CHAPTER 1

The Issue of Intelligence

So it is that gods do not give all men the gifts of grace . . . neither good looks nor intelligence nor eloquence.

Homer, *The Odyssey*

There's many a man has more hair than wit.

Shakespeare, *Comedy of Errors*, act 2, scene 3

1.1. The Idea of Intelligence

Homer and Shakespeare lived in very different times, more than two thousand years apart, but they both captured the same idea; we are not all equally intelligent. I suspect that anyone who has failed to notice this is somewhat out of touch with the species. However, we cannot simply sort people into the "intelligent" and the "not-so-intelligent." Homer observed that few people have great gifts. Shakespeare, more pithily, observed that all too many of us do not do terribly well at problem solving. Most of us, though, fall in between Homer's desire for eloquence and Shakespeare's worry about lack of wit.

In this book I will talk about the nature of intelligence, its causes, who has it, and how it is used. I will do so without the eloquence of Homer and Shakespeare. I will take a scientific view. Modern psychology has a great deal to say about intelligence, and somehow a great deal that has been said has been seriously misunderstood. The popular media sometimes report that the psychologists who study intelligence say almost the opposite of what the psychologists actually said.[1]

There is a reason for this. The study of intelligence is not an isolated academic topic; our intelligence has social consequences. We want our leaders to be intelligent, and exhibit concern if we think they are not. There were politically motivated attacks on the intelligence of Presidents Lincoln, Truman, Harding, and Ford. Serious concerns about mental competence

1 See Tannenbaum (1996) for a discussion of this issue and references to earlier discussions of the topic. Gottfredson (2005) provides a spirited discussion of how failing to consider the implications of psychological research on intelligence can be costly to society.

were raised about Wilson and Eisenhower, following strokes, and Reagan, due to early symptoms of Alzheimer's disease. Lincoln and Truman, who received the most vicious attacks, are now considered two of our finest presidents. Eisenhower recovered to function well; Wilson did not. The effect of Reagan's illness upon his second term is still a matter of debate.

Concerns about intelligence are not confined to concerns about our leaders. Our school systems use cognitive tests to stream high school students into different programs. Colleges use cognitive tests to screen applicants for higher education programs. These tests are never called "intelligence tests," but they correlate highly with them.

Testing is not confined to the educational system. Volunteers for the United States military services must obtain passing scores on a test of general mental competence, the Armed Forces Qualifying Test (AFQT). Similar tests are used in many other countries. Toward the bottom end of the scale there are a variety of special assistance programs for people who simply do not have the cognitive competence to cope with the complexities of the modern world. Low intelligence test scores can be offered as evidence of diminished mental capacity during the penalty phase of a criminal trial.

While there is broad agreement that some people are smarter than others, things become more complex when we try to be precise. I think that every knowledgeable person would agree that Albert Einstein and Thomas Jefferson were both highly intelligent. Who was the more intelligent? That is hard to say; they were brilliant in different ways at different times. It would be easy to find other examples of the same point. There are clearly varieties of cognitive skill, especially at the top. As a result some modern observers have concluded that there is no single dimension of intelligence.

This idea is not new. In the sixteenth century the Spanish physician/philosopher Juan Huarte de San Juan[2] drew a remarkably cogent picture of individual differences in human thought. Huarte believed that when people attack problems some will use their imaginations to envisage how a solution might work out, while others will rely on their memories of solutions that have worked in the past. Huarte also defined "understanding" (*entendimiento*) as a separate capacity, implying that one can be bright without having a good understanding of a situation. Huarte's distinction between problem solving by imagination or by memory is mirrored in contemporary theories that distinguish between the ability to do abstract reasoning and the ability to apply previously learned solution methods.[3] Robert Sternberg, a prolific modern writer on intelligence, has emphasized the distinction between analytic intelligence and the ability to understand complex social situations.[4]

Huarte anticipated another modern idea, the need to have a biological explanation for intelligence. Huarte offered a theory based on the sixteenth-century notion that the body is governed by four "humors" – blood, bile, black bile, and phlegm. This theory of biology has long since been discarded. The idea that there should be a biological explanation for individual variations in cognition has been retained. One of the most active areas of modern intelligence research deals with the relation between intelligent behavior and the brain.

Let us leap from the sixteenth century to the nineteenth, and to one of the most colorful characters in the history of science, the Victorian physician, mathematician, and explorer Sir Francis Galton. Galton explored in Africa, made major contributions to the development of statistics, and conducted research in psychology, most noticeably on intelligence. He wholeheartedly endorsed the theory of evolution proposed by his cousin Charles Darwin. Galton believed that human intelligence was largely inherited. He also maintained that intelligence was one manifestation of a person's overall constitutional fitness. Therefore,

2 Huarte de San Juan, 1575/1991.

3 Horn & Noll, 1994.
4 Sternberg, 2003.

it should be possible to learn something about a person's intelligence by examining his or her physique, including brain size, and by determining the efficiency of the person's nervous system by doing such things as recording the speed with which he or she reacted to a signal to strike a bag. These ideas are alive, in much expanded form, today.

The next step was taken at the start of the twentieth century, when the Frenchman Alfred Binet developed the first intelligence tests to be used in schools. Testing has dominated the study of intelligence since then, so we need to look more closely at the idea.

1.2. Testing

If you want to go beyond saying "people are different" you have to offer some way of measuring those differences. There is an imperative to develop such measures if a society wants to assign different roles to different people, based on their personal characteristics. Not all societies do this, all the time. There was no intelligence test for the ruler in the hereditary kingships of medieval Europe. The Hindu caste system pre-assigned people to social roles, based upon their birth. It is notable, though, that both these societies experienced a good deal of conflict due to their restricting people's social roles.[5]

In village-based societies personal knowledge of individuals plays a major role in assigning people to jobs. When American pioneers began to move into the northern plains states Sitting Bull, the paramount chief of the Lakota (Sioux) Indians, selected Crazy Horse to be war leader from among people whom he knew personally.[6] That technique does not work in today's large societies, where there are many positions to be filled in both government and industries. Leaders cannot possibly know all their subordinates, let alone the subordinates' subordinates. Our society requires formal machinery for selecting candidates either into employment, directly, or into educational systems that serve as channels to future employment.

Many societies solve this problem by an elaborate form of recommendations. A boy or (historically less often) a girl who is thought to be talented is sent off for training and/or apprenticeships. While the details have been lost, this appears to be the way the ancient Egyptians selected children to be trained as scribes. It was also the way in which second, third, and fourth sons were recruited for the priesthood (or the army) in medieval Europe. The person was not needed at home, and somebody had connections enough to start them on a career. The use of connections is certainly not unknown in modern times. But we do rely on another method of personnel selection: testing.

There have been many objections to testing. In evaluating them it is well to keep one thing in mind. Society needs a mechanism for personnel selection. Not everyone can have whatever they want. Students have to be selected, jobs have to be filled, and when behavioral problems arise, mental competence must be assessed. If you do not like testing, what is your alternative?

1.2.1. *Testing Before Psychology*

Modern psychologists did not invent testing. In the days when the Chinese emperor claimed to rule "the Earth, the Moon, and three quarters of the Sun" an elaborate series of local, regional, and nationwide tests was used to select officers for the imperial bureaucracy. Candidates had to write traditional poetry and to explain the importance of fearing the will of heaven and knowing the words of the sages. Evidently it was assumed that a person who could do these things could collect the imperial taxes or be an ambassador to the Mongols.

5 During the Hundred Years' War between England and France (roughly 1350–1450) the French court was disrupted when Charles the Foolish inherited the throne. He probably suffered from a bipolar psychosis. In India the Sikhs were formed largely as a protest against the rigid social structure enforced by the Hindu caste system.

6 He did well. Crazy Horse defeated General Custer at the Battle of the Little Big Horn. This was the most stunning Native American victory in the history of the western expansion.

Some centuries later the British Empire emulated the Chinese Empire. Until after World War II career positions in both the Indian empire and the British upper bureaucracy were filled largely from the ranks of people who had read history, classics, and occasionally economics at the elite universities, especially Oxford and Cambridge. The British assumed that someone who could do well on oral and written examinations of the writings of Horace and the wars of Caesar would have the ability to administer India or to ferret out secrets while on Her Majesty's Service.[7]

Such techniques of personnel selection may sound quaint to us, but, on the whole, they worked. The Chinese bureaucracy held the empire together in a way that no succession of able emperors could ever have done alone. Properly educated British gentlemen administered India reasonably well for two hundred years. The classics program at Cambridge University produced at least four remarkably effective twentieth-century spies. Unfortunately, they spied against Britain for the Soviet Union, but that is a motivational rather than a cognitive aberration.[8] They were good at what they did.

Why did these exercises in testing apparently irrelevant knowledge do a reasonably good job of selecting people able to run very large, complex empires? Or for that matter, able to fool a modern counterintelligence agency?

What the British and the Chinese had stumbled on, and what we today attempt to evaluate, was a collection of mental traits that, collectively, we call *intelligence*. These traits define individual differences in skills that have broad application in many settings. One of the most important aspects of intelligence is an ability to learn. You demonstrate this by showing that after exposure to knowledge you have learned something. The skills needed to learn the wisdom of Confucius or the philosophical ideas of Socrates are not exactly the skills you need to run an empire, *but there is an overlap*. For that matter, the skills needed to do well on a college entrance test are not exactly the skills you need to acquire a bachelor's degree, *but there is an overlap*. That is why both the classic and the modern testing systems work. It is also why they work imperfectly.

1.2.2. *Alfred Binet Invents Modern Intelligence Testing*

At the start of the twentieth century the French Ministry of Education had a problem. The idea of universal public education had been accepted, but the schools did not seem to work for some students. How did this problem arise?

France, like all modern democracies, was (and is) committed to providing public education for all its citizens, so that all children have an opportunity to compete for desirable positions in society. This goal is not easy to achieve.

Modern schooling is an historically unusual form of education. Before 1800 most humans were educated "on the job" – observing and then helping adults, and serving as apprentices. Universal education, the requirement that every child learn by practicing seemingly esoteric exercises in a setting divorced from everyday life, is a late nineteenth-/early twentieth-century idea. By 1900 it was apparent to educators that some children have a great deal of trouble learning in this manner. The French educational administration needed to have a way of identifying such children, so that they could either be dropped from the system or channeled into an educational program more suited to their capabilities.

The fact that France was a democracy imposed an added constraint. French educators needed an objective method, in preference to the subjective impressions of "persons in authority," the teachers and principals. In an authoritarian regime there is no

7 Gardner, Kornhaber, & Wake, 1996, pp. 12–16.

8 Anthony Blunt, George Burgess, David MacLean, and Kim Philby. Burgess, MacLean, and Philby fled to the Soviet Union when they were about to be exposed. Blunt stayed in England, undiscovered, and became a distinguished art historian. His espionage role, which seems to have lasted through the 1940s and 50s, was not publicly revealed until 1979.

need for such a method; if the authorities don't like you, you're out.

So from the very first, testing was embedded in the society that required it.[9]

In order to meet this challenge the Education Ministry hired Alfred Binet, who had worked for a while with Galton. Binet began by making two important assumptions. The first was that mental competence increases over the childhood years. The typical six-year-old can solve problems a four-year-old cannot; a four-year-old can solve problems a two-year-old cannot, and so on, at least from birth to the late teenage years. Therefore, it makes sense to talk about *mental age* (MA) – the level of mental competence at which a child is operating.

Binet took a pragmatic approach to the measurement of mental age. He asked experienced teachers what sorts of problems children could solve at different ages. Once he had a set of problems typical of the ones children could solve at age six, seven, eight, and so on, he could determine a person's mental age by finding the most difficult problems that a child could solve. Mental age could then be compared to chronological age (CA), to determine whether a child has been performing below, at, or above the cognitive level that would be expected.

Binet then made his second, more arguable, assumption. He assumed that a child's relative standing in mental development, compared to his age group, will remain fairly constant as the child grows up. If Sammy and Tommy are both six, but Sammy has a mental age of eight and Tommy one of five, Binet assumed that four years later, when they are both ten, Sammy would have a mental age higher than ten, and Tommy a mental age lower than ten. Therefore, it follows that if you test children on entrance to school (age six or seven), and you find that some are markedly behind (i.e., have mental ages in the three-to-four range), those children are likely to be behind their classmates at all ages, and therefore are candidates for removal from the normal school program. That is what the French education system wanted to know.[10]

The Education Ministry accepted Binet's argument. The modern era of intelligence testing had been born.

1.2.3. *The Intelligence Quotient (IQ)*

Binet did not use the term *Intelligence Quotient* (IQ) because the concept of mental age was sufficient for classifying children who were entering school. As mental testing expanded to the evaluation of adolescents and adults, however, there was a need for a measure of intelligence that did not depend upon mental age. Accordingly the intelligence quotient (IQ) was developed. There have been changes in the definition and use of the term since its introduction. The details are provided in panel 1.1. Here we proceed directly to the modern use of the term.

The term IQ is used in two ways, which I will call the *narrow* and *broad* uses of the term.

The narrow definition of IQ is a score on an intelligence test, developed according to a scoring protocol where "average" intelligence, that is, the median level of performance on an intelligence test, receives a score of 100, and other scores are assigned so that the scores are distributed normally about 100, with a standard deviation of 15. Some of the implications are that:

1. Approximately two-thirds of all scores lie between 85 and 115.
2. Five percent (1/20) of all scores are above 125, and one percent (1/100) are above 135. Similarly, five percent are below 75 and one percent below 65.

9 The contrast between the French and the Chinese and British imperial systems is informative. The Chinese and British systems were designed to select a sufficient number of qualified candidates for government functions. So long as the supply of young officers and bureaucrats was adequate, there was little concern that the system might have shut out potential candidates. The French testing program was designed to staff society, without favoring some citizens over others. Any government has to solve its staffing problems. Only democracies have to justify the staffing system to the citizens.

10 Binet & Simon, 1905.

Panel 1.1. The Intelligence Quotient (IQ)

Mental age is inadequate as a means of comparing the intelligence of two children of different chronological ages (CA). Suppose a six-year-old and a ten-year-old both have a mental age (MA) of eight. Cognitively, they are likely to be very different individuals, for one is developing rapidly and the other is developing slowly. The need for a measure of mental development that is independent of chronological age led to the concept of the *Intelligence Quotient*, IQ, which was originally defined as the ratio of mental age to chronological age, multiplied by 100.*

$$IQ = 100 \cdot \frac{MA}{CA}. \tag{1.1}$$

To illustrate, a ten-year-old who can solve problems at the level of difficulty expected of a twelve-year-old would have MA = 12, CA = 10, IQ = 120. An IQ of 100 indicates that the child's cognitive development is proceeding on schedule, an IQ above 100 indicates acceleration, and an IQ below 100 indicates slowed development. In the case of our hypothetical six-year-old and ten-year-old, both of whom have a mental age of eight, the first child would have an IQ of 133 and the second an IQ of 80. In modern educational terms the first child might be considered for an accelerated program, while if the second's IQ score were accompanied by difficulties with schoolwork he or she would be a candidate for a special education program.

This method of calculating IQ will not work with adults, because intelligence does not increase linearly with age past childhood. A man of sixty whose mental powers are equal to those expected of a forty-year-old would not be considered a case of retarded development! Therefore, the IQ ratio just described has been replaced by a measure based on the notion that IQ should reflect a person's relative standing within his or her own age group.

Intelligence tests are *standardized* by giving them to a relatively large sample of people chosen to be representative of the population for whom the test is intended. In the case of a test intended for broad use, such as establishing the mental competency of adults, this is essentially the entire population of the country, so attempts are made to obtain a large representative sample of the United States, the United Kingdom (for the British version), Spain (for the Spanish version), and so forth.† The sample should be sufficiently large that a distribution of scores can be obtained for different age groups.

The *raw* test score is based on the number of items correctly answered and/or the difficulty level of the item. (The scoring algorithm varies somewhat with the test, as discussed in Chapter 2). The IQ score is derived from the raw score in the following way.

Let Y be a raw score for a person in a particular age group, and let M be the mean and S the standard deviation for all scores in the reference group.‡ The corresponding *standard score* is

$$z = \frac{Y - M}{S}. \tag{1.2}$$

The IQ score is derived from this by the conversion

$$IQ = 15 \cdot z + 100. \tag{1.3}$$

If the raw scores were normally distributed, then the resulting IQ scores will be normally distributed with a mean of 100 and a standard deviation of 15. The graphic depiction of this distribution is the famous "Bell Curve." The Bell Curve for IQ distributions is shown in Figure 1.1.

Why not record scores in the standard score format? Statisticians, psychometricians, and research workers would probably prefer to do this. The advantages

have to be weighed against two countervailing trends: tradition and public relations. IQ scores were introduced almost a century ago; people are used to them. Standard scoring is a bit esoteric for the nonstatistician. For a statistician, having a scale with a mean of zero and a standard deviation of 1 is a convenience, nothing more. If a (nonstatistical) parent were to be told that their child had scored a zero on an intelligence test, they might interpret this as a claim that their pride and joy had the intelligence of a rock! Half the people who took the test would receive negative scores, which could lead to all sorts of misunderstanding. Appearances can be important.

* The concept of IQ was developed by the German psychologist William Stern, not by Binet.

† This procedure contains the implicit assumption that the distributions of intelligence are the same across populations. Thus if you consider a score of 100 on the Spanish form of the test to indicate the same thing as a score of 100 on the American version of the test, you are implicitly assuming that Americans and Spaniards have equivalent intelligences, on a population basis. Such assumptions have been vigorously debated. The controversies are discussed in detail in Chapter 11.

‡ The standard deviation is a measure of the amount of variation in a population. More details are given in Chapter 2.

Thus IQ, in the narrow sense, is a score indicating a person's relative performance on an intelligence test, compared to the performance of people in an appropriately chosen comparison group. This does not completely clarify the matter, because there can be debate about what counts as an intelligence test. This matter is also discussed on page 8. I will attempt to be clear about how the term is being used in various contexts.

In the broad sense the term IQ is used as a synonym for intelligence, that is, as a shorthand term for individual differences in cognition. This can lead to confusion, so I shall attempt to use IQ only in the sense of a test score. The term *intelligence* will be used to refer to the broader concept of individual differences in mental ability. In my usage, a person who has high intelligence will probably have a high IQ score, but the distinction between the two is important.

In interpreting IQ scores it is often useful to think of percentiles, which indicate the percentage of people in the referent group whose scores are below a certain level. What that level is will be determined by the IQ score and by the properties of the Bell (normal) Curve itself. Table 1.1 gives some important reference scores. The properties of these scores follow from the assumption that IQ scores will follow the normal distribution, which is illustrated in Figure 1.1.

As a result, in terms of the modern scoring, if someone says that their child has an IQ of, say, 120, this does not mean that the child's mental age is 20% higher than his chronological age. It means that the child has a test score in the top 9% of test scores at the child's age.

Why can we assume that IQ scores are distributed normally? The answer is simple. IQ tests (and many other tests) are constructed by choosing appropriate numbers of easy, intermediate, and hard items, so that the total scores will be normally distributed in the population for which the test was intended. The Bell Curve is an artifact of the way the test is constructed! There is no definition of IQ independent of the tests themselves. This contrasts with a variable like height, which is defined independently of yard sticks or meter sticks. Height happens to be distributed approximately normally, within the populations of adult males and females. The distribution of height is a fact of nature. The fact that IQ test scores are normally distributed is an outcome of the test construction procedure. Nevertheless, it is a reasonable thing to do. Why?

IQ scores are used to describe people *relative to each other*. They are also used to

Table 1.1. The distributions of standard and IQ scores in terms of the percentage of people above or below selected scores. An IQ of 65 would, if accompanied by other indications of mental incompetence, be cause for considering a person mentally disabled. If IQ is distributed normally, about one percent of all people have IQ scores this low. Average IQ is, by definition, 100. Approximately half of all scores lie between 90 and 110. About 16% of the scores lie above 115, 2% above 130, and 1% above 135. *MENSA*, an organization whose members have high IQ scores, defines the *4 sigma* group, people with IQs over 160. (*Sigma* is a term frequently used to refer to the standard deviation.) This level of score would be expected three times in every 100,000 observations.

Standard Score (z)	*IQ Score*	*% Below*	*% Above*
−2.33	65	0.982	99.018
−2.00	70	2.275	97.725
−1.00	85	15.866	84.134
−.67	90	25.249	74.751
0.00	100	50.000	50.000
.67	110	74.751	25.249
1.00	115	84.134	15.866
2.00	130	97.725	2.275
2.33	135	99.018	0.982
3.00	145	99.865	0.135
4.00	160	99.997	0.003

make predictions and to indicate associations, as in predicting a student's likely academic progress or investigating the association between intelligence and income. There are technical reasons for wanting to deal with normally distributed scores when we apply the statistical methods used for making predictions and analyzing associations.

There is another, less technical reason for requiring that IQ scores be normally distributed. Many other human qualities that can be measured on scales with physical interpretations, like height and weight, are distributed normally. It seemed to many of the early researchers that if we could measure intelligence in some physical manner, such as measuring the efficiency of the nervous system, these measures would probably turn out to be normally distributed. Therefore, it seemed appropriate to require that IQ scores be normally distributed.

In the late nineteenth and early twentieth century this reasoning seemed compelling, because the normal distribution itself was regarded (almost mystically) as a Law of Nature. Today we are a bit more skeptical, but there is still a good argument for assuming normality. If a person's intelligence is due to a large number of independent causes, each of which has a small effect, intelligence would be distributed normally across the population.

A certain amount of the confusion between the broad and narrow senses of IQ is due to the way in which cognitive tests are described. Some tests are explicitly marketed as intelligence tests. But because the term IQ, and sometimes even the term *intelligence*, have acquired a bad taste in certain circles, many tests of cognitive skills are not marketed as intelligence tests, even though these tests are highly correlated with tests that *are* marketed as intelligence tests! For instance, in a widely read and highly controversial report, Richard Herrnstein and Charles Murray used the Armed

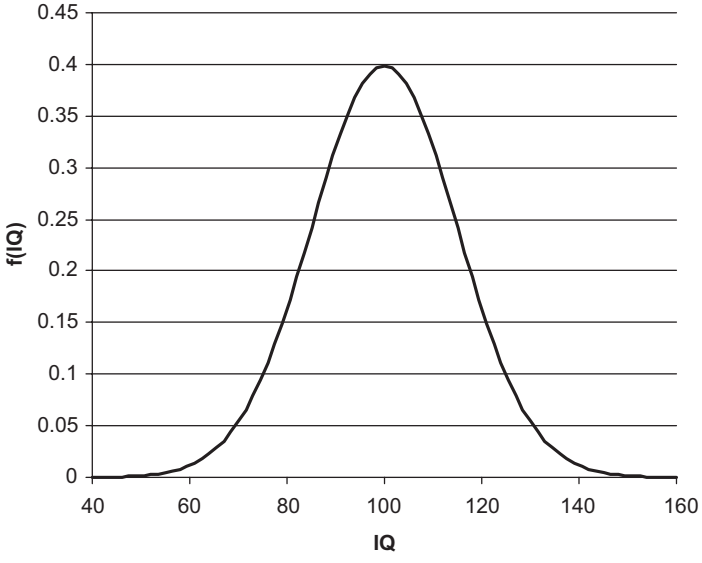

Figure 1.1. The "Bell Curve" for IQ. The area under this curve represents 100% of the population. The area under the curve and to the left of a given IQ value (on the abcissa) represents the fraction of people in a population who have IQs lower than the indicated IQ. Conversely, the area to the right indicates the fraction of people who have this IQ or a higher one. For example, 50% of the area under the curve lies to the left of IQ = 100, indicating that half the population has an IQ of less than 100. Nine percent of the area under the curve lies to the right of 120, indicating that only nine percent of all people have IQs of 120 or higher. The Bell Curve for IQ scores is a special example of the normal, or Gaussian, distribution. At the extremes, the curve never quite touches the abscissa ("x axis"), but this cannot be shown on the graph.

Services Qualifying Test (AFQT) as a measure of intelligence, and treated AFQT and IQ scores as being virtually synonymous.[11] The US Department of Defense never refers to the AFQT as an intelligence test. Similar confusions arise with the SAT. Many research projects have used SAT scores as a measure of intelligence, although the test's publisher, the Educational Testing Service, does not describe it as an intelligence test.

There is great controversy over whether or not IQ scores should be treated as real indicators of mental ability. Panel 1.2 presents a historical debate that took place in the 1920s, but in many ways foreshadowed contemporary arguments. I shall come down squarely in the middle of the controversy. I will argue that the scores certainly do mean something, but they may not mean as much as some enthusiasts claim.

1.2.4. *What Binet Discovered: "Drop in from the Sky" Testing Works*

Let us take a closer look at what Binet assumed and what he found.

Binet's assumption that mental competence increases as children grow older is certainly correct. Mental competence may decrease in old age, but that is another story, and was of no concern to Binet. He was also correct that there are marked individual differences in the rate at which mental competence increases.

His second assumption, that relative standings remain constant as children age, is true on the whole, but there are exceptions. As a toddler, Albert Einstein was a relatively

11 Herrnstein & Murray, 1994.

Panel 1.2. Defining Intelligence: The Debate between Mr. Lippmann and Professor Boring

In science clarity of definition is essential, for good definitions make clear what the important questions are. The study of intelligence has been plagued by a lack of precise definitions. The debate between Lippmann and Boring, early in the twentieth century, shows how the failure to define terms introduced confusions that continue to this day.

Following the use of tests in World War I, intelligence testing became a growth industry. So, inevitably, it attracted the attention of learned commentators – people who, if there had been TV in those days, would have appeared as talking heads on the Sunday morning pundit shows. One of the most respected of these commentators, Walter Lippmann, did not at all like the new technology. He was particularly incensed by a claim, based on analyses of the Army Alpha data, that the average American had a mental age of fourteen. In Lippmann's own words:

> *The intelligence test, then, is an instrument for classifying a group of people, rather than "a measure of intelligence." People are classified within a group according to their success in solving problems which may or may not be tests of intelligence. They are classified according to the performance of some Californians in the years 1910 to about 1916 with Mr. Terman's notion of the problems that reveal intelligence. They are not classified according to their ability in dealing with the problems of real life that call for intelligence.*
>
> *(Lippmann, 1922a)*

Lippmann argued that the test developers had produced a barrage of statistics that had the trappings of science, but were not scientific. The tests themselves were based on hunches by people such as Terman that this or that behavior indicated intelligence, rather than on any scientific theory of what constituted intelligence. Lippmann also doubted that a classification of people on the basis of test scores would map onto a classification according to their ability to "deal with problems of real life that call for intelligence."

Academia responded. The Harvard professor E. G. Boring* clarified the matter by asserting:

> *Intelligence is what the intelligence tests test.*
>
> *(Boring, 1923)*

So there!

The exchange between Lippmann and Boring foreshadowed a debate that is active today. Should a test be developed inductively, by the pragmatic procedure of identifying people who are believed to vary in intelligence, seeing what behaviors distinguish those who are intelligent from those who are not, and then incorporating these behaviors into a test? Or should a test be based on a theory of how individual differences in cognitive power come to arise?

The answer is not a simple one. To see why, consider the following analogy. I think that everyone would agree that people differ in their physical fitness. But what is physical fitness? You could take the approach of an athletic coach, and ask people to run, jump, throw weights, and so forth in order to determine physical fitness. Alternatively, you could take a medical approach. Physical fitness depends on muscular adequacy, reaction, and the ability of the heart and lungs to provide fuel to the muscles. So let us take measures of cardio-pulmonary capacity and construct tests of the strength of isolated muscle groups and the speed of neural impulses.

Binet, Terman, and their many successors took the coach's approach. Lippmann seems to have wanted a more theoretically justified approach, although he