "catching on," "making sense" of things, or "figuring out" what to do. (Gottfredson, 1997)

The concept of intelligence as a general mental ability is widely accepted among many researchers but it is not the only concept. What evidence supports the concept of intelligence as a general mental ability, and what other mental abilities are relevant for defining intelligence? How do we reconcile intelligence as a general ability with the specific abilities of savants?

1.3 The Structure of Mental Abilities and the *g*-factor

We all know from our experience that there are many mental abilities. Some are specific, such as spelling or the ability to mentally rotate 3D objects or to rapidly calculate winning probabilities of various poker hands. There are many tests of specific mental abilities. We have over 100 years of research about how such tests relate to each other. Here's what we know: Different mental abilities are not independent. They are all related to each other and the correlations among mental tests are always positive. That means that if you do well in one kind of mental
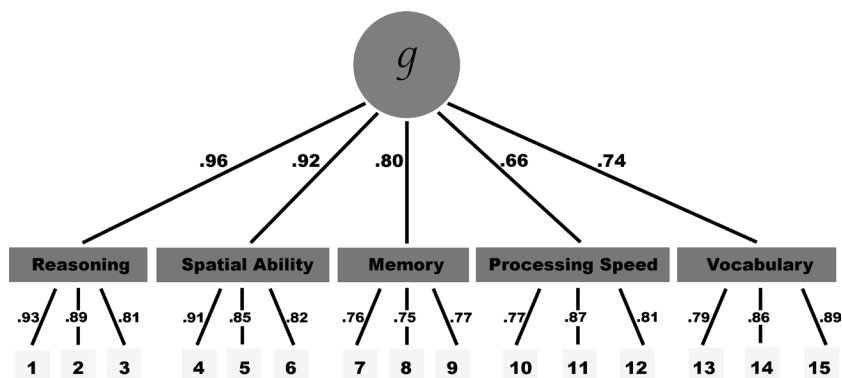
Figure 1.1 The structure of mental abilities. The *g*-factor is common to all mental tests. Numbers are correlations that show the strength of relationship between tests, factors, and *g*. Note all correlations are positive; these are simulated data. (Courtesy Richard Haier)

ability test, you tend to do well in other tests. This may not be the case for any specific person but it is true statistically for populations.

This is the core finding about intelligence assessment and, as we'll see throughout this book, it is the basis for most modern research. Please note this important point: *tend* means there is a higher probability, not a perfect prediction. Whenever we say that one score *predicts* something, we always mean that the score predicts a higher probability for the something.

The relationship among mental tests is called the structure of mental abilities. To picture one possible structure, imagine a three-level pyramid, as shown in Figure 1.1.

At the bottom of Figure 1.1, we have a row of 15 different tests of specific abilities. At the next level up, tests of similar abilities are grouped into more specific factors: reasoning, spatial ability, memory, speed of information processing, and vocabulary. In the illustration, tests 1, 2, and 3, for example, are all reasoning tests and tests 7, 8, and 9 are all memory tests. But all these more specific factors are also related to each other. Basically, people who score high on one test or factor tend to score high on the others (the numbers in the figure are illustrative correlations that show the strength of relationship between tests and factors; see more about correlations in Textbox 1.1). This is a key finding that is demonstrated over and over again. It strongly implies that all the factors derived from individual tests have something in common, and this common factor is called the general factor of intelligence or *g* for short: *g* sits at the highest point on the pyramid in Figure 1.1. The *g*-factor provides

a bridge between the definitions of intelligence that emphasize a general mental ability and individual tests that measure (or, more accurately, estimate) specific abilities.

Most theories about factors of intelligence start with the empirical observation that all tests of mental abilities are positively correlated with each other. This is called the "positive manifold," and Charles Spearman first described it more than 115 years ago (Spearman, 1904). Spearman worked out statistical procedures for identifying the relationships among tests based on their correlations with one another. The basic method is called factor analysis. It works essentially by analyzing correlations among tests. You probably already know about correlations, but see the brief review in Textbox 1.1.

Textbox 1.1: Correlations

*Many of you know about correlations. Since they are ubiquitous throughout this book, here is a brief explanation so everyone starts with an understanding of the concept. Let's say we measure height and weight in many people. We can graph each person by locating the height and weight as a single point with height ranges on the **y**-axis and weight ranges on the **x**-axis. When we add points on the graph for each person, we begin to see an association. Taller people tend to weigh more. You can see this in Figure 1.2. This association is obvious without needing to plot the points, but associations between other variables are not so obvious. Moreover, correlations quantify the strength of association.*

If height and weight were perfectly related, the points would all fall on a straight line and we could predict one from the other without error. A correlation has a value of +1 if a high value on one variable goes perfectly with a high value on the other variable. A strong but not perfect positive correlation is shown in Figure 1.2. A perfect negative correlation is where a high value on one variable predicts a low value on the other without error. A strong but not perfect negative correlation (also called an inverse correlation) is also shown in Figure 1.2. A perfect negative correlation has a value of minus 1. In the Figure 1.2 example, the higher the family income, the lower the rate of infant mortality. Finally, in Figure 1.2 the bottom panel shows no relationship at all (zero correlation) between height and hours of video game playing.

Correlations between two variables are calculated based on how much each point deviates from the perfect line. The higher the correlation, positive or negative, the stronger the relationship and the better one variable predicts the other. Correlations always fall between plus and minus 1. Here is a critical point: A correlation between two variables does not mean one causes the
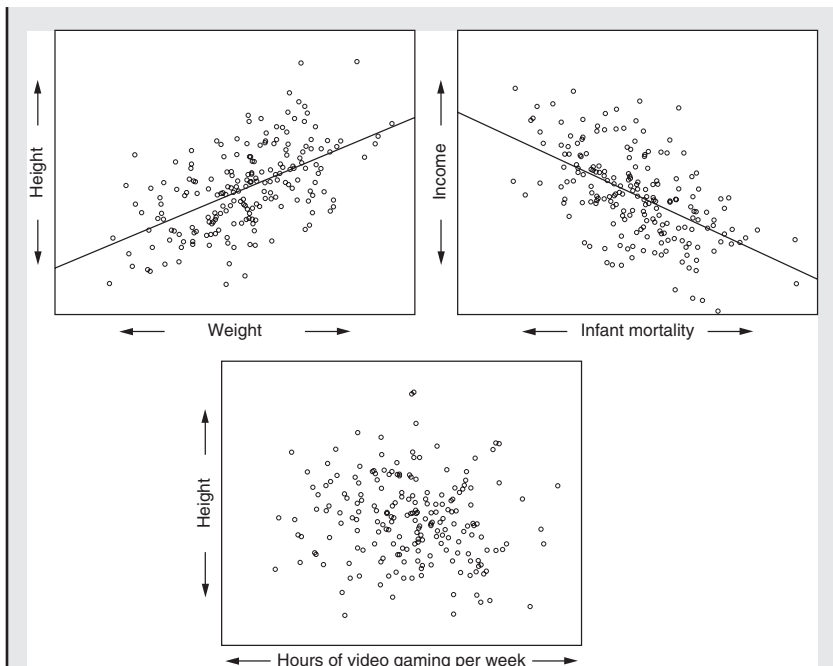

Figure 1.2 An example of a positive correlation is on the left, showing that as height increases weight also increases. A negative correlation is on the right, showing that as family income goes up infant mortality goes down (simulated data). No correlation between height and hours spent playing video games is shown on the bottom. For all of these scatterplots, each circle is a data point. The solid line shows a perfect correlation; the amount that points scatter above and below this line is used to calculate the correlation. (Courtesy Richard Haier)

other. The correlation only means there is a relationship such that as one goes up or down so does the other. To repeat, correlation does not mean causality. Two variables may be correlated to each other but neither causes the other. For example, salt consumption and cholesterol level in the blood may be somewhat correlated but that does not mean one causes the other. The correlation could be caused by a third factor common to both, such as poor diet.

Factor analysis is based on the pattern of correlations among multiple variables. In our case we are interested in the correlations among different tests of mental abilities. So the point of factor analysis is to identify what tests go with other tests, based not on content but rather on

correlations of scores irrespective of content. The set of tests that go with each other define a factor because they have something in common that causes the correlation. Studies in this field typically apply factor analysis to data sets where hundreds or thousands of people have completed dozens of tests.

There are many forms of factor analysis but this is the basic concept, the basis for models of the structure of mental abilities such as the pyramid described in Figure 1.1. Going back to that, note that the correlation values show how strong the associations are among tests, factors, and *g*. Note that all the correlations are positive and illustrative of Spearman's positive manifold.

Let's look at some details of this example in Figure 1.1. The reasoning factor is related to *g* with the strongest correlation of 0.96. This indicates that the reasoning factor is the strongest factor related to *g*, so tests of reasoning are regarded as among the best estimates of *g*. Another way of saying this is that reasoning tests have high *g*-loadings. Note that test 1 has the single highest loading of 0.93 on the reasoning factor so it might provide the single best estimate of *g* if only one test is used rather than a battery of tests. The second strongest correlation is between the spatial ability factor and *g*. It turns out that spatial ability tests are also good estimates of *g*. The vocabulary factor is fairly strong at 0.74, followed by the other factors including memory. In this example, memory tests are good but not the best estimators of *g* with a correlation of 0.80, although other research shows much stronger correlations between working memory and *g* (see Section 6.2).

1.4 Alternative Models

Other statisticians and researchers worked out alternative factor analysis methods. The details don't concern us, but different factor analysis models of intelligence were derived using these various methods. Each identified a different factor structure for intelligence. These various factors emphasize that the *g*-factor alone is not the whole story about intelligence; no intelligence researcher ever asserted otherwise or claimed that a single score captures all aspects of intelligence. The other broad factors and specific mental abilities are important. Depending on how researchers derive factors from a battery of tests, a different number of factors secondary to *g* emerge. In the pyramid structure diagram example, there are five broad factors. Another widely used model is based on only two core factors: crystalized intelligence and fluid intelligence (Cattell, 1971, 1987). Crystalized intelligence refers to the ability to learn facts and

absorb information based on knowledge and experience. This is the kind of intelligence shown by some savants. Fluid intelligence refers to inductive and deductive reasoning for novel problem solving. This is the kind of intelligence we associate with Einstein or Newton. Measures of fluid intelligence are typically highly correlated to measures of *g*, and the two are often used synonymously. Crystalized intelligence is relatively stable over the life span with little deterioration with age, whereas fluid intelligence decreases slowly with age (Schaie, 1993). The distinction between fluid and crystalized intelligence is widely recognized as an important evolution in the definition of intelligence. Both are related so they are not in conflict with the *g*-factor. They represent factors just below *g* in the pyramid structure of mental abilities.

Another factor analysis model focuses on three core factors – verbal, perceptual, and spatial rotation – in addition to *g* (Johnson & Bouchard, 2005). There are also models with less empirical evidence such as those of Robert Sternberg (Brody, 2003; Gottfredson, 2003; Sternberg, 2000, 2003, 2014) that deemphasize *g*, and Howard Gardner (Ferrero, Vadillo, & León, 2021; Gardner, 1987; Gardner & Moran, 2006; Waterhouse, 2006) that ignore the *g*-factor. Virtually all of the neuroscience studies of intelligence, however, use various measures with high *g*-loadings. We will focus on these, but also include several neuroscience studies that investigate factors and specific abilities other than *g*.

1.5 Focus on the *g*-factor

The *g*-factor is the basis of most intelligence assessment used in research today because it alone accounts for about half of the intelligence test score variability among people. It is not the same as IQ, but IQ scores are good estimates of *g* because most IQ tests are based on a battery of tests that sample many mental factors, an important aspect of *g*. Many of the controversies about intelligence have their origins in confusion about how we use words such as mental abilities, intelligence, the *g*-factor, and IQ. Figure 1.3 shows a diagram that will help clarify how I use these words throughout this book.

We have many mental abilities – all the things you can think of from multiplying in your head to picking stocks to naming state capitals. The large circle in Figure 1.3 represents all mental abilities. Intelligence is a catchall word that means the mental abilities most related to responding to everyday problems and navigating the environment, as per the American Psychological Association and the Gottfredson definitions. The circle labeled intelligence is smaller than all mental abilities. IQ is a
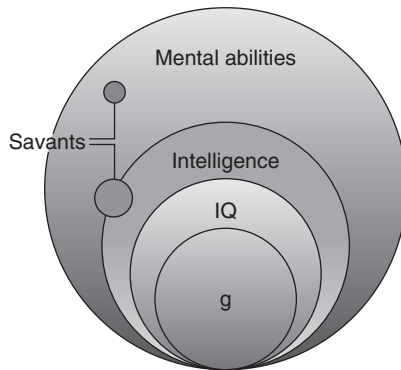
Figure 1.3 Conceptual relationships among mental abilities, intelligence, IQ, and the *g*-factor. (The Intelligent Brain, copyright 2013 The Teaching Company. LLC. Reproduced with permission of The Teaching Company, LLC, www.thegreatcourses.com)

test score based on a subset of the mental abilities that relate to everyday intelligence. The IQ circle is a fairly large part of the intelligence circle because IQ is a good predictor of everyday intelligence. This circle also includes broad factors such as those shown in the diagram of the pyramid structure in Figure 1.1. We describe IQ in more detail in Section 1.6. Finally, the *g*-factor is what is common to all mental abilities. The *g*-factor is a fairly large part of IQ. Whereas everyday intelligence and IQ test scores can be influenced by many factors, including social and cultural ones, the *g*-factor is thought to be relatively more biological and genetic, as we discuss in Chapters 2 and 6.

The savant examples described earlier speak to the level of very specific abilities with little if any *g* in many cases, such as Kim and Derek. They show that powerful independent abilities can exist, but they also show the problems when *g* is lacking. The IBM computer Watson demonstrates a specific ability to analyze verbal information and solve problems based on the meanings of words. This is an amazing accomplishment, but, in my view, Watson does not show the *g*-factor. Watson is more like Kim Peek than Albert Einstein – at least for now. There is a concerted effort among artificial intelligence (AI) researchers to develop general AI, but it is a daunting challenge (Chollet, 2019). Perhaps psychometric or neuroscience insights from the human *g*-factor will be helpful (see Section 6.4).

The savant examples are exceedingly rare cases. Most people have *g* and independent factors to varying degrees, and two people with the same level of *g* can have different patterns of mental strengths and weaknesses

across different mental abilities. Can we ever hope to learn how savants do amazing mental feats, and why we can't? Is it possible that we all have the potential to memorize 22,514 digits or the potential for musical or artistic genius? And why are some people just smarter than others? Does everyone have equal potential for learning all subjects? There are many questions and, as in every scientific field, the answers depend entirely on measurement.

1.6 Measuring Intelligence and IQ

IQ is what most people associate with measuring intelligence. Criticism of IQ and all mental tests is widespread, and has been so for decades (Lerner, 1980). It is worth remembering that the concept of testing mental ability arose to help children get special education. It is also worth stating that intelligence tests are regarded as one of the great achievements of psychology despite many concerns. Let's briefly discuss both these points. Informative, detailed discussions about IQ testing are also found in two classic textbooks (Hunt, 2011; Mackintosh, 2011) and a recent one (Haier & Colom, 2023; see also Coyle, 2021).

In the early part of the twentieth century, the minister of education in France was concerned about identifying children with low school achievement who needed special attention. The problem was how to distinguish children who were "mentally defective" from other children who were low achievers owing to behavioral or other reasons. They wanted the distinction to be made objectively by means of testing so a teacher could not assign a child with discipline issues to a special school as a punishment, as was apparently common at the time.

In this context, Alfred Binet and his collaborator, Theodore Simon, devised the first IQ test to identify children who could not benefit mentally from ordinary school instruction. So the IQ test was born as an objective means for identifying low mental ability in children so they could get special attention, and also to identify children erroneously sent to special schools not because of low mental ability but as a punishment for bad behavior. Both goals were admirable.

The test constructed by Binet and Simon consisted of several subtests that sampled different mental abilities with an emphasis on tests of judgment because Binet felt that judgment was a key aspect of intelligence. He gave each test to many children and developed average scores for each age and sex. He was then able to say at what age level any individual child scored. This was called the child's mental age. A German psychologist named William Stern took the concept of mental age another

step, by dividing mental age by chronological age. This resulted in an IQ score that was the ratio of a child's mental age (averaged across all the subtests) divided by the child's chronological age. Multiplying this ratio by 100 avoided fractions.

For example, if a child was reading at the level of an average 9-year-old, the child's mental age was nine. If this child actually had a chronological age of 9, the IQ would be 9 divided by $9 = 1 \times 100$, or an IQ of 100. If a child had a mental age of 10, but was only 9 years old, the IQ would be 10 divided by $9 = 1.11 \times 100$, or 111. A 9-year-old with a mental age of 8 would have an IQ of 8 divided by $9 = 0.89 \times 100$, IQ= 89.

The point of these early tests was to find children who were not doing so well in school *relative to their peers*, and get them special attention. The Binet–Simon test actually worked reasonably well for this purpose. However, one problem with the concept of mental age is that it is hard to assess after about age 16. Can we really see a mental age difference between a 19-year-old and a 21-year-old? We're not talking about maturity here. The mental age of a 30-year-old really isn't much different than a 40-year-old, so the Binet–Simon test was not really useful or intended for adults.

But there is a much more important measurement problem to keep in mind. Note that the IQ score is a measure of a child *relative to his or her peers*. Even today, newer IQ tests based on a different calculation, discussed later, show how an individual scores *relative to his or her peers.* IQ scores are not absolute measures of a quantity, such as pints of water or feet of distance. *IQ scores are meaningful only relative to other people.* Note that intelligence differences among people are quite real, but our methods of measuring these differences depend on test scores that are interpretable only in a relative way. We elaborate this key point shortly and return to it throughout this book.

Nonetheless, the Binet–Simon test was an important advance for assessing the abilities of children in an objective way. The Binet–Simon test was translated to English and redone at Stanford University in the 1920s by Professor Louis Terman and the test is now known as the Stanford–Binet test. Professor Terman used very high IQ scores from this test to identify a sample for a longitudinal study of "genius," which we discuss in Section 1.10.4.

The Wechsler Adult Intelligence Scale (WAIS) was designed with subtests such as the Stanford–Binet, but as its name states, it was designed for adults. It is the most widely used intelligence test today. The current version consists of a battery of ten core subtests and another five supplemental subtests. Together, they sample a broad range of mental abilities.
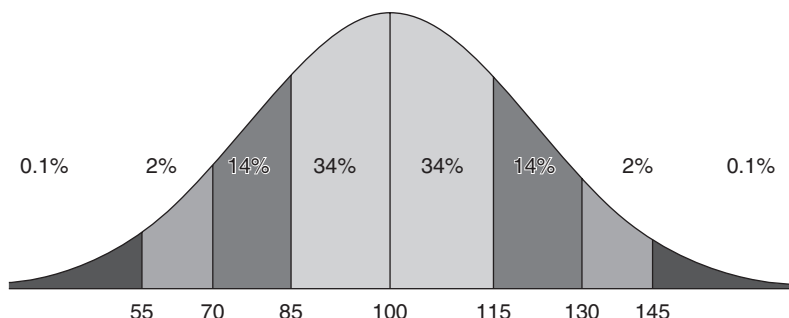
Figure 1.4 The normal distribution of IQ scores and the percentage of people within each level. (Courtesy Richard Haier)

One key change is in the way IQ is calculated in both the WAIS and the Stanford–Binet tests. Mental age is no longer used. IQ is now based on the statistical properties of the normal distribution and deviation scores. The concept is simple: How far from the norm does an individual's score deviate?

Here's how deviation scores work. Let's start with the properties of a normal distribution (also called a bell curve because of its shape), as shown in Figure 1.4.

Many variables and characteristics such as height or income or IQ scores are normally distributed in large populations of randomly selected individuals. Most people have middle values, and the number of individuals decreases toward the low and high extremes of the distribution. Any normal distribution has specific statistical properties in that any individual score can be expressed as a percentile relative to other people. This is shown in the illustration of IQ scores where the mean score is 100 and the standard deviation is 15 points. Standard deviations show the degree of spread around the mean and are calculated as a function of how much each person deviates from the group mean. In a normal distribution, 50 percent of people score below 100, while 68 percent of individuals fall between plus one and minus one standard deviations, so scores between 85 and 115 are regarded as the range of average IQ. A score of 130, two standard deviations above the mean, would be at about the 98th percentile, which is the top 2 percent. A score of 70 would be two standard deviations below the mean and represent about the second percentile. A score of 145 represents the top 10th of 1 percent. Scores over 145 are often considered to be in the genius range, although few tests are accurate at this extreme high end of the distribution.

IQ tests were developed so scores would be normally distributed. Each subtest has been taken by a large number of males and females of different ages. These are the norm groups. Each norm has an average score called the mean, and the spread of scores around the mean is measured by a statistic called the standard deviation.

Let's say a subtest has a perfect possible score of 20 points. Each norm group may have a different average score on this test depending, say, on age. Younger test takers may average 8 points if they are 10 years old, and older children taking the same test, say at age 12, may average a score of 14 points. This is why it's important to have norm groups for each age. If a new 12-year-old takes the subtest and scores 14, he is scoring at the average for his age. If he scores above or below 14, the deviation from the norm average can be calculated and his score can be expressed by how much it deviates from the mean. The average deviation across all the subtests is used to calculate the deviation IQ for the full battery. As illustrated, deviation scores are easily convertible into percentiles.

Each deviation point is equal, but these scores only have meaning relative to other people. In technical terms, these scores are not a ratio scale because there is no actual zero point. This is unlike quantitative units of weight or distance or liquid, which are ratio scales. IQ scores and their interpretation depend on having good normative groups. This is one reason that new norms are generated periodically for these tests. It is also why there is a separate version of the test for children called the Wechsler Intelligence Scale for Children.

The WAIS can be divided into specific factors other than the full scale IQ (FSIQ) score that closely resemble the pyramid structure of mental abilities shown in Figure 1.1. The individual subtests are grouped at the next highest level into factors of verbal comprehension, working memory, perceptual organization, and processing speed. These four specific factors are grouped into more general factors of verbal IQ and performance IQ, and these two broad factors have a common general factor defined by the total IQ score, or FSIQ. This is based on several tests that sample a range of different mental abilities, and is therefore a good estimate of the *g*-factor. Each of the factor scores can be used for other predictions, but FSIQ is the most widely used score in research.

1.7 Some Other Intelligence Tests

So far, the IQ tests we have described are administered by a trained test-giver interacting with one individual at a time until the test is completed, often taking 90 minutes or more. Other kinds of psychometric intelligence
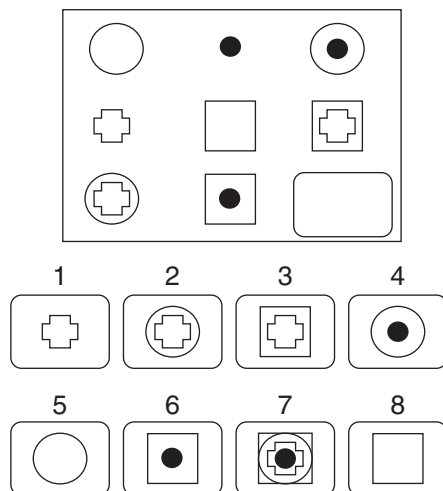
Figure 1.5 Simulated problem from the RAPM test. The lower right corner is missing from the matrix of symbols. Only one of the eight choices fits that spot once you infer the pattern or rule. In this case the answer is seven (add one row or column to the next). (Courtesy of Rex Jung)

tests can be given in a group setting or without direct interaction with the test-giver. Some tests are designed to assess specific mental abilities and others are designed to assess general intelligence. Typically, the more a test requires complex reasoning, the better it estimates the *g*-factor. Such tests have a "high *g*-loading." Here, briefly, are three important high-*g* tests used in neuroscience studies in addition to IQ.

1. The Raven's Advanced Progressive Matrices (RAPM) test (named for its developer, Dr. Raven) can be given in a group format and usually has a time limit of 40 minutes. It's regarded as a good estimate of the *g*-factor, especially because of the time constraint. Tests with a time limit tend to separate individuals better. It's a nonverbal test of abstract reasoning. Figure 1.5 is an example of one item. In the large rectangle, you see a matrix of eight symbols and a blank spot in the lower right corner. The eight symbols are not arranged randomly. There is a pattern or a rule linking them. Once you deduce the pattern or rule, you can decide which of the eight choices below the matrix completes the pattern or rule and goes in the lower right corner.

In this example, the answer is seven. If you add the left column to the middle column in the matrix, you get the symbols in the right column. If you add the top row to the middle row, you get the bottom row. The

actual test items get progressively more difficult. The underlying pattern or rule can be quite hard to infer and there are different versions of the test that vary in difficulty. But because of its simple administration, this test has been used in many research studies. Performance on a test like this is fairly independent of education or culture. Scores are reasonable estimates of the *g*-factor, but they should not be mistaken as the *g*-factor (Gignac, 2015).

2. Analogy tests also are very good estimators of *g*. For example, wing is to bird as window is to _____ (answer is house). Or, helium is to balloon as yeast is to _____ (answer is dough). Or how about Monet is to art as Mozart is to _____ (answer is music). Analogy tests look as if they could be easily influenced by education and culture, so they have been dropped from many assessment test batteries despite the fact that empirically they are good estimates of *g*.

3. The Scholastic Assessment Test (SAT) is an interesting example. Until recently, it was widely used for college admission (see Textbox 1.2). Is it an achievement test, an aptitude test, or an intelligence test? The SAT was originally called the Scholastic Aptitude Test, then it was renamed the Scholastic Achievement Test, and now it's called the Scholastic Assessment Test. Achievement tests measure what you have learned. Aptitude tests measure what you might learn, especially in a specific area, such as music or a foreign language. It turns out that the SAT, especially the overall total score, is a good estimator of *g* because the problems require reasoning (Coyle, 2015; Frey & Detterman, 2004); see also (Beaujean et al., 2006; Coyle, 2021; Koenig, Frey, & Detterman, 2008). Like IQ scores, SAT scores are normally distributed and interpreted best as percentiles. For example, people in the top 2 percent of the SAT distribution tend to be in at least the top 2 percent of the IQ distribution. Sometimes this surprises people, but why should intelligence not be related to how much someone learns?

Achievement, aptitude, and intelligence test scores are all related to each other. They are not independent. Remember, the *g*-factor is common to all tests of mental ability. It would be unusual if learning and intelligence were unrelated. So your performance on achievement tests is related to the general factor, just as IQ scores and aptitude test scores are related to *g*. It can be confusing because we all know examples of bright students who are underachievers, and not-so-bright students who are overachievers. However, such examples are exceptions. In reality, there are some valid distinctions among achievement, aptitude, and intelligence testing. Each kind of test is useful in different settings, but they are also all related to *g*.

> **Textbox 1.2: Is this a case for or against using standardized tests for college admission?**
>
> The University of California (UC) decided in 2021 to end the use of standardized tests but this was not consistent with findings from a UC task force charged with evaluating the issues. Part of the task force conclusion was that they found "that standardized test scores aid in predicting important aspects of student success, including undergraduate grade point average (UGPA), retention, and completion. At UC, test scores are currently better predictors of first-year grade point average (GPA) than high school grade point average (HSGPA), and about as good at predicting first-year retention, UGPA, and graduation. For students within any given (HSGPA) band, higher standardized test scores correlate with a higher freshman UGPA, a higher graduation UGPA, and higher likelihood of graduating within either four years (for transfers) or seven years (for freshmen). Further, the amount of variance in student outcomes explained by test scores has increased since 2007, while variance explained by high school grades has decreased, although altogether does not exceed 26 percent. Test scores are predictive for all demographic groups and disciplines, even after controlling for HSGPA. In fact, test scores are better predictors of success for students who are Underrepresented Minority students (URMs), who are first-generation, or whose families are low-income: that is, test scores explain more of the variance in UGPA and completion rates for students in these groups. One consequence of dropping test scores would be increased reliance on HSGPA in admissions. The [task force] found that California high schools vary greatly in grading standards, and that grade inflation is part of why the predictive power of HSGPA has decreased since the last UC study."
>
> They also "noted the average differences in test scores among groups and expected to find that test score differences explain differences in admission rates. That is not what we found. Instead, the [task force] found that UC admissions practices compensated well for the observed differences in average test scores among demographic groups. This likely reflects UC's use of comprehensive review, as well as UC's practice of referencing each student's performance to the context of their school" (https://senate.universityofcalifornia.edu/_files/underreview/sttf-report.pdf). These empirical findings appear to undermine the administrative decision that ended the use of these tests.

1.8 Myth: Intelligence Tests Are Biased or Meaningless

Are intelligence test questions fair or do correct answers depend on an individual's education, social class, or factors other than intelligence? A professor I had in graduate school used to say that most people define

a fair question as one they can answer correctly. Is a question unfair or biased because you don't know the answer?

Just what do intelligence test scores actually mean? Low test scores result because a person doesn't know the answers to many questions. There are numerous possible reasons for not knowing the answer to a question: You were never were taught it, never learned it on your own, learned it but forgot it a long time ago, learned it but forgot it during the test, were taught it but couldn't learn it, didn't know how to reason it out, or couldn't reason it out. Most, but not all, of these reasons seem related to general intelligence in some way. High test scores, on the other hand, mean the person knows the answers. Does it matter how you came to know the answer? Is it better education, just good memory, good test-taking skills, or good learning? The definitions of general intelligence combine all these things.

Test bias has a specific meaning. If scores on a test consistently over or underpredict actual performance, the test is biased. For example, if people in a particular group with high SAT scores consistently fail college courses, the test is overpredicting success and it is a biased test. Similarly, if people with low SAT scores consistently excel in college courses, the test is underpredicting success and it is biased. A test is not inherently biased just because it may show an average difference between two groups. A spatial ability test, for example, may have a different mean score for men and women, but that does not make the test biased. If scores for men and for women predict spatial ability equally well, the test is not biased even if there is a mean difference. Note that a few cases of incorrect prediction do not constitute bias. For a test to be biased, there needs be a consistent failure of prediction in the wrong direction. The lack of any prediction is not bias; it means the test is not valid.

Considerable research on test bias for decades shows this is not the case for IQ and other intelligence test scores (A. Jensen, 1980). Test scores do predict academic success irrespective of social economic status (SES), age, sex, race, and other variables. Scores also predict many other important variables, including brain characteristics such as regional cortical thickness or cerebral glucose metabolic rate, as we detail in Chapters 3 and 4. If intelligence test scores were meaningless, they would not predict any other measures, especially quantifiable brain characteristics. In this context, "predict" also has a specific meaning. To say a test score predicts something only refers to a higher probability of the something occurring. No test is 100 percent accurate in its predictions, but the reason intelligence tests are considered by many psychologists to be a great achievement is that the scores are good predictors for success in many areas, and in some areas test scores are excellent predictors. Before we

review key research that is the basis for this conclusion, there is a funda-
mental problem to discuss.

1.9 The Key Problem for "Measuring" Intelligence

As briefly noted earlier in this chapter, the main problem with all intel-
ligence test scores is that they are not on a ratio scale. This means there
is no true zero, unlike measures for height and weight. For example, a
person who weighs 200 pounds is literally twice the weight of a person
who weighs 100 pounds because a pound is a standard unit on a scale
with an actual zero point. Ten miles is twice the distance of five miles.
This is not the case for IQ scores. A person with an IQ score of 140 is not
literally twice as smart as a person with a score of 70. Even if you believe
you have encountered at least one person with zero intelligence, zero is
certainly not the case. For IQ, it's the percentile that counts. Someone
with an IQ of 140 is in the top 1 percent and someone with a score of 70
is in the bottom 2 percent. A person with an IQ of 130 is not 30 percent
smarter than a person whose score is 100. The person with an IQ of 100
is at the 50th percentile and the person with an IQ of 130 is at the 98th
percentile. No psychometric test score is based on a ratio scale. All IQ
test scores have meaning only relative to other people.

Here's the key point about this limitation of all intelligence test scores:
They only estimate intelligence because we don't yet know how to mea-
sure intelligence as a quantity like we measure liquid in liters or distance
in feet (Haier, 2014). If you take an intelligence test when you are sick
and unable to concentrate, your score may be a bad estimate of your
intelligence. If you retake the test when you are well, your score is a
better estimate. However, just because your score goes up, it does not
mean your intelligence increased in the interval between the two tests.
This becomes an issue in Chapter 5 when we talk about why claims of
increasing intelligence are not yet meaningful.

Despite this fundamental problem, researchers have made considerable
progress. The main point is that measurement is required to do scientific
research on intelligence. No one test may be a perfect measure of a single
definition, but as research findings accumulate, both definition and mea-
surement evolve and our understanding of the complexities increases. The
empirical robustness of research on the *g*-factor essentially negates the myth
that intelligence cannot be defined or measured for scientific study. It is
this research base that allows neuroscience approaches to take intelligence
research to the next level, as detailed in subsequent chapters. But first, we
will summarize some compelling studies of intelligence test validity.

1.10 **Four Kinds of Predictive Validity for Intelligence Tests**

1.10.1 **Learning Ability**

IQ scores predict general learning ability, which is central to academic and vocational success and to navigating the complexities of everyday life (Gottfredson, 2003). For people with lower IQs of around 70, simple learning typically is slow and requires concrete, step-by-step teaching with individual instruction. Learning complex material is quite difficult or not possible. IQs around 80–90 still require very explicit, structured individual instruction. When it comes to learning by written materials, IQs of at least 100 are usually required, and college-level learning usually works best at 115 and over. Higher IQs over 130 usually mean that more abstract material can be learned relatively quickly, and often independently. There are exceptions, and there is good evidence that lower-IQ (< 90) students who complete college also benefit in later life success as much as higher-IQ students, possibly because of strong compensatory factors such as personality and family support (McGue et al., 2022).

The US military uses their own test but the rough equivalent IQ score minimum cutoff is about 85–90 for recruits, although this moves down a bit when recruitment is strained; sometimes with tragic results (Gregory, 2015). Most graduate programs in the United States require tests such as the Graduate Record Exam or the Medical College Admission Test for medical school or the Law School Admission Test for law school. Cutoffs for these tests usually ensure that individuals with IQs over 120 are most likely to be accepted, and the top programs have higher cutoffs to maximize accepting applicants in the top 1 or 2 percent of the normal distribution. This doesn't mean that people with lower scores cannot complete these programs, but the higher-scoring students are usually more efficient, faster learners and more likely to successfully finish the program.

Keep in mind, these are not perfect relationships and there are exceptions, as noted (McGue et al., 2022). The relationship between IQ scores and learning ability, however, is strong. Many people find this disturbing because it indicates a limitation on personal achievement that runs counter to a prevalent notion expressed in the phrase: "You can be anything you want to be if you work hard." This is a restatement of another notion: "If you work hard, you can be successful." The latter may often be true because success comes in many forms and for many reasons, but the former is seldom true unless a caveat is added: "You can be anything you want to be if you work hard and have the ability." Not everyone has the ability to do everything successfully, although, surprisingly, many students arrive at college determined to succeed but naïve about the role

ability plays. Few students with low SAT-Math scores, for example, are successful majors in the physical sciences even if they are highly motivated and work hard.

Given the powerful influence of *g* on educational success, it is surprising that intelligence is rarely considered explicitly in vigorous debates about why pre-college education appears to be failing many students. Doug Detterman has noted, "[a]s long as educational research fails to focus on students' characteristics, we will never understand education or be able to improve it" (Detterman, 2016: 1). The best teachers cannot be expected to attain educational objectives beyond the capabilities of students. The best teachers can maximize a student's learning, but the intelligence level of the student creates some limitations, although it is fashionable to assert that no student has inherent limitations. Many factors that limit educational achievement can be addressed, including poverty, poor motivation, lack of role models, family dysfunction, and so on, but so far there is no evidence that alleviation of these factors increases *g*. As we will see in Chapter 2, early childhood education has a number of beneficial effects but increasing intelligence is not one of them. Imagine a pie chart with all the factors that influence a student's school achievement. Surely the *g*-factor would deserve representation as a slice greater than zero. The strong correlations between intelligence test scores and academic achievement indicate that the slice could represent a sizable portion of the whole. In my view, this alone should justify more research on intelligence and how it develops.

1.10.2 Job Performance

In addition to academic success, IQ scores also predict job performance (F. Schmidt, 2016; F. L. Schmidt & Hunter, 2004; F. L. Schmidt & Hunter, 1998), especially when jobs require complex skills. In fact, for complex jobs the *g*-factor predicts success more than any other cognitive ability (Gottfredson, 2003). A large study conducted by the US Air Force, for example, found that *g* predicted virtually all the variance in pilot performance (M. J. Ree & Carretta, 1996; Malcolm James Ree & Carretta, 2022; M. J. Ree & Earles, 1991). Most of us are not pilots, but in general, lower IQ is sufficient for jobs that require a minimum of complex, independent reasoning. These jobs tend to follow specific routines such as assembling a simple product, food service, or nurse's aide. IQs around a 100 are minimally necessary for more complex jobs such as bank teller and police officer. Successful managers, teachers, accountants, and others in similar professions usually have IQs of at least 115. Professions

such as attorney, chemist, doctor, engineer, and business executive usually require higher IQs to finish the advanced schooling that is required and to perform at a high level of complexity.

Complex job performance is largely *g*-dependent, but of course there are other factors, including how well one deals with other people. This is the concept of emotional intelligence. Emotional intelligence, that is, the personality and social skills one has, may contribute to greater success compared to a person of equal *g* but lacking people skills. This does not diminish the importance of the *g*-factor. In some circumstances emotional intelligence might compensate for a lack of job-appropriate *g* but only for so long, if at all. Good evidence shows that IQ scores are more predictive of educational attainment (years of education) than personality measures (Zisman & Ganzach, 2022).

As with academic success, intelligence–job performance relationships are general trends, and there are always exceptions. But, from a practical point of view, a person with an IQ under 100 is not very likely to complete medical school or engineering school. Of course, it's possible, especially if the IQ score is not a good estimate of intelligence for that person, or if that person has a very specific ability such as memorization to compensate for low or average general intelligence. Similarly, a high score does not guarantee success. This is why an IQ score by itself is not usually used to make education or employment decisions. IQ is usually considered in the context of other information but a low score is typically a red flag in many areas that require complex, independent reasoning.

Here's another point about predicting job success. Some researchers suggest that expertise in any area requires at least 10,000 hours of practice (Ericsson, 2014; Ericsson & Towne, 2010). That's 1,250 8-hour days, or about 3.4 years of constant effort. This implies that expertise can be achieved in any field with this level of practice irrespective of intelligence or talent. Studies of chess grand masters, for example, report that the group average IQ is about 100. This suggests that becoming a grand master may depend more on practice of a specific ability such as spatial memory than on general intelligence. Grand masters may actually have a savant-like spatial memory, but the idea of a chess grand master being a super all-purpose giant intellect is not necessarily correct. Many studies refute the idea that 10,000 hours of practice can lead to expertise if there is no pre-existing talent to build on (Detterman, 2014; Grabner, 2014; Grabner, Stern, & Neubauer, 2007; Gullich, Macnamara, & Hambrick, 2022; Hambrick, Macnamara, & Oswald, 2020; Macnamara & Maitra, 2019; Plomin et al., 2014a, 2014b).

1.10.3 Everyday Life

The importance of general intelligence in everyday life often is not obvious but it is profound. As Professor Earl Hunt has pointed out, if you are a college-educated person, it is highly likely that most of your friends and acquaintances are as well. When is the last time you invited someone to your home for dinner that was not college-educated? Professor Hunt calls this cognitive segregation and it is powerful in fostering the erroneous belief that everyone has a similar capacity or potential for reasoning about daily problems and issues. Most people with high *g* cannot easily imagine what daily life is like for a person with low *g*.

The complexity of everyday life is often quite challenging, especially when a nonroutine or a novel problem presents itself. Professor Robert Gordon summarizes this with a simple statement: "Life is a long mental test battery" (Gordon, 1997). This was true as early humans navigated unforgiving natural environments and solved continuous problems of finding food, water, shelter, and safety. It was true as early civilizations developed and great thinkers (likely with high *g*) solved even more complex problems (e.g., just how does one build a seaworthy ship or a pyramid?). And it is still true today as we grapple with connecting our new television sets and audio systems with HDMI cables or using all the functions in our word processor or on our "smart" phones or digital cameras beyond the auto mode. Do you know how to use the scanners in the self-checkout lines at the supermarket or do you wait in a long slow line for a human cashier? How much do you understand about money management and investing in stocks, bonds, and mutual funds? Do you do your own taxes? Many people grapple daily with the challenges of navigating nearly impenetrable systems for healthcare, social support, or justice. Poverty presents myriad daily problems to solve. It could be said that in the modern world, nothing is simple for anyone all the time.

Consider some statistics comparing low- and high-IQ groups (low = 75–90; high = 110–125) on relative risk of several life events. For example, the odds of being a high school dropout are 133 times more likely if you're in the low group. People in the low group are 10 times more at risk for being a chronic welfare recipient. The risk is 7.5 times greater in the low group for incarceration, and 6.2 times more for living in poverty. Unemployment and even divorce are a bit more likely in the low group. IQ even predicts traffic accidents. In the high-IQ group, the death rate from traffic accidents is about 51 per 10,000 drivers, but in the low-IQ group this almost *triples* to about 147. This may be telling us that people with lower IQ, on average, have a poorer ability to assess risk and may take more chances when driving or performing other activities (Gottfredson, 2002, 2003).

Textbox 1.3: Functional literarcy

Another way to look at the role of thinking skills and everyday life is based on functional literacy data. Functional literacy is assessed by the complexity of everyday tasks that a person can complete. Like IQ scores, functional literacy scores are meaningful relative to other people, but they provide more concrete examples of ability. The last comprehensive US national survey of functional literacy was done in 1992.

The chart in Table 1.1 is from that survey. On the left side, we have five levels of functional literacy: 1 is the lowest, 5 is the highest. In the middle we have the percentage of people who are in each category, and on the left we have some sample tasks that people in each category can complete successfully. Let's look at the top row. If you're like me, you will be quite surprised to see that only 4 percent of the white population is in the top category and can complete tasks such as using a calculator to figure out the cost of carpeting a room. This requires determining the area, converting to square yards, and multiplying by the price. In the next row down, 21 percent of people are at level four of functional literacy. They can calculate social security benefits from a table and understand basic issues of how employee benefits work. Then 36 percent are in the middle category. They can calculate miles per gallon from a chart, and they can write a letter explaining a credit card

Table 1.1 Everyday literacy levels from the National Adult Literacy Survey along with sample problems from each level

Everyday literacy (NALS)

NALS level	% pop. (white)	Simulated everyday tasks
5	4	• Use calculator to determine cost of carpet for a room • Use table of information to compare two credit cards
4	21	• Use eligibility pamphlet to calculate Supplemental Security Income benefits • Explain difference between two types of employee benefits
3	36	• Calculate miles per gallon from mileage record chart • Write brief letter explaining error on credit card bill
2	25	• Determine difference in price between two show tickets • Locate intersection on street map
1	14	• Total bank deposit entry • Locate expiration date on driver's license

> error. Twenty-five percent are in category two. They can determine price differences between two tickets, and they can locate an intersection on a map. Fourteen percent are in the lowest category. They can accomplish tasks such as filling out a bank deposit slip, but more complex tasks, such as locating an intersection on a map, would present difficulty. Note that these data are more than 30 years old and it may be that the percentages of people in each category differ now, but the main point is still the same: intelligence matters in everyday life.

The examples in Textbox 1.3 and Table 1.1 demonstrate that intelligence helps us navigate the problems of everyday life. It's not a shocking idea. But this is easy to take for granted, especially if you are navigating reasonably well and most of the people you spend time with are like you. The key point here is that functional literacy is another indicator of intelligence, and you can see from the functional literacy data that intelligence matters for daily tasks. But, of course, the *g*-factor does not predict many other important things such as being a kind or likable or honest person. No intelligence researcher has ever asserted otherwise.

Let's talk for a moment about a controversial book in 1994 that explored the role of intelligence in social policy, *The Bell Curve* by Richard Herrnstein and Charles Murray (Herrnstein & Murray, 1994). The main theme was that modern society increasingly requires and rewards people with the best reasoning skills. This is to say people with high intelligence. Therefore, people in the bottom part of the normal distribution of IQ (a normal distribution is also called a bell curve because of its shape) will be at a serious disadvantage for succeeding, especially in school and some vocations. Herrnstein had introduced this theme in an earlier book, *IQ in the Meritocracy* (Herrnstein, 1973), that also generated considerable acrimony (see the detailed description of hostility on the Harvard campus recounted in the preface to get a sense of the times); a few years later another Harvard professor, Edward O. Wilson, encountered similar outrage when he proposed the concept of sociobiology (Wilson, 1975). *The Bell Curve* continued the argument with over 900 pages of data and statistical analyses mostly comparing high and low intelligence groups, but the one chapter that discussed black/white IQ differences aroused the fiercest controversy (please note that the terms black and white are used here because most of the research, from America and other countries, uses these terms). This issue of group differences haunts all intelligence research and I refer the reader to in-depth accounts of the complexities involved (see Further Reading).

My point about *The Bell Curve* is whether public policy discussions benefit by recognizing that people with low *g* may need help in navigating life, irrespective of race, background, or why they might have low *g*. This is a fundamental issue today in politics, although the role of intelligence is hardly mentioned as explicitly as it was in *The Bell Curve*, or in a later book by Murray expanding the theme of societal implications of cognitive segregation (C. A. Murray, 2013). Most researchers would agree that research data on intelligence can only inform policy decisions, but the goals of the policy need to be determined through democratic means; we return to this issue in Section 6.6. Unfortunately, psychometric research on intelligence has often been portrayed as damaging to a progressive social agenda because there are substantial average test score differences among some racial and ethnic groups. These relative average group differences often motivate a general disregard for empirical research on intelligence although neuroscience approaches are advancing the field, as the following chapters discuss. Before we get to those, let's continue with more data about IQ scores and what they mean.

1.10.4 Longitudinal Studies of IQ and Talent

The predictive power of a single test score in childhood is also demonstrated dramatically in three classic longitudinal studies. Each one starts with children and tests their mental abilities and subsequent life successes at various intervals over decades. One study started in California the 1920s, one started in Scotland the 1930s, and one started in Baltimore in the 1970s.

Study 1. Professor Lewis Terman at Stanford University initiated a long-term study of high-IQ individuals in the 1920s. This is the same Louis Terman who brought Binet's IQ test to the United States and revised it into the Stanford–Binet intelligence test. Terman designed a straightforward study. It started by testing many school children with the Stanford–Binet test. Children with very high IQ scores were selected and then studied extensively for decades. Terman's study had two goals: to find the traits that characterized high-IQ children, and to see what kind of adults they would become. The common stereotype of intelligent adults was not so different then as it is now. Francis Galton, for example, wrote in his 1884 book, *Hereditary Genius* (Galton & Prinzmetal, 1884): "There is a prevalent belief that men of genius are unhealthy, puny beings – all brain and no muscle – weak-sighted, and generally of poor constitutions" (Galton, 2006: 321).

Here's how Terman's project started (Terman, 1925): In 1920–1921, 1,470 children with IQs of 135–196 were selected from over 250,000 children in California's public schools and they were retested and interviewed every seven years. Their average IQ was about 150, and 80 children had IQs over 170 (these were in the top 0.1 percent). This entire group became known unofficially as the Termites. They completed extensive medical tests, physical measurements, achievement tests, character and interest tests, and trait ratings, and both parents and teachers supplied additional information. A control group with average IQ scores was also tested. The results of Terman's study were published over time in five volumes. The data were quite extensive.

Here's a summary of the key findings about the lives of the Termites. Overall, they completely refute the stereotypes both for children and adults. The negative, nerdy attributes were basically unfounded. They were not odd or puny. On average, they were actually physically quite robust and more physically and emotionally mature than their age-mates. On average, the Termites were happier and better adjusted than the controls over the course of the study. Although they had their share of life problems, follow-up studies showed considerable achievement with respect to publishing books, scientific papers, short stories and poems, musical compositions, television and movie scripts, and patents (Terman, 1954). However, further follow-up indicated that high IQ alone did not necessarily predict life success. Motivation was also important, and Terman believed that while genes played an important role in high IQ, he also believed that exceptional ability required exceptional education to maximize a student's potential. This may not sound so radical, but even today there is a debate about whether any educational resources at all should be allocated to the most gifted students to develop their high ability.

Terman's project also demonstrated the predictive validity of the IQ score. That is, one IQ score in childhood can identify individuals who will excel in later life. Like all studies, however, there were some major flaws: (1) Terman intervened in the lives of these "subjects" and helped them with letters of reference for college and for employment; (2) strong sex bias in education and employment resulted in female Termites mostly becoming housewives, so valid male–female comparisons were not possible. Similarly, there are no data about minorities. Do these problems invalidate the main findings? Not likely (Warne, 2019). Overall, the level of success and the achievement of these very high-IQ individuals stand on their own. But fortunately, we have more data from a newer study that modified Terman's approach.

Study 2. The second longitudinal study is the Study of Mathematically and Scientifically Precocious Youth at Johns Hopkins University. This was an ambitious, longitudinal project initiated by Professor Julian Stanley in 1971 (Stanley, 1974). Dr. Stanley repeated Terman's approach, but instead of IQ scores he used extremely high SAT-Math scores obtained by junior high school students aged 11–13 in special testing sessions called "talent searches." So instead of general intelligence, Stanley focused on a very specific mental ability. This project also had two major goals. First, identify precocious students early, and second, foster their special talent.

I started graduate school at Hopkins in 1971 and I worked on this study in its early years. I must say that this experience was an early influence on my interest in intelligence, and Dr. Stanley was one of the most important and interesting mentors I had at Hopkins.

This project had its origins in the late 1960s. Dr. Stanley started working with a precocious student, and after he gave the student a battery of psychometric tests, Dr. Stanley helped the student to get into Hopkins at the early age of 13. Dr. Stanley subsequently referred to this young man as the first "Radical Accelerant," identified as Joseph B. In his first year at Hopkins, at age 13, Joseph took honors calculus, sophomore physics, and computer science, and his GPA was 3.69 out of 4.0. He lived at home during this time but he also made friends with other college students and adjusted well to his accelerated studies. In four years, he received a BA *and* a MSc degree in computer science. He began a PhD program in computer science at Cornell before he was 18 years old, and Joseph went on to a productive career.

From the beginning, a main goal for Dr. Stanley was to not only identify and follow such precocious students but also to select the best candidates for education acceleration, including early college admission. So was born the idea of using the SAT-Math test for screening junior high school students to find precocious individuals with talent for math and science. The Spencer Foundation provided multiple-year funding to Dr. Stanley beginning in 1971, and the first talent search was in 1972. For that search, junior high school students in the Baltimore area had to be nominated by their math teachers to participate. Actual SAT-Math tests were given in the standard way. In that first search, 396 seventh- and eight-grade students took the SAT-Math. Here are two fascinating results of that first talent search. Twenty-two of the 396 scored at least 660, which was higher than the average Hopkins freshman at the time. And all of these 22 were boys; none of the 173 girls scored over 600.

The male–female ratio has improved considerably over the years, but at the time, this huge disparity was surprising. What about the 22 boys who scored higher than a Hopkins freshman? What were they like? The early data analyses confirmed Terman's results with respect to stereotype. These mathematically precocious students were more physically and emotionally mature than age-peers. One of my first research projects was to give this precocious group some standardized tests of personality. On average, they scored more like college students than their age-peers (Weiss, Haier, & Keating, 1974).

Professor Stanley believed that enriched classes were not as productive as actual college classes, so he helped many of these very talented students go to college early. Over the years, many of the most precocious students did get early admission, usually living at home. And there was no evidence that they suffered any emotional harm from an accelerated program. Like the Termites, many went on to have successful and very productive careers (Bernstein, Lubinski, & Benbow, 2019; Lubinski, Benbow, & Kell, 2014; Makel et al., 2016; McCabe, Lubinski, & Benbow, 2020).

The original talent searches have evolved dramatically and now include many programs for enrichment in addition to early college admission, including summer camps that emphasize math and science experiences. You can find out more details about these programs using Google. Actually, one of the students associated with the talent searches co-founded Google: Sergey Brin. Mark Zuckerberg of Facebook was also identified in a talent search, as was Lady Gaga. Seriously. Look it up.

There are now detailed follow-up studies of thousands of the students who participated in several of the original searches. Follow-up results show that many of these mathematically precocious children, as determined by a single test score when they were in their early teens, became exceptionally successful in terms of occupational and life success (Bernstein et al., 2019; Lubinski et al., 2006; Lubinski et al., 2014; Lubinski, Schmidt, & Benbow, 1996; McCabe et al., 2020; Robertson et al., 2010; Wai, Lubinski, & Benbow, 2005). Figure 1.6 shows professional achievement based on a 25-year follow-up study of the top 1 percent of the original searches that included 2,385 students (Robertson et al., 2010). All these students in the top 1 percent are divided into quartiles – Q1, Q2, Q3, and Q4 – based on their SAT-Math score at age 13. On the x-axis, we have SAT-Math score at age 13. On the y-axis, we have the proportion of the quartile with an outcome such as getting a PhD, a JD, or an MD. Another outcome is having any peer-reviewed publications. Another would be getting a PhD and tenure in a STEM
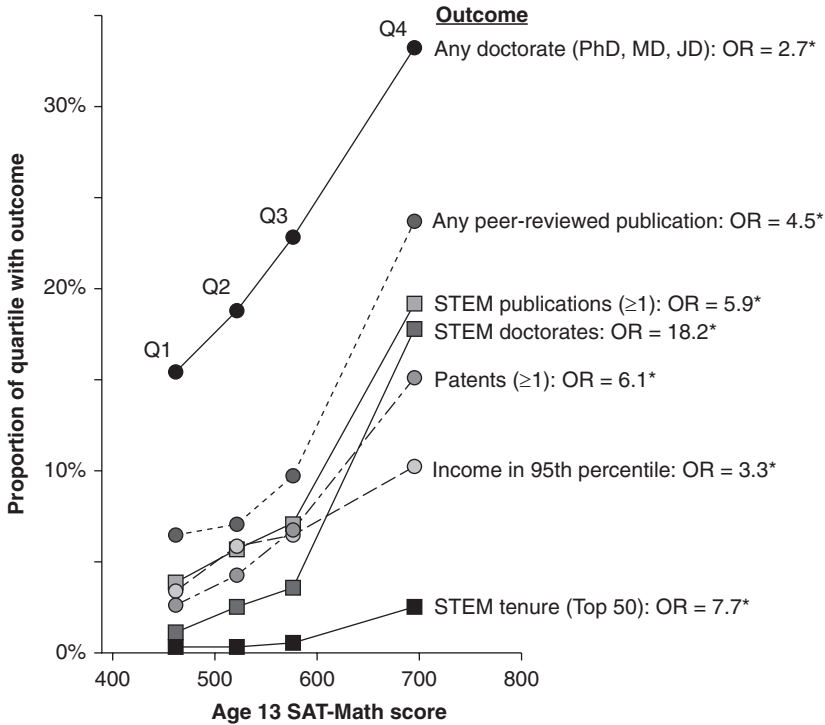
Figure 1.6 SAT-Math scores at age 13 predict adult outcomes of academic success. (Reprinted with permission (Robertson et al., 2010))

field, which includes science, technology, engineering, or math. Patents are another outcome and so is high income defined in the 95th percentile.

What we see in this chart is that for students with age 13 SAT-Math scores in the 400–500 range, which is in the top 1 percent for 13-year-olds but in the lowest quartile 1 for this sample, about 15 percent got a doctorate in any field, and this percentage increases with higher scores. In the top SAT-Math quartile 4, the percentage of advanced degrees is about 35 percent. This is all shown in the line with black dots at the top of the chart. You see this same trend for all the other outcomes.

The OR after each outcome stands for "odds ratio" and compares the top quartile proportion to the bottom quartile for each outcome. For example, the greatest disparity is 18.2 for getting a doctorate in a STEM field. This means the upper quartile within the top 1 percent were 18 times more likely to get a STEM doctorate than the bottom quartile within the top 1 percent. So even in this rarified group of the

top 1 percent, the individuals with the highest scores did the best based on these outcomes.

Remember, a single test taken at age 13 identified these individuals. Again, you can see the predictive validity of this standardized test score is reasonably strong. Clearly, individuals in the top 1 percent of scores obtained in childhood have notable future achievements, but even within this rarified group, the higher the scores, the more likely there will be these kinds of achievements. The longitudinal study of the original talent search participants is continuing, with additional follow-ups conducted by researchers Professor Camilla Benbow and Professor David Lubinski at Vanderbilt University.

Study 3. The third longitudinal study is the Scottish Mental Survey. This was a truly massive project conducted by the Scottish government. All children born in Scotland in 1921 and in 1936 completed intelligence testing at age 11 years and were retested again in old age. This study differed from the other two in that it included virtually *all* children in the country on a test of general intelligence rather than identifying samples of very high scorers (von Stumm & Deary, 2013). The total number of children in the study was about 160,000.

At the time this study began in the 1930s, there was considerable debate around the world about national intelligence and eugenics. This had profoundly evil consequences in Nazi Germany. It's one of the reasons intelligence testing became a negative topic in academia following World War II. But another reason for using intelligence tests in some countries was the desire to open opportunities for better schooling to all social classes by using test scores as an objective evaluation to give all students an opportunity to attend the best schools irrespective of background or wealth. This actually happened in the United Kingdom after the war, and this motivation was important in the development and use of the SAT in the United States (Wooldridge, 2021).

But the Scottish survey was over after the second round of testing in 1936. It only became a longitudinal follow-up study somewhat by accident when the original records were rediscovered in an old storage room. A team of researchers, directed by Professor Ian Deary at the University of Edinburgh, used this database and follow-up evaluations to study the impact of intelligence on aging. Several years ago, Dr. Deary got a new grant from the Scottish government, restored the physical handwritten records as much as possible, and then computerized them all. He also identified 550 original participants who were still living and willing to be retested. So there is now follow-up data. Let's look at two interesting results from the longitudinal analyses:

1. IQ scores were fairly stable over time as demonstrated by showing scores at age 11 correlated to scores at age 80 (r = 0.72) (Deary et al., 2004). The intelligence test used at the beginning of the survey and for follow-up is called the Moray House Test. It gives an IQ score essentially equivalent to the Stanford–Binet or the WAIS. Recall that fluid intelligence decreases with age. Crystallized intelligence is more stable, and the IQ score from the test used in this study combined both fluid and crystallized intelligence. Although not part of this study, it should be noted that different components of IQ might rise and fall at different times across the lifespan (Hartshorne & Germaine 2015).
2. Individuals with higher intelligence scores at age 11 lived longer than their classmates with lower scores, as shown in Figure 1.7 (Batty, Deary, & Gottfredson, 2007; C. Murray et al., 2012; Whalley & Deary, 2001).

The top graph in Figure 1.7 shows the data for women, and the bottom graph shows men. Both show the same trends. On the *x*-axis, we see the ages of participants by decade from age 10 to age 80, and on the *y*-axis, we see the percentage of the group originally tested who are still alive at each age. The data are shown separately for the lowest and the highest quartile based on IQ.

So, for example, in Figure 1.7 let's look at the top graph of women, and let's focus on the data points at the far-right side of the graph (about 80 years old). You can see that more women are alive in the highest IQ quartile, about 70 percent compared to the bottom quartile, where about 45 percent are still alive. This is quite a large difference. And this is true starting around age 20. It's the same for men, but starting later at around age 40 and the trend is not quite as strong. Since the United Kingdom has universal healthcare, differential rates of insurance coverage do not influence these data. But why should IQ be related to longevity? Here is one possible explanation. Before age 11, several factors, both genetic and environmental, may influence IQ, and then higher IQ leads to healthier environments and behaviors and to a possibly better understanding of physician instructions, and these in turn influence age at death. However, there is compelling evidence that a better explanation is that mortality and IQ have genetic influences in common. An estimated 84–95 percent of the variance in the mortality–IQ correlation may be due to genes (Arden et al., 2015).

To recap the evidence from these three classic studies, Terman's project helped popularize the importance of IQ scores and demolished the
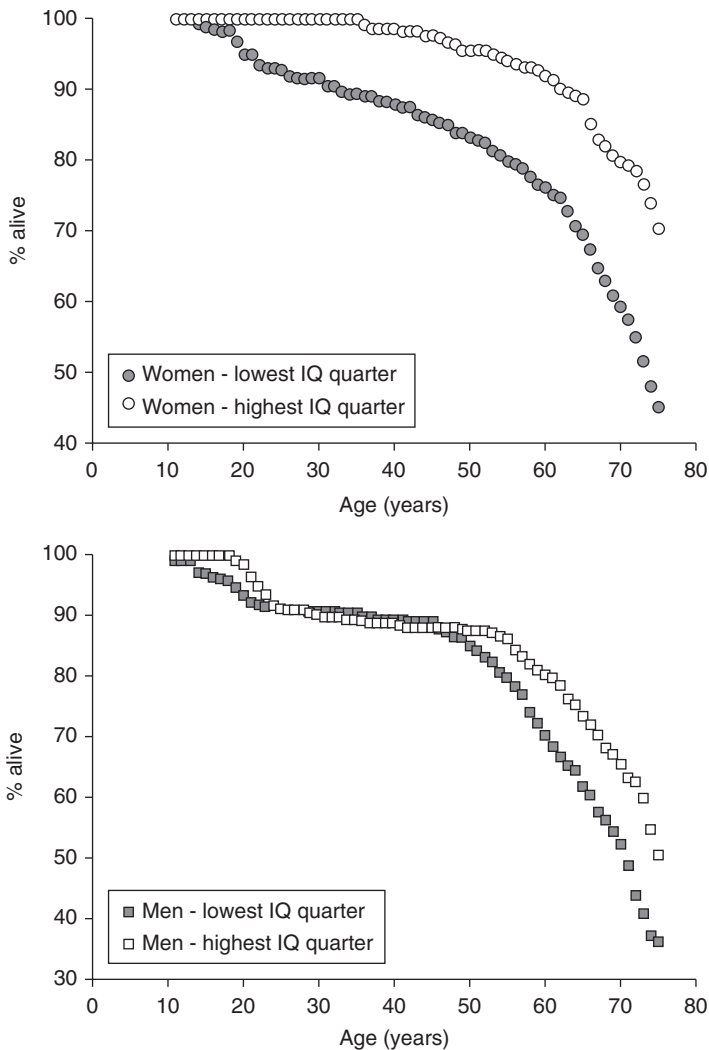
Figure 1.7 Childhood IQ scores predict adult mortality. Note that many more people in the highest IQ group are alive recently compared to the lowest IQ group. (Reprinted with permission (Whalley & & Deary, 2001))

popular but negative stereotype of childhood genius. Gifted student education essentially started with this study. Stanley's project went further and incorporated ways to foster academic achievement in the most gifted and talented students. Deary's analyses of the National Survey data in

Scotland provided new insights about the stability of IQ scores and the importance of general intelligence for a number of social and health outcomes.

These studies provide compelling data that one psychometric test score at an early age predicts many aspects of later life including professional success, income, healthy aging, and even mortality. The bottom line is that it's better to be smart, even if defined by test scores that have meaning only relative to other people.

1.11 Why Do Myths about Intelligence Definitions and Measurement Persist?

Given all this strong empirical evidence that intelligence test scores are meaningful, why does the myth persist that these scores have little if any validity? Here is an informative example. From time to time, a college or university admissions representative will assert that in their institution they find no relationship between GPA and SAT scores. Such observations are virtually always based on a lack of understanding of a basic statistical principle regarding the correlation between two variables. To calculate a correlation between any two variables, there must be a wide range of scores for each variable. At a place like MIT, for example, most students fall in a narrow range of high SAT scores. This is a classic problem of restriction of range. There is little variance among the students, so in this case, the relationship between GPA and SAT scores will not be very strong. Sampling from just the high end or just the low end or just the middle of a distribution restricts range and results in spuriously low or zero correlations. Restriction of range actually accounts for many claims about what intelligence test scores "fail" to predict.[1]

Here's another classic example of an erroneous finding due to restriction of range. In the 1930s, Louis Thurstone challenged Spearman's finding of a *g*-factor (Thurstone, 1938) and proposed an alternative model of "7 Primary Abilities" that he claimed were independent of each other. That is, they were not correlated to each other and there

---

[1] Nonetheless, this is one argument used by critics of testing, who have succeeded in eliminating the SAT and other standardized testing in the admissions process for many colleges and universities. Other arguments include alleged test bias and unequal education opportunities that lead to lower scores. The evidence, however, supports the idea that SATs can provide a fair evaluation that is useful for the admissions process (Wai, Brown, & Chabris, 2019; and see Textbox 1.2).

was no common *g*-factor. There's spatial ability, as measured by tests that require mental rotation of pictures and objects. There's perceptual speed, as measured by tests of finding small differences in pictures as fast as possible. There's number facility, as measured by tests of computation. There's verbal comprehension, as measured by tests of vocabulary. There's word fluency, as measured by tests that require generating as many words as possible for a given category within a time limit. There's memory, as tested by recall for digits and objects. And finally, there is inductive reasoning, as measured by tests of analogies and logic.

However, Thurstone's model was not supported by subsequent research. It turns out that the original research was flawed because the samples he used did not include individuals across the full range of possible scores. That is, the range was restricted, so there was no variance to predict any test from any other. When additional research corrected this problem, tests of the Thurstone "primary" abilities, in fact, were correlated to each other and there was a *g*-factor. Thurstone retracted his original conclusion (Thurstone & Thurstone, 1941). So why include this example from the 1930s in a modern book? As we will see in later chapters, a surprising number of studies still report erroneous findings because of restricted range.

Differences in factor structure among many models based on factor analysis have given some critics the idea that *g* is merely a statistical artifact of factor analysis methodology. We now have hundreds of factor analysis studies of intelligence on hundreds of mental tests completed by tens of thousands of people and using many varieties of factor analysis methods. The bottom line is that there is always a *g*-factor. Here's a key point: *g*-factors derived from different test batteries correlate nearly perfectly with each other as long as each battery has a sufficient number of tests that sample a broad range of mental abilities and the tests are given to people sampled from the wide range of abilities (Johnson et al., 2004; Johnson, te Nijenhuis, & Bouchard, 2008). A study of 180 college students reported that a *g*-factor derived from their performance on a battery of video games correlated highly (0.93) with a *g*-factor extracted from their performance on a battery of cognitive tests (A. M. Quiroga et al., 2015; M. A. Quiroga et al., 2019). Such studies provide strong evidence that *g* is not a statistical artifact, even though its meaning is limited as an interval scale. And, logically, if it were merely an artifact, *g*-scores would not correlate with other measures of the complexity of everyday life, as we noted, nor with genetic and brain parameters, as we detail in subsequent chapters.

Finally, perhaps the major motivation for diminishing the validity of intelligence tests, and other tests of mental abilities including the SAT, is the desire, shared by many, to explain away group differences in average scores as a mere artifact of the tests. In my view, this motivation is misplaced. The causes of average test score differences among groups are not yet clear, but the differences are a major concern in education and other areas. They deserve full attention with the most sophisticated research possible so causes and potential remediation can be developed based on empirical studies. Imaging studies of brain development and intelligence are beginning to address some issues, as detailed in Chapters 3 and 4, and the goal of enhancing intelligence, discussed in Chapters 5 and 6, is something to consider beyond science fiction (Haier, 2021).

Before we get into the brain itself, in Chapter 2 we summarize the overwhelming evidence that intelligence is strongly influenced by genetics and how genes may affect the brain. We also introduce the concept of the epigenetic influences of environmental/social/cultural factors on gene expression, all of which work through biological processes to affect the brain. Altogether, this evidence supports our primary assumption that intelligence is 100 percent biological.

Chapter 1 Summary

- Intelligence can be defined and assessed for scientific research.
- The *g*-factor is a key concept for estimating a person's intelligence compared to other people.
- It is surprising that intelligence is rarely considered explicitly in vigorous debates about why pre-college education appears to be failing many students. The best teachers cannot be expected to attain educational objectives beyond the capabilities of students.
- At least four kinds of study demonstrate the predictive validity of intelligence test scores and the importance of intelligence for academic and life success.
- Intelligence tests are the basis for many important empirical research findings, but going forward, the key problem for assessment is that there is no ratio scale for intelligence, so test scores are meaningful only relative to other people.
- Despite widespread but erroneous beliefs about definition and assessment, neuroscience studies seek to understand the brain processes that underlie intelligence and how they develop.

Review Questions

1. Is a precise definition of intelligence required for scientific research?
2. What is the difference between specific mental abilities that define savants and the *g*-factor?
3. Why is an intelligence test score not like a measure of length, liquid, or weight?
4. What is restricted range and why is it an important concept for intelligence research?
5. What are two myths about intelligence and why do they persist?
6. Why do you suppose this chapter begins with a quote from 1980?

Further Reading

Straight Talk about Mental Tests (Arthur Robert Jensen, 1981). This is a clear examination of all the issues surrounding mental testing, written without jargon by a real expert for students and the general public. It is still a classic, but you may find it only in libraries or from online sellers.

The g-Factor (A. R. Jensen, 1998). This is a more technical and thorough text on all aspects of the *g*-factor. It is considered the classic in the field but it may be hard to find.

IQ in the Meritocracy (Herrnstein, 1973). This controversial book put forth an early argument about how the genetic basis of IQ was stratifying society. The preface is a hair-raising account of the acrimonious climate of the times at Harvard University for unpopular ideas. This book is hard to find but try online sellers.

The Bell Curve (Herrnstein & Murray, 1994). This is possibly the most controversial book about intelligence ever written. It expands arguments first articulated in *IQ in the Meritocracy*. It is loaded with data and possible implications about what intelligence research could mean for public policy.

"The Neuroscience of Human Intelligence Differences" (Deary, Penke, & Johnson, 2010) This is a concise review article written by longtime intelligence researchers.

Coming Apart (C. A. Murray, 2013). This book advances Charles Murray's analysis of the implications of society segregating by levels of intelligence. It is prescient about today's political and social divisions.

The Nature of Human Intelligence (R. J. Sternberg, 2018). This edited book includes contributions from 19 of the most cited living intelligence researchers. Each chapter covers a different topic, including many not covered in this book, which is focused exclusively on neuroscience approaches.

The Aristocracy of Talent: How Meritocracy Made the Modern World (Wooldridge, 2021). This is a scholarly discussion of the social implications of intelligence and a broad defense of meritocracy.

The Cambridge Handbook of Intelligence and Cognitive Neuroscience (Barbey, Karama, & Haier, 2021). This is an edited book with chapters written by experts in cognitive neuroscience; many chapters are written for advanced students.

The Science of Human Intelligence (Haier & Colom, 2023). This is a textbook for students that covers a wide range of intelligence topics. It is the second edition of Hunt's classic textbook, *Human Intelligence* (Hunt, 2011).