CHAPTER 1

A History of Intelligence Assessment
The Unfinished Tapestry

John D. Wasserman

When our intelligence scales have become more accurate and the laws governing IQ changes have been more definitively established it will then be possible to say that there is nothing about an individual as important as his IQ, except possibly his morals; that the greatest educational problem is to determine the kind of education best suited to each IQ level; that the first concern of a nation should be the average IQ of its citizens, and the eugenic and dysgenic influences which are capable of raising or lowering that level; that the great test problem of democracy is how to adjust itself to the large IQ differences which can be demonstrated to exist among the members of any race or nationality group.

—LEWIS M. TERMAN (1922b)

This bold statement by the author of the first Stanford–Binet intelligence scale captures much of both the promise and the controversy that have historically surrounded, and that *still* surround, the assessment of intelligence. Intelligence tests and their applications have been associated with some of the very best and very worst human behaviors. On the one hand, intelligence assessment can provide a meaningful basis for understanding the strengths and weaknesses of misunderstood children, adolescents, or adults—thereby providing data that can be used to design and implement interventions to help people reach their potential more effectively. On the other hand, intelligence assessment can be used to segregate and label people—treating their future as a fixed outcome, an unchangeable fate. The history of forced sterilizations of individuals with intellectual disabilities in the United States and many other countries is a tragic example of how intelligence tests may be misused, exceeded only by the systematic extermination of intellectually disabled individuals in Nazi Germany (e.g., Friedlander, 1995). The topic of intelligence and its assessment deservedly elicits many strong feelings.

Intelligence is arguably the most researched topic in the history of psychology, and the concept of *general intelligence* has been described as “one of the most central phenomena in all of behavioral science, with broad explanatory powers” (Jensen, 1998, p. xii). Still, social, legal, and political forces have in some instances excluded intelligence test results from important types of educational and personnel decision-making processes. Tangible advances in assessment practices have been slow and episodic. Following Alfred Binet’s initial successes, the beginning of the 20th century saw an accelerated pace of small- and large-scale applied intelligence testing, but many anticipated educational and occupational benefits were never realized. Buros (1977) considered 1927 as the “ banner year” in which “the testing movement reached maturity” (p. 9). The middle of the century saw only incremental gains in testing, such as electronic scoring, analysis, and reporting of test results, but with comparatively “little progress” (Buros, 1977,

p. 10) and more than a little "stagnation" (Carroll, 1978, p. 93). A landmark quantitative review of factor-analytic investigations near the end of the 20th century (i.e., Carroll, 1993) stimulated a new school of thinking about intelligence assessment, but the story remains unfinished. In the United States, federal educational reforms and civil rights legislation have had pronounced effects upon the use of intelligence tests in education. It is possible to see the history of intelligence assessment as an unfinished tapestry depicting the rich saga of a developing discipline, with recurrent characters interwoven through different narratives, as well as more than a few loose and unresolved thematic threads.

In this chapter, the origins of intelligence assessment are recounted, with an emphasis on milestone events and seminal individuals. Thematic strands present from the early days are traced, including some that were resolved and some that remain unresolved. An effort has been made whenever possible to provide samples of primary source material. Finally, although we all tend to view history through the lens of our own experiences, it is helpful to appreciate the sociocultural context, institutional traditions, and professional *Zeitgeist* associated with historical events, as well as the experiences and personal motivations that may have driven the ideas and behaviors of historical figures.

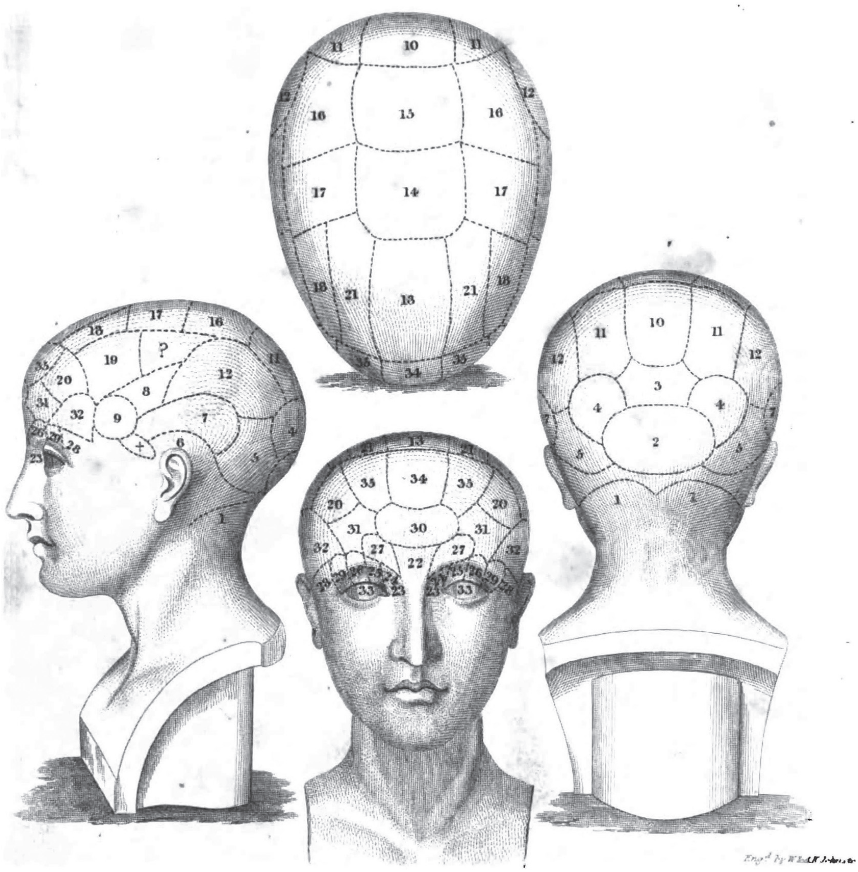
PSEUDOSCIENTIFIC ANTECEDENTS: PHRENOLOGY IN THE 19TH CENTURY

The first science purporting to be a "true science of mind" that could measure mental qualities and functions was *cranioscopy*, introduced at the beginning of the 19th century by Franz Joseph Gall, and later renamed *phrenology* by Gall's associate, Johann Gaspar Spurzheim. Gall (1758–1828) was a Viennese physician and neuroanatomist, and Spurzheim (1776–1832) was a physician and colleague who would ultimately be responsible for the widespread dissemination of phrenology. But it would be a Scotsman, George Combe (1788–1858)—who developed and published a two-volume system of phrenology in 1824, as well as launching a phrenology journal with his brother—who would prove most instrumental in the popularization of phrenology. Combe's system appears in Figure 1.1. He also wrote the immensely successful book *The Constitution of Man*, which advanced the idea that all the laws of nature were in harmony with one another, and that people could best fulfill God's will and obtain the greatest happiness by discovering these laws and obeying them. The book went through eight editions and sold approximately 350,000 copies between 1828 and 1900.

The basic tenets of phrenology can be summarized easily. In a letter to a Viennese official, Gall (1798/1857) asserted that the brain was the organ of the mind, that the mind could be reduced to a number of faculties, that the faculties were innate, that the faculties were located in distinct and particular organs of the brain, that the surface of the skull was determined by the external form of the brain, and that phrenologists could judge the development of individual faculties merely by examining the form of the skull. A well-developed faculty was thought to have a large cerebral organ that corresponded to a cranial protuberance. Gall originally described and localized 27 distinct faculties; Spurzheim (1815) increased the list to 32 faculties; Combe (1853) further expanded the list to 35; and others expanded the list to 43 (e.g., Payne, 1920).

Gall and Spurzheim traveled through Europe promoting phrenology, which Gall advocated as a science and Spurzheim as a way to reform education, religion, and penology. It quickly became popular in the United Kingdom, and Spurzheim came to the United States in 1832 to promote phrenology to a scientific community that was already quite familiar with it. By the time Combe conducted his 1839 American phrenology lecture tour, audiences averaged over 500 across each of the 16 lectures (Walsh, 1976). A satirical depiction of a phrenological examination from about the same time appears in Figure 1.2.

Gall and Spurzheim are today credited with recognizing the significance of gray matter as the source of nerve fibers; most importantly, they are credited with introducing the neuroscientific concept of functional localization in the cerebral cortex (Simpson, 2005). Dallenbach (1915) provides evidence that they should be credited with the terms *mental functions* and *faculties*. British philosopher and critic G. H. Lewes (1867) went a step further, asserting that Gall laid the groundwork for psychology as a science rather than philosophy: "Gall rescued the problem of mental functions from Metaphysics and made it one of Biology" (p. 407). Even so, there is a long history of disparaging efforts to localize mental functions in specific regions in the brain by calling them a new "phrenology" (e.g., Franz, 1912; Fuster, 2008, p. 346).



Names of the Phrenological Organs

REFERRING TO THE FIGURES INDICATING THEIR RELATIVE POSITIONS.

AFFECTIVE						INTELLECTUAL					
I. PROPENSITIES			II. SENTIMENTS			I. PERCEPTIVE			II. REFLECTIVE		
1	Amativeness	Page 116	10	Self-esteem	231	22	Individuality	380	34	Comparison	460
2	Philoprogenitiveness	121	11	Love of approbation	243	23	Form	385	35	Causality	474
3	Concentrativeness	134	12	Cautiousness	252	24	Size	389			
4	Adhesiveness	151	13	Benevolence	261	25	Weight	393			
5	Combativeness	157	14	Veneration	274	26	Colouring	399			
6	Destructiveness	165	15	Firmness	285	27	Locality	414			
+	Alimentiveness	184	16	Conscientiousness	288	28	Number	420			
7	Secretiveness	190	17	Hope	304	29	Order	424			
8	Acquisitiveness	203	18	Wonder	309	30	Eventuality	425			
9	Constructiveness	217	19	Ideality	322	31	Time	434			
			?	Unascertained	330	32	Tune	430			
			20	Wit or Mirthfulness	340	33	Language	446			
			21	Imitation	353						

FIGURE 1.1. George Combe, the best-known phrenologist of the 19th century, divided the brain into intellectual faculties and feelings. The plate of the phrenological bust faces the title page in Combe (1830). In the public domain.

FIGURE 1.2. Illustration from a fictional story of a member of a phrenology society who decides to use phrenology to identify a possible thief in his household. The drawing shows a servant who was paid five shillings to shave his head so that the phrenological organs could be traced in ink, not a standard part of phrenology practice. From Prendergast (1844, p. 17). In the public domain.

PHILOSOPHICAL AND SCIENTIFIC ANTECEDENTS

The most prominent British philosopher of his era, Herbert Spencer (1820–1903) sought to synthesize universal natural laws (especially *evolution*) across the disciplines of biology, psychology, sociology, and ethics. Spencer coined the phrase "survival of the fittest" (p. 444) in *The Principles of Biology* (1864) after reading Charles Darwin (1859), although he was reluctant to accept Darwin's evolutionary mechanism of natural selection. In *The Principles of Psychology* (1855), Spencer described how the behavior of the individual organism adapts through interaction with the environment, and defined *intelligence* as a "continuous adjustment" of "inner to outer relations" (p. 486). Spencer's ideas persist in a number of ways to this day. Intelligence, as we shall see, is still widely considered to represent a capacity associated with adaptation to one's environment. In a critical review of Spencer's synthesis, John Dewey (1904) was struck by the luck that Spencer and Darwin published almost simultaneously, thereby making their very different concepts of evolution indistinguishable to the public.

Beyond Spencer's philosophical influence, the foundation for psychology as a science, as well as for the scholarly study of intelligence, was laid by naturalist Charles Darwin (1809–1882), who is most remembered for his theory of evolution by natural selection. In his writings, Darwin frequently referred to adaptive behavior in animals and humans as "intelligent"; more importantly, he argued that the same forces that act on animal evolution also apply to human mental abilities: "There is no fundamental difference between man and the higher mammals in their mental faculties" (Darwin, 1871, p. 35). In *The Descent of Man*, Darwin (1871) went even further in applying his evolutionary theory to human mental characteristics—probably after reading the work of his half-cousin Francis Galton, the Victorian polymath, whose drive for scientific measurement of human capabilities would start the race to develop measures of intelligence in motion.

It is difficult to overstate the impact of Darwin's theory of evolution on psychology. By considering human behavior in an evolutionary context, Darwin treated the study of psychology as no less a science than biology and other natural sciences. His influence was substantial and may be seen, for example, in Joseph Jastrow's (1901) American Psychological Association (APA) presidential address to start the 20th century. Jastrow described psychology as both a laboratory science and an applied science, setting the study of intelligence in a somewhat Spencerian evolutionary context:

> Intelligence must first be realized as an advantage-gaining factor in the evolutionary struggle; that struggle is not merely, and indeed in all the stages that here come into consideration, not mainly a conflict of tooth and nail, a contest of strength of claw and fleetness of foot, but a war of wits, an encounter of skill and cunning, a measure of strategy and foresight. (p. 9)

Francis Galton and the Anthropometric Laboratory

If you lived in London in the mid-1880s or 1890s, you could pay three- or fourpence for you or your children to undergo a variety of physical measurements and tests, with the option to register results for future reference and follow-up. The measurements were available from Francis Galton's Anthropometric Laboratory, first located at the International Health Exhibition (see Figure 1.3), then at the University of Cambridge, the South Kensington Museum, and finally at the Clarendon Museum at Oxford. *Anthropometry* referred to the "measurement of man," and Galton's laboratory was, according to Diamond (1977), "a device to tease the public into providing the data he needed for his research" (p. 52). As a lifelong advocate for objective scientific measurement, Galton (1822–1911; see Figure 1.4) was a pioneer in the use of test batteries and questionnaires for data collection, the concept of control groups in research, and statistical methods (as developer of the techniques of regression and correlation).

Galton introduced his system of anthropometric measurements in *Inquiries into Human Faculty and Its Development* (1883), where he wrote, "It is needless for me to speak here about the differences in intellectual power between different men and different races, or about the convertibility of genius as shown by different members of the same gifted family achieving eminence in varied ways" (pp. 82–83). He conceptualized his measurements as constituting indicators of physical efficiency to complement performance on formal academic written literary examinations, which he thought were the best available measures of intelligence (e.g., Galton, 1884, 1891).

FIGURE 1.3. Francis Galton's first Anthropometric Laboratory was featured at the International Health Exhibition held in London in 1884–1885. Nearly 10,000 people paid threepence each to be examined and receive a copy of their measurements. From Pearson (1924, Plate L). Reprinted by permission of Cambridge University Press.

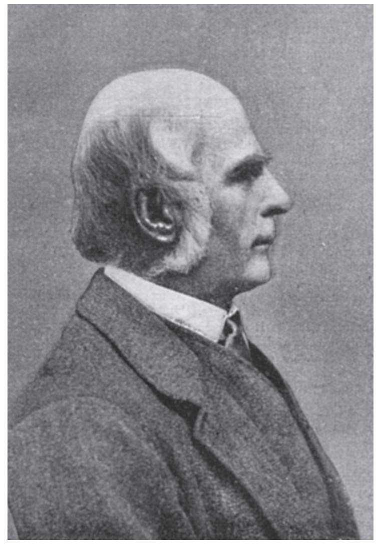


FIGURE 1.4. Francis Galton in 1888 at the age of 66, when the Anthropometric Laboratory remained active. From the copperplate prepared for *Biometrika*. In the public domain.

The examination took less than 1 hour. Although the makeup of the battery changed slightly over time, each session began with the examinee's completing a card recording age, birthplace, marital status (married, unmarried, or widowed), residence (urban, suburban, or country), and occupation. The examinee's name, birth date, and initials were collected in the laboratory's later years, with the full name indexed in a separate list. The examiner then recorded the color of the examinee's eyes and hair, followed by tests and measurements of sensory acuity, stature, strength, and lung capacity:

- Eyesight keenness, color sense, and judgment in estimating length and squareness
- Hearing keenness and highest audible note
- Height standing, without shoes
- Height sitting, from seat of chair
- Span of arms (between opposite fingertips, with arms fully extended)
- Weight, in usual indoor clothing
- Breathing capacity (volume of air exhaled after a deep breath)
- Strength of pull (as an archer draws a bow)
- Strength of grasp (squeeze with the strongest hand)
- Swiftness of blow with fist (against a flat bar with pad at one end)

Specialized instruments (some invented by Galton) were employed, such as the spirometer, which required exhaling into a tube to measure the number of cubic inches of water displaced in a tank. Galton (1890b) interpreted breathing (lung) capacity as an indicator of energy level:

> The possession of a considerable amount of breathing capacity and of muscular strength is an important element of success in an active life, and the rank that a youth holds among his fellows in these respects is a valuable guide to the selection of the occupation for which he is naturally fitted, whether it should be an active or a sedentary one. (p. 238)

Galton constructed normative distributions for each measurement, including mean values and *percentile ranks* (i.e., the percentage of cases falling below the obtained score) in specified age ranges, differentiated by gender. Some measures, like breathing capacity and strength of grip, were assessed in relation to stature. It was possible to look at a normative chart and instantly know your approximate percentile rank. After collecting data on nearly 10,000 examinees at the International Health Exhibition, Galton's laboratory at South Kensington collected data on an additional 3,678 examinees (Galton, 1892), so adult norms were based on fairly large samples.

Galton never directly asserted that his tests measured *intelligence*. Instead, he observed that sensory measures are relevant in determining the breadth of experience upon which intelligence can operate:

> The only information that reaches us concerning outward events appears to pass through the avenue of our senses; and the more perceptive our senses are of difference, the larger is the field upon which our judgment and intelligence can act. (Galton, 1907, p. 19)

In 1890, he acknowledged that only research could reveal the most important areas of human functioning to measure, through careful examination of test results and correlations with external criteria:

> One of the most important objects of measurement is hardly if at all alluded to here and should be emphasized. It is to obtain a general knowledge of the capacities of a man by sinking shafts, as it were, at a few critical points. In order to ascertain the best points for the purpose, the sets of measures should be compared with an independent estimate of the man's powers. We thus may learn which of the measures are the most instructive. (Galton, 1890a, p. 380)

The uncertainty, of course, was where to sink the "shafts"—or, in other words, which abilities to measure.

With the methods proposed by Galton, a belief in scientifically based mental measurement of individual differences began to crystallize in the 1890s, and many independent research efforts were launched in the United States and Europe. Charles E. Spearman (1904, pp. 206–219) counted over 30 international investigators studying mental tests, and this was probably an underestimate. The quest for mental tests is generally agreed to have started in Great Britain with Galton's initiatives, but Spearman's (1904) discovery of a *general intellectual factor*, described in a later section, would almost immediately begin to guide theory development. The earliest U.S. efforts in mental testing came through large-scale studies from James McKeen Cattell (Cattell & Farrand, 1896) at Columbia University; Franz Boas, then at Clark University (see Bolton, 1892); J. Allen Gilbert (1894) at Yale University; and Joseph Jastrow (1893) at the University of Wisconsin, Madison. In France, Alfred Binet and his colleagues (principally Victor Henri and then Théodore Simon) were the pioneers. Germany's early contributors included Hermann Ebbinghaus and Emil Kraepelin, especially his student Axel Oehrn (1896).

It is debatable whether efforts to develop a working intelligence test ever became a scientific race, like the competitive quest to identify the molecular structure of deoxyribonucleic acid (DNA) or the pursuit of space travel technology to become the first nation to land a person on the moon. Certainly, there was constant comparison between test development efforts in different nations. For example, Sharp (1899) reviewed the competing perspectives of "M. Binet and the French psychologists," "Prof. Kraepelin and the German psychologists," and the American psychologists (p. 334), pitting the assertions of each research group against one another. In journals like *L'Année Psychologique*, Binet and his colleagues could be found reviewing work by all competing laboratories, even commenting on Sharp's paper. After Spearman (1904) described his statistical method of quantifying "general" intelligence, competition between research groups may have become even more pronounced because a more focused end goal had been specified (i.e., a test of "general" intelligence per se, rather than random tests of associated mental processes). As shown in Figure 1.5, the practice of intelligence assessment in the earliest years of the 20th century essentially consisted of an array of sensory

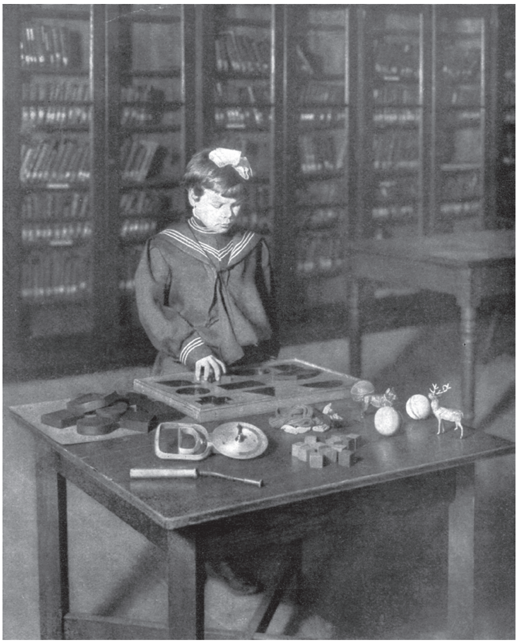


FIGURE 1.5. A photograph depicting the array of tasks used to measure intelligence at Lightner Witmer's Psychological Clinic at the University of Pennsylvania in about 1908. On the table are a Galton whistle for testing the upper limit of sound perception; a dynamometer for testing hand strength; colored yarns and blocks for testing counting skills and color perception; toys to test common knowledge, play, instinctive reactions, and coordination; and the formboard for identifying nonverbal problem solving and detecting feeble-mindedness. From Carter (1909, p. 166). In the public domain.

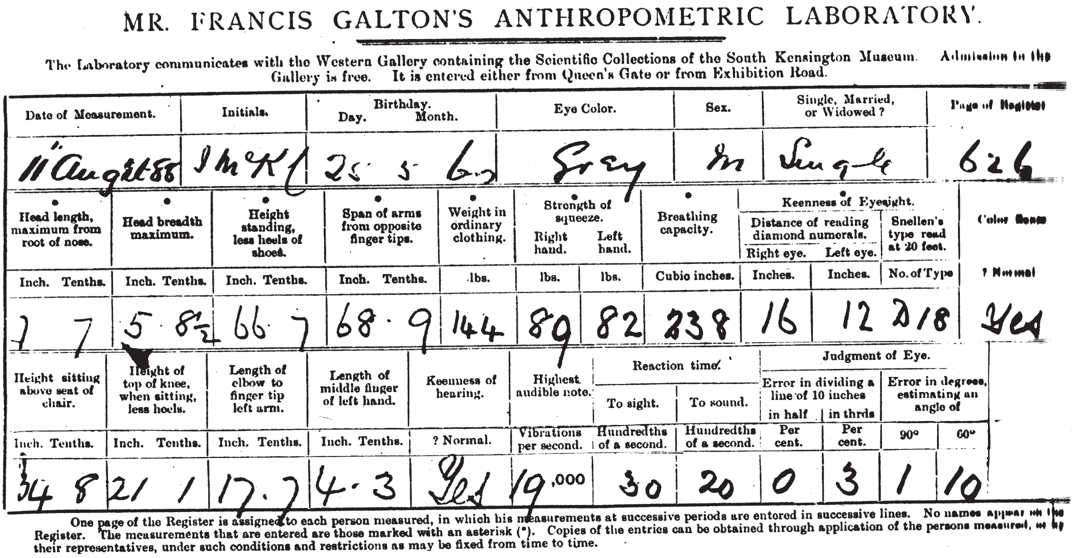
and motor measures, with a formboard to measure higher mental processes.

James McKeen Cattell and the End of Anthropometrics

If there were royalty in the field of psychology, James McKeen Cattell (1860–1944) might qualify. He was the son of a professor at (and later the president of) Lafayette College in Easton, Pennsylvania, where he graduated as valedictorian in 1880. After studying for 2 years in Germany, he won a fellowship at Johns Hopkins University, where he began researching the timing of various mental processes in G. Stanley Hall's "physiologico-psychological laboratory" (Sokal, 1981, p. 64). He left to study with Wilhelm Wundt, the father of experimental psychology, at the University of Leipzig, Germany, where he worked from 1883 to 1886 before receiving his doctorate. His article "The Time It Takes to See and Name Objects" (Cattell, 1886) summarized two of his studies on basic reading processes, which are now considered to be the first research studies to support a whole-word, sight-reading approach to reading instruction (Venezky, 2002, p. 6). Rejecting Wundt's reliance on experimenter introspection, Cattell conducted reaction time experiments with some of his own instruments, growing interested in the measurement of individual differences. According to his biographer, Michael M. Sokal, Cattell "refocused psychological research away from experimenters' self-observation of their mental activity and toward subjects' behavior in a laboratory setting precisely defined by experimenters" (Sokal, 2006, p. 25). In just a few years, Cattell would become the leading American experimental psychologist of his time.

In 1887, Cattell took a position at the University of Cambridge, where he came to know and work closely with Francis Galton. Cattell's data card from his personal anthropometric measurements appears in Figure 1.6. Cattell helped Galton set up the Anthropometric Laboratory at South Kensington. Cattell would remain devoted to Galton for the rest of his life, acknowledging in his late 60s that Galton was "the greatest man whom I have known" (Cattell, 1930, p. 116). For 2 years, Cattell split his time between work in Galton's laboratory, lecturing and establishing a laboratory at Cambridge University, and lecturing also at Bryn Mawr College and the University of Pennsylvania in the United States. In 1888, Cattell became a professor of psychology at the University of Pennsylvania (the first such professorship established anywhere, he claimed). In 1891, Cattell relocated to Columbia University, where he became the administrative head—beginning Columbia's experimental psychology laboratory and mentoring doctoral students like Edward L. Thorndike, Robert S. Woodworth, and Harry L. Hollingworth, who would themselves become faculty and leading figures in psychology. Over 40 students would take their doctorates with Cattell, seven of them becoming presidents of the APA. Cattell himself served as president of the APA in 1895.

With respect to intelligence testing, Cattell is a seminal historical figure due to his tireless advocacy for psychology as a science, his own test development efforts, and his advocacy for psychometrics and testing, as well as his emphasis on statistical analyses of individual differences, all of



MR. FRANCIS GALTON'S ANTHROPOMETRIC LABORATORY.

The Laboratory communicates with the Western Gallery containing the Scientific Collections of the South Kensington Museum. [illegible] Gallery is free. It is entered either from Queen's Gate or from Exhibition Road.

Date of Measurement.	Initials.	Birthday. Day. Month.	Eye Color.	Sex.	Single, Married, or Widowed?	[illegible]
11 Aug 88	JMcKC	25 5 60	Grey	M	Single	626

* Head length, maximum from root of nose.	* Head breadth maximum.	* Height standing, less heels of shoes.	* Span of arms from opposite finger tips.	* Weight in ordinary clothing.	* Strength of squeeze. Right hand.	Left hand.	* Breathing capacity.	* Keenness of Eyesight. Distance of reading diamond numerals. Right eye.	Left eye.	Snellen's type read at 20 feet.	[illegible]
Inch. Tenths.	Inch. Tenths.	Inch. Tenths.	Inch. Tenths.	lbs.	lbs.	lbs.	Cubic inches.	Inches.	Inches.	No. of Type	? Normal
7 7	5 8½	66 7	68 9	144	89	82	238	16	12	D18	Yes

Height sitting above seat of chair.	Height of top of knee, when sitting, less heels.	Length of elbow to finger tip left arm.	Length of middle finger of left hand.	Keenness of hearing.	Highest audible note.	Reaction time. To sight.	To sound.	Judgment of Eye. Error in dividing a line of 10 inches in half	in thrds	Error in degrees, estimating an angle of 90°	60°
Inch. Tenths.	Inch. Tenths.	Inch. Tenths.	Inch. Tenths.	? Normal.	Vibrations per second.	Hundredths of a second.	Hundredths of a second.	Per cent.	Per cent.		
34 8	21 1	17 7	4 3	Yes	19,000	30	20	0	3	1	10

One page of the Register is assigned to each person measured, in which his measurements at successive periods are entered in successive lines. No names [illegible] Register. The measurements that are entered are those marked with an asterisk (*). Copies of the entries can be obtained through application of the persons measured, [illegible] their representatives, under such conditions and restrictions as may be fixed from time to time.

FIGURE 1.6. Measurement data card recorded in 1888 at Galton's Anthropometric Laboratory for J. McKeen Cattell, who was deeply influenced by Francis Galton. Papers of James McKeen Cattell, 1835–1948, Library of Congress, Washington, D.C. In the public domain.

which established a fertile environment for test development at Columbia University and in American psychology in general. In the British journal *Mind*, Cattell (1890) used the term *mental tests* for the first time:

> Psychology cannot attain the certainty and exactness of the physical sciences, unless it rests on a foundation of experiment and measurement. A step in this direction could be made by applying a series of mental tests and measurements to a large number of individuals. The results would be of considerable scientific value in discovering the constancy of mental processes, their interdependence, and their variation under different circumstances. Individuals, besides, would find their tests interesting, and perhaps, useful in regard to training, mode of life or indication of disease. (p. 373)

Cattell made his principal research initiative at Columbia an investigation to determine whether a battery of Galtonian anthropometric tests and sensory, motor, and higher cognitive tasks could constitute a measure of intelligence. Beginning in 1894, the Cattell–Columbia Tests (as Cattell referred to them in 1924) were given to freshmen at Columbia's School of Arts and School of Mines. With student consent, the tests were to be repeated at the end of the sophomore and senior years. In the course of an hour, 26 measurements were made in the laboratory, and 44 observations were recorded. Later, each student sent in answers to 50 questions with regard to background, health, physical condition, habits (including coffee, smoking, alcohol use, and exercise), and interests. Cattell also had access to student academic records and athletic accomplishments.

Tests and measurements conducted in the laboratory included some of Galton's sensory measures; some of Cattell's reaction time measures; and some newer measures, including letter cancellation, rapid color naming, memory for digits, logical memory, self-reported retrieval of mental images, and a word association test. The battery was something of a hybrid between anthropometric, lower-order, and higher-order measures. Cattell had always relied on the experimental approach as producing descriptive results that would speak for themselves; he did not offer a priori hypotheses or even articulate his concept of intelligence. Cattell's commitment to quantitative measurement and statistical analysis of experimental results was unshakeable, and as late as 1924, Cattell, pictured in Figure 1.7, still expressed a belief that his test battery might correlate with long-term student accomplishments. He would not have the chance to find out, as two studies would put a conclusive end to Cattell's approach to intelligence testing and his experimental research efforts.

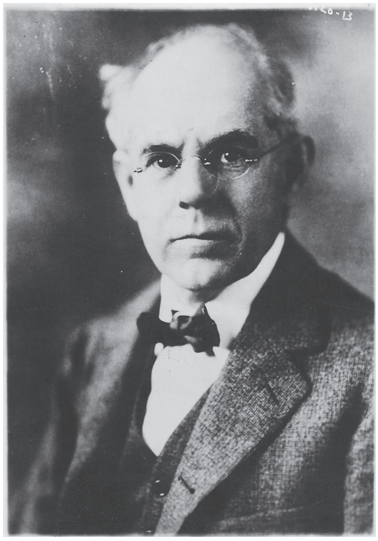


FIGURE 1.7. James McKeen Cattell at the age of 63 in December 1923. Long after the failure of his anthropometric testing program and after his 1917 dismissal from Columbia University, Cattell founded The Psychological Corporation and continued to edit several scientific journals. From the chapter author's collection.

First, a dissertation completed by Stella Sharp (1899) in Edward B. Titchener's laboratory at Cornell University sought to examine the variability of complex mental processes and the relations between complex mental processes, with the intention of demonstrating the practicality of testing complex processes rather than the simpler mental processes endorsed by Cattell and Galton. She assessed seven advanced philosophy students at the university with the test battery formulated by Binet and Henri (1895), including measures of memory, mental images, imagination, attention, observation/comprehension, suggestibility, and aesthetic tastes. Her results listed the scores of individual participants and described these results in terms of rank order and variability. Sharp concluded:

> We concur with Mm. Binet and Henri in believing that individual psychical differences should be sought for in the complex rather than in the elementary processes of mind, and that the test method is the most workable one that has yet been proposed for investigating these processes. (Sharp, 1899, p. 390)

She further concluded that the Binet–Henri measures required modification but were practical and yielded considerable variation in scores. She offered only qualitative observations about relations between tests of different mental processes, however. Although she did not collect data on other assessment approaches, she was critical of the anthropometric tests as unproven and lacking an explanatory theory.

The second blow to Cattell's testing program, and its *coup de grâce*, came from a Columbia University psychology graduate student, Clark Wissler (1901). Wissler examined the correlations between the Cattell–Columbia Tests and student grades for 300 undergraduates at Columbia and Barnard Colleges. He reported that while isolated correlations were large (e.g., height and weight r = .66; Latin and Greek grades r = .75), the laboratory mental tests had negligible correlations with each other and with college class grades. The failure to correlate with academic grades was considered fatal to Cattell's testing program because academic performance had long been considered an independent criterion measure of intelligence. In the words of Cattell's biographer, Wissler's analysis would definitively "discredit anthropometric testing" (Sokal, 2006, p. 29).

It remains to note that over a century after Galton's and Cattell's testing programs were discredited, the relations of elementary cognitive processes (reaction time and sensory discrimination) to mental abilities and intelligence are now being revisited. Jensen (2006) has effectively summarized the literature relating reaction time to intelligence, while Deary and his colleagues (Deary, 1994; Deary, Bell, Bell, Campbell, & Fazal, 2004) have documented findings with sensory discrimination and intelligence. There is uniform agreement as to the serious methodological flaws in the Sharp and Wissler studies, including small sample size, restriction of range, and unreliability of measures (e.g., Buckhalt, 1991; Deary, 1994; Jensen, 2006).

THE ORIGINS OF CONTEMPORARY INTELLIGENCE TESTING

I have described the pseudoscience of phrenology and the visionary science of Galton, who inspired the search for effective ways to measure intelligence and who pioneered many statistical methods that would be critical for norm-referenced assessment. I have also recounted the tale of the psychologist that Galton so profoundly influenced, J. McKeen Cattell, who threw down a gauntlet of sorts when proposing that *mental tests* should constitute part of establishing psychology as a science that can measure individual differences. The unfortunate fate of Cattell's test battery has been told. Even after his assessment work was discredited, however, Cattell remained a highly connected institutional scientist and a pioneer in the development of scientific psychology in American universities.

Ironically, the problem of developing a working intelligence test would be solved by an outsider, a man with few friends, who worked without pay and who had no institutional connections of any benefit. He did have his own journal, however, where he reviewed the work of his contemporaries. His name was Alfred Binet.

Alfred Binet: The Innovative Outsider

Alfred Binet (1857–1911) is generally acknowledged as the father of intelligence tests, having developed the first working measure of intelligence. Binet's remarkable history has been most definitively documented by biographer Theta Wolf (1973). He was educated as a lawyer but chose not to practice; some historical accounts also report that Binet studied medicine until his father, a physician, traumatized him by showing him a cadaver. As an only child of a wealthy family he could afford to pursue a career with little remuneration, and he developed a consuming interest in the study of psychology. He was a voracious reader across several languages who educated himself as a psychologist, spending considerable time studying in the Bibliothèque Nationale [French National Library]. He wrote his first article at age 23 and completed a doctorate in the natural sciences at age 37. According to long-time colleague Théodore Simon, for most of Binet's career "psychology was his sole occupation" (quoted by Wolf, 1973, p. 9).

Although he is remembered for his intelligence test (from which he does not appear to have profited financially), Alfred Binet was a remarkably productive and versatile researcher, authoring nearly 300 works during his career; he is now credited with pioneering experimental investigations in areas of abnormal, cognitive, developmental, educational, forensic, personality, and social psychology (e.g., Siegler, 1992; Wolf, 1973). Regrettably, most of his work has never been translated into English, although nearly all of it has been brought back into print in the last decade. Personally, he has been described as a loner, "a reserved man with few friends" (Tuddenham, 1974, p. 1071), and as a domineering individual who antagonized many of his coworkers (cf. Henri Piéron, according to an interview with Wolf, 1961, p. 246). In 1901, Binet wrote a friend, "I educated myself all alone, without any teachers; I have arrived at my present scientific situation by the sole force of my fists; *no one*, you understand, no one, has ever helped me" (quoted by Wolf, 1973, p. 23). Lacking patronage, he was denied academic positions in France (Nicolas & Ferrand, 2002), and his efforts for educational reform and mental measurement in the military were resisted by a rigid French establishment (e.g., Carson, 2007; Zazzo, 1993). Several scholars have sought to explain why so much of his work was forgotten after his death (e.g., Fancher, 1998; Schneider, 1992; Siegler, 1992; Wolf, 1973), and the answer seems to lie in his disconnection from the professional and academic community in France: He did not present at conferences, he did not leave students to continue his work, and he preferred to work alone or with a collaborator. Only at the 2005 centennial of the first Binet–Simon scale did he begin to garner national recognition in his native France for his remarkable contributions.

A devastating early career setback left Binet careful to avoid preconceptions and reticent to form theories in his experimental work. In one of his first positions, Binet's work with collaborator Charles Féré at Jean Martin Charcot's clinic in the Salpêtrière Hospital in Paris was publicly discredited. In working with Charcot's hysterical patients in the 1880s (at a time when Sigmund Freud was also studying with Charcot), Binet and Féré thought they had proven that movements and perceptions could be shifted from one side of the body to the other, and that emotions could be reversed (e.g., from love to hate) through the application of magnets while participants were hypnotized. Their demonstrations could not be independently replicated, presumably because participants were prone to dissimulation or demand effects. Charcot's approach and the Binet–Féré studies were effectively challenged in a series of articles by a Belgian psychologist, Franz Joseph Delboeuf. Gradually Binet realized he had been "taken in" by Charcot's reputation; suitably chastened, he learned painful lessons about the need for careful experimentation, objective observation, and skepticism about a priori theoretical assumptions. Binet left the Salpêtrière in 1890, ending his connections with Charcot and Féré. For the rest of his life, he remained wary of theories that might bias his research findings.

From 1890 through his death, Binet published more than 200 articles, many in the journal *L'Année Psychologique*, which he cofounded and edited. In 1891, he became an unpaid staff member at the Laboratory of Physiological Psychology at the Sorbonne. Three years later, he became director of the laboratory, a position he held until his death. Between 1894 and 1898, Binet and Victor Henri sought new methods that would "substitute for vague notions of man in general, of the archetypal man, precise observations of individuals considered in all the complexity and variety of their aptitudes" (Binet & Henri, 1894, p. 167; translated and cited by Carson, 2007, p. 132). In 1899, Binet was approached by a young intern physician and *aliéniste* (psychiatrist), Théodore Simon, who had access to clinical populations (Wolf, 1961). Simon completed his doctoral thesis under Binet's supervision, and their subsequent collaborations included Binet's most important work in intelligence assessment.

The creative work that culminated in the intelligence scales began in 1890, when Binet published three papers describing experimental studies with his two young daughters, Madeleine and Alice (given the pseudonyms Marguerite and Armande), whom he had carefully observed and tested with a variety of cognitive and personality tasks. Binet's wife and daughters appear in Figure 1.8. In describing their attentional styles, he wrote that Madeleine was "silent, cool, concentrated, while Alice was a laugher, gay, thoughtless, giddy, and turbulent" (translated by Wolf, 1966, p. 234). Many of the tasks Binet gave his daughters will be familiar to contemporary psychologists—word generation, word association, sentence generation, sentence completion, thematic composition for various pictures, description of objects or pictures, design reproduction, letter cancellation, digit repetition, reaction time after stimulation, recall of unrelated words, recall of prose passages, recall of pictured objects, recall of nonsense sentences (i.e., sentences presented in a foreign language unknown to the girls), and verbal responses to inkblots. Binet was careful to analyze the quality of response content. For example, he classified verbal responses as personal, unelaborated, abstract, or imagined; prose recall was scored according for verbatim recall and gist recall; thought processes were described according to linearity, conventionality, and originality. Binet continually made qualitative observations of performance styles that differentiated his daughters. Madeleine had greater "stability" and had better voluntary control of her attention; she could more effectively focus on assigned work and memorize neutral, uninteresting material; and she tended to respond with shorter and more constant reaction times. Alice presented more "variability"; she was more imaginative and emotional; and material to be learned had to be of interest to her, or she would have difficulty. She could not memorize long literary passages verbatim as her sister could, but she could accurately remember a series of ideas provided just once (see Wolf, 1973, p. 132). Binet continued to test his daughters through midadolescence with a battery of cognitive and personality tests, including measures of attention, language, reasoning, and memory (many repeated multiple times in alternative

FIGURE 1.8. Alfred Binet; his wife, Laure Balbiani Binet; and his daughters, Madeleine and Alice, whom he tested extensively with cognitive and personality tasks through their midadolescence. In his writings, Binet described the girls under the pseudonyms Marguerite and Armande. Madeleine was said to be "silent, cool, concentrated, while Alice was a laugher, gay, thoughtless, giddy, and turbulent" (translated by Wolf, 1966, p. 234). Theta H. Wolf Papers, Archives of the History of American Psychology, Center for the History of Psychology, University of Akron. Reprinted by permission.

forms over several years), always accompanied by careful qualitative observation and interview inquiries. He reported the results in 1903 in *L'étude Expérimentale de l'Intelligence [The Experimental Study of Intelligence]*.

Comparison of his children's performances with each other and with those of adults led Binet to conclude that complex, multidimensional tasks were more sensitive to developmental changes than narrow, unidimensional tasks. He further concluded that a mental developmental progression from childhood through adulthood should be reflected in task performance:

> In case one should succeed in measuring intelligence—that is to say, reasoning, judgment, memory, the ability to make abstractions—which appears not absolutely impossible to me, the figure that would represent the average intellectual development of an adult would present an entirely different relation to that of the figure representing the intellectual development of a child. (Binet, 1890, p. 74, translated by Wolf, 1966, p. 235)

In 1895, Binet and Henri outlined the project for the development of an intelligence test, specifying 10 discrete mental faculties that would be measured: memory, imagery, imagination, attention, comprehension, suggestibility, aesthetic sentiment, moral sentiment, muscular strength/willpower, and motor ability/hand–eye coordination. Binet and Henri, along with other colleagues from Binet's laboratory, appear in Figure 1.9. Higher-order, complex processes were considered to show greater variability among individuals and to constitute better measures of intelligence than simpler sensory and motor processes:

> The higher and more complex a process is, the more it varies in individuals; sensations vary from one individual to another, but less so than memory; memory of sensations varies less than memories of ideas, etc. The result is, that if one wishes to study the differences existing between two individuals, it is necessary to begin with the most intellectual and complex processes, and it is only secondarily necessary to consider the simple and elementary processes. (Binet & Henri, 1895, p. 417; translated by Sharp, 1899, p. 335)

In a passage that made direct reference to the work of Galton and Cattell, Binet and Henri (1895) rebutted the claim that greater experimental precision was possible in the measurement of simpler mental processes:

> If one looks at the series of experiments made—the *mental tests* as the English say—one is astonished by the considerable place reserved to the sensations and the simple processes, and by the little attention lent to the superior processes. . . . The objection will be made that the elementary processes can be determined with much more precision than the superior processes. This is certain, but people differ in these elementary ones much more feebly than in the complex ones; there is no need, therefore, for as precise a method for determining the latter as for the former. . . . Anyway, it is only by applying one's self to this point that one can approach the study of individual differences. (Binet & Henri, 1895, pp. 426, 429; translated by Siegler, 1992, p. 181)

The formal mandate that led to the development of the intelligence test came in October 1904, when Joseph Chaumié, the Minister of Public Instruction, established a commission chaired

FIGURE 1.9. Alfred Binet and colleagues in the 1890s: J. Courtier, J. Philippe, and Victor Henri. This group collaborated in the work involved for Binet's (1894) book on experimental psychology. Photo from Binet's Laboratoire de Psychologie Physiologique de la Sorbonne (École des Hautes-Études). Theta H. Wolf Papers, Archives of the History of American Psychology, Center for the History of Psychology, University of Akron. Reprinted by permission.

by Léon Bourgeois and charged it with studying how France's 1882 mandatory public education laws could be applied to abnormal [*anormaux*] children, including students who were blind, deaf-mute, and backward [*arriérés*] (Carson, 2007; Wolf, 1969). The ministry was persuaded to take this initiative by public pressure, including a resolution from the 1903 Third National Congress of Public and Private Welfare held at Bordeaux, where critics noted France's failure to comply with its own special education laws. A resolution from an educational advocacy group, La Société Libre pour l'Étude Psychologique de l'Enfant [Free Society for the Psychological Study of the Child], was also a reason for creation of the Bourgeois Commission. Binet was a leader of La Société Libre and an author of the resolution, and he became a member of the commission, which began its work by circulating questionnaires to teachers and principals throughout France. The commission met on numerous occasions in 1904 and 1905, issuing its report in 1906.

Binet saw in the commission's mandate an opportunity to complete his efforts toward a norm-referenced standard for diagnosis and educational decision making. Building on the earlier work with Henri (who had departed), Binet and Simon developed and tested a series of cognitive tests. Their collaborations worked in tandem: One of them would talk with and question the examinee, while the other wrote the replies and noted salient behaviors. The assessments had "the air of a game" for children, with encouragement being constantly provided (Binet & Simon, 1905/1916a, p. 141). The work culminated in the 1905 publication of the Binet–Simon Intelligence Scale (Binet & Simon, 1905/1916c), consisting of 30 items that could be given in about 20 minutes; it was normed on some 50 children from ages 3 through 11 years; and one of its chief advances may have been to combine a wide range of cognitive tasks to obtain a global estimate of intelligence (e.g., DuBois, 1970). Binet and Simon (1905/1916c) sequenced tasks in a cognitive-developmental order from easy to hard and from simpler to more complex, while sampling a wide range of tasks tapping various abilities. In general, they sought tasks that tapped the higher-order ability of *judgment*—especially procedures that had demonstrated the capacity to differentiate groups on the basis of intelligence. For example, individuals considered *idiots* generally could not move beyond the 6th of the 30 tasks; individuals considered *imbeciles* rarely went beyond the 15th task (Binet & Simon, 1905/1916a).

The Bourgeois Commission issued its report early in 1906, based primarily on a subcommittee report drafted by Binet. Recommendations were that the *anormaux* be educated through *classes spéciales* annexed to ordinary primary schools and, in certain situations, through separate institutions. A five-part classification of exceptional students was proposed, identifying students who were blind, deaf, medically abnormal, intellectually backward, and emotionally unstable. The commission recommended that students who did not benefit from education, teaching, or discipline should receive a "medico-pedagogical examination" before being removed from primary schools, and that such children, if educable, should be placed in special classes. The examination was to be overseen by an examination committee consisting of an inspector of primary schools, a physician, and a director of the separate special school. The commission did not offer any specific content for the examination, recommending that the Minister of Public Instruction appoint a competent person to draw up a scientific guide for the school examination committee (Carson, 2007). Undoubtedly Binet hoped to draw up the scientific guide, and Binet and Simon's (1907/1914) book *Les Enfants Anormaux* was probably intended to serve as the guide; it even contained a preface by Léon Bourgeois, the head of the commission.

Unfortunately, Binet's efforts were almost completely rebuffed by the French establishment. When the French legislature enacted the law of April 15, 1909, on the education of the *anormaux*, it stated that the commission determining eligibility for special education should be composed of a physician, school inspector, and director or teacher at an *école perfectionnement*. It highlighted the medical examination and made no mention of any role for psychologists or use of special methods (i.e., intelligence tests) for assessing students (Carson, 2007). Binet's efforts had little visible impact on practice in his native France.

In the 1908 revision, the Binet–Simon Scale took its definitive revolutionary form, the "graded scale of intelligence" [*L'échelle métrique de l' intelligence*], which was easier to use and interpret (Binet & Simon, 1908/1916b). It featured 56 tests arranged by difficulty so that tests were placed at levels, or grades, corresponding to approximately a 75% pass rate for children of a given age, based on normative performances of about 200 children between the ages of 3 and 15. The 1908 scale permitted a student's mental level [*niveau mental*] to be estimated through what later became interpreted in the United States as a mental age level. The mental level was determined by the highest age at which a child passed four or five tests (the basal year), with an additional year credited for each of the five tests passed beyond the basal. By the completion of the 1911 edition (Binet, 1911/1916), the scale was extended from age 3 through adulthood, with 11 levels and five items administered at each level. Table 1.1 lists content from the final 1911 scale. The Binet–Simon Scale never yielded an intelligence quotient (IQ), but Binet endorsed the convention of identifying intellectual disability [*arriérés*] for a mental level delay of "two years when the child is under [age] nine, and three

TABLE 1.1. Contents of the Binet–Simon (Binet, 1911/1916) Intelligence Scale [*L'Échelle Métrique de l' Intelligence*]

Three years	*Nine years*
Show eyes, nose, mouth	Give change out of 20 sous
Name objects in a picture	Definitions superior to use
Repeat 2 figures	Recognize the value of 9 pieces of money
Repeat a sentence of 6 syllables	Name the months
Give last name	Comprehend easy questions
Four years	*Ten years*
Give sex	Place 5 weights in order
Name key, knife, penny	Copy a design from memory
Repeat 3 figures	Criticize absurd statements
Compare 2 lines	Comprehend difficult questions
	Place 3 words in 2 sentences
Five years	*Twelve years*
Compare 2 boxes of different weights	Resist the suggestion of lines
Copy a square	Place 3 words in 1 sentence
Repeat a sentence of 10 syllables	Give more than 60 words in 3 minutes
Count 4 sous	Define 3 abstract words
Put together two pieces in a "game of patience"	Comprehend a disarranged sentence
Six years	*Fifteen years*
Distinguish morning and evening	Repeat 7 figures
Define by use	Find 3 rhymes
Copy diamond	Repeat a sentence of 26 syllables
Count 13 pennies	Interpret a picture
Compare 2 pictures esthetically	Solve a problem composed of several facts
Seven years	*Adults*
Right hand, left ear	Comprehend a cut in a folded paper
Describe a picture	Reversed triangle
Execute 3 commissions	Answer the question about the President
Count 3 single and 3 double sous	Distinguish abstract words
Name 4 colors	Give the sense of the quotation from Hervieu
Eight years	
Compare 2 objects from memory	
Count from 20 to 0	
Indicate omission in pictures	
Give the date	
Repeat 5 digits	

Note. The final 1911 Binet–Simon Scale extended from 3 years into adulthood. In this edition, an individual's mental level [*niveau mental*] was estimated by identifying the highest age at which all the tests were passed (the basal year), to which is added one-fifth of a year for every test passed. The Binet–Simon Scale never yielded an intelligence quotient (IQ), but Binet endorsed the convention of identifying intellectual disability for a mental-level delay of 2 years when a child is under age 9, and 3 years past his or her 9th birthday. From Binet (1911/1916). In the public domain.

years when he is past his ninth birthday" (Binet & Simon, 1907/1914, p. 42). Long after Binet's death, Simon indicated that the use of a summary IQ score was a betrayal [*trahison*] of the scale's objective (cited by Wolf, 1973, p. 203).

In the spring of 1908, Henry H. Goddard, director of the psychological research laboratory at the New Jersey Training School for Feeble-Minded Girls and Boys (later known as the Vineland Training School), traveled to Europe. He visited doctors and teachers working in 19 different institutions and 93 special classes. Ironically, he did not even look up Binet in Paris, having been told by Pierre Janet that "Binet's Lab. is largely a myth . . . Not much being done—says Janet," according to his journal (cited by Zenderland, 1998, pp. 92–93). In Brussels, he met Ovide Decroly, a Belgian teacher, physician, and psychologist, engaged in a tryout of the 1905 Binet–Simon Scale. Decroly provided him with a copy of the test, and upon his return home, Goddard began to use the test on the children at the training school. In the words of Goddard's biographer Leila Zenderland (1998), Goddard immediately understood the significance of the Binet–Simon Scale:

> Two years of frustrating institutional experience had prepared him to see what Janet, Cattell, and even [G. Stanley] Hall, the most prescient of contemporary psychological entrepreneurs, had missed. Contained within Binet's articles, Goddard quickly realized, was an entirely new psychological approach toward diagnosing and classifying feeble minds. (p. 93)

In a short time, Goddard would become the United States' leading advocate for Binet's approach to assessment and diagnosing intellectually disabled individuals. He described his evaluation of the ease, simplicity, and the utility of the 1908 scale as "a surprise and a gratification" (Goddard, 1916, p. 5), and he promoted the test widely. The Binet–Simon Scale was both praised and criticized widely in professional journals; for example, several consecutive issues of the *Journal of Educational Psychology* in April, May, and June 1916 were dedicated to "Mentality Tests: A Symposium," a wide-ranging exchange of experiences with the Binet–Simon Scale (and other tests) by 16 leading psychologists. Goddard arranged for Elizabeth S. Kite, his laboratory's field worker and contributor to the famous Kallikak study, to complete the definitive translations into English of Binet and Simon's writings on their intelligence scale. By 1916, the Vineland laboratory had distributed 22,000 copies of a pamphlet describing administration of the Binet–Simon Scale and 88,000 record forms, as well as publishing a two-volume translation of the Binet–Simon articles (Goddard, 1916). By 1939, there were some 77 available adaptations and translations of the Binet–Simon Scale (Hildreth, 1939), including the most used psychological test of all, the Stanford–Binet. According to Théodore Simon, Binet gave Lewis M. Terman at Stanford University the rights to publish an American revision of the Binet–Simon Scale "for a token of one dollar" (cited by Wolf, 1973, p. 35). Terman's work would change the landscape for mental testing in the United States.

The Binet–Simon Intelligence Scale represented a major paradigm shift for the young field of psychology. It tapped intelligence through assessment of complex mental abilities, as opposed to the narrow sensory and motor measures dominating the Galton–Cattell batteries. It was standardized, with explicit procedures for administration and objective scoring guidelines. It was norm-referenced, permitting an individual's performance to be compared with that of his or her age peers. It was reliable, yielding consistent scores from one occasion to another. It was developmentally sensitive, recognizing that mental abilities in children develop in a meaningful progression and that the abilities of children differ substantially from that of adults. It was efficient and engaging, administered in an adaptive format in which content changed frequently. It offered clinical assessment, aimed at diagnosing intellectual disabilities, identifying cognitively advanced students, and describing the characteristics of both "normal" and "abnormal" individuals. Finally and most importantly, it seemed to work fairly well, providing an empirical foundation for the nascent study of intelligence and cognitive abilities.

Lewis M. Terman: Defender of the Discipline

> I hate the impudence of a claim that in fifty minutes you can judge and classify a human being's predestined fitness in life. I hate the pretentiousness of that claim. I hate the abuse of scientific method which it involves. I hate the sense of superiority which it creates, and the sense of inferiority which it imposes.
>
> —Walter Lippmann (1923)

When journalist Walter Lippmann launched the first high-profile public attack on intelligence testing in a series of articles in *The New Republic* (Lippmann, 1922a, 1922b, 1922c, 1922d, 1922e, 1922f, 1923), it was Lewis M. Terman (1922a) who responded and defended the new discipline. He was the natural choice—developer of the Stanford University revision of the Binet–Simon Intelligence Scale (later called the Stanford–Binet Intelligence Scale); member of the National Research Council team that created the Army mental tests in 1917 and 1918; coauthor of the National Intelligence Tests and Terman Group Test of Mental Ability, released in 1920; principal investigator on the longitudinal Genetic Studies of Genius, initiated in 1921–1922; and coauthor of the Stanford Achievement Test, which would be released in 1923. For decades, Terman would be the living American most strongly associated with intelligence testing and its value for educational decision making.

The 12th of 14 children from a rural Indiana farming family, Lewis M. Terman (1877–1956) was a brilliant, hard-working, and determined student from an early age; he accelerated from first grade to third grade and memorized most of his textbooks. Graduating early from eighth grade (the conclusion of education in typical Midwest farming communities of that era), he began teacher's college at the age of 15, attending when he could and taking breaks to earn enough money to return. He pursued training in education, as teaching was the "only avenue of escape for the youth who aspired to anything beyond farm life" (Terman, 1932, p. 300); eventually he would teach for one year in a one-room schoolhouse. By the age of 21, he had earned three baccalaureate degrees from Central Normal College in Danville, Indiana, and he became a principal of a small high school. He then pursued a master's degree in psychology at Indiana University, followed by a doctorate at Clark University. In 1905, recurrent tubercular hemorrhages in his lungs (eventually the cause of his death) forced Terman to relocate his family to Southern California, where he worked again as a high school principal and then as a professor of pedagogy at Los Angeles State Normal School (later UCLA) before accepting a position in 1909 at Stanford University, where he remained for the duration of his career. Figure 1.10 shows Terman at about the time he started his career at Stanford University.

Terman is described by two biographers, Henry L. Minton (1988) and May V. Seagoe (1975), as having been a highly gifted man and voracious learner, who was tirelessly persistent, intense, and sensitive. As a rigorous and careful researcher, he became a pioneer in mental testing by creating the best of many adaptations of the Binet–Simon Scale. He also harbored a progressive vision of large-scale testing to identify the individual differences and needs of schoolchildren, as well as to identify intellectually gifted children (Chapman, 1988). Like Cattell, Terman had been seen by a phrenologist as a child; he was deeply impressed by the experience and remembered that the phrenologist "predicted great things of me" (Terman, 1932, p. 303). Having spent 6 months each year during his adolescence toiling at farmwork from 5:00 A.M. through about 7:00 or 8:00 P.M., Terman considered his intellectual abilities to have been inherited; he remembered his lengthy stints at farmwork as periods without mental stimulation, contributing to his conviction that environment was substantially less important than heredity in explaining intelligence.

FIGURE 1.10. Lewis M. Terman in 1910, the year he arrived at Stanford University. Terman was the leading advocate for intelligence testing in the first half of the 20th century. Reprinted by courtesy of the Department of Special Collections and University Archives, Stanford University Libraries.

Terman's master's thesis on leadership, his doctoral dissertation on genius, and his longitudinal study of gifted children beginning in 1921–1922 all contributed to his status as founder of the "gifted child" movement. Terman's thesis, published as a journal article in 1904, used experimental methodology (from Binet's suggestibility studies), teacher ratings, and questionnaires to examine leadership in male and female schoolchildren from grades 2 through 8. It is a qualitatively rich study that identifies different types of leaders and subtly links leadership with perceived intelligence. Terman's dissertation, completed in 1905 and published as a journal article in 1906, was entitled "Genius and Stupidity: A Study of Some of the Intellectual Processes of Seven 'Bright' and Seven 'Stupid' Boys." For his dissertation, Terman administered a variety of higher-order mental tests to seven boys identified by teachers as the "brightest" and seven boys identified as the "dullest," based upon a holistic review (i.e., not merely based on classwork) of willing boys. All of the boys were 10–13 years of age. Terman tested the boys for about 20–40 hours in each of eight areas: creative imagination, logical processes, mathematical ability, mastery of language, interpretation of fables, ease of learning to play chess, powers of memory, and motor abilities. Some tests were culled from the literature, including measures from Binet and Henri, Ebbinghaus, and others; other tests were tasks developed by Terman that would reappear in the Stanford–Binet. Terman found that the bright boys were superior to the dull boys in all but the motor tests, with creative imagination tests showing modest differences between bright and dull boys. Most tests administered tended to agree with one another—a finding that Terman interpreted as supporting the presence of Spearman's general factor. Bright children preferred to read, while dull children preferred to play games; there was little difference between the two groups in terms of persistence.

In 1910, Terman began his revision of the Binet–Simon Scale, a technical *tour de force* that would be published in 1916. Terman began initial studies by administering the 1908 Binet–Simon Scale to some 400 schoolchildren, as well as examining all available published studies of age-level placement for the Binet tests. It soon became evident that some tests were misplaced, with tests at the lower age levels too easy and those at the upper age levels too hard. He also wanted to add tests to reach six at each age level, eventually augmenting the Binet–Simon with 36 new tasks and clarifying administration and scoring criteria. Terman, his students, and his colleagues tested some 700 additional children in pilot studies. Some of Terman's new tasks were noteworthy, including a 100-word vocabulary test yielding full credit for correct definitions, half credit for partially correct definitions, and no credit for incorrect responses; and (arguably) the first executive function measure, the Ball and Field Test of Practical Judgment (see Littman, 2004, for an account of its origins). Terman and Childs (1912a, 1912b, 1912c, 1912d) published a "tentative revision and extension" of the Binet–Simon Scale, but further revision was necessary, given the 1911 extension of the Binet–Simon Scale through adulthood. As Seagoe (1975) reports, Terman's "unfamiliarity with statistics" and dislike of the "drudgery of computation" (p. 47) caused him to rely heavily on Arthur S. Otis, and for later editions on Truman L. Kelley and Quinn McNemar, for statistical analyses and data management. Dahlstrom (1985) notes the contribution of Otis's statistical knowledge and skills for the 1916 Stanford–Binet: "The efforts of Arthur S. Otis . . . were particularly important in this entire venture. He carried out the work on the item analyses and try-outs of the various early combinations of these items in tentative scales" (p. 76). Otis would later make important contributions to the development of the Army mental tests.

Terman's final standardization sample for the 1916 Stanford–Binet included 905 participants between the ages of 5 and 14 years, all within 2 months of a birthday and drawn from public schools in California and Nevada. No foreign-born or minority children were included. Special population studies included 200 "defective" and "superior" children. The adult sample consisted of 150 adolescent delinquents, 150 unemployed men, 50 high school students, and 30 businessmen across California and Oregon. The overall sample was predominantly white, urban, and middle-class, with an average adult mental age of 15–17 years. The final 1916 Stanford–Binet consisted of 90 items—six at each age level from ages 3 to 10; eight items at age 12; six items at age 14; and six items at each of two adult levels (average adult, superior adult). Sixteen alternative tests were available for tests that were inappropriate or otherwise spoiled. Of the final 90 items, 60% were drawn from the Binet–Simon and 40% from Terman and other sources. Terman adapted William Stern's (1914) "mental quotient" to generate the IQ (mental age divided by chronological age, with the product multiplied by 100 to remove decimals). Although Terman was critical of Spearman's work, he explicitly stated that the Stanford–Binet measured general intelligence, in effect making the single IQ score a functional estimate of Spearman's *g* and treating intelligence as a unitary construct:

> The scale does not pretend to measure the entire mentality of the subject, but only *general intelligence*. There is no pretence of testing the emotions or the will beyond the extent to which these naturally dis-

play themselves in the tests of intelligence. (Terman, 1916, p. 48; emphasis in original)

Terman retained Binet's adaptive testing format, which permitted flexibility in determining at which level to start the test, and different item types were intermixed to make the testing experience a fast-moving experience with tasks changing frequently.

Terman's Stanford–Binet was a resounding success, becoming the most frequently used psychological test (and intelligence test) in the United States for decades (Louttit & Browne, 1947). The Stanford–Binet would be renormed and expanded to create two parallel forms (Form L for Lewis, and Form M for coauthor Maud A. Merrill) spanning the ages 2 years through Superior Adult III in a remarkable 1937 revision (Terman & Merrill, 1937). The best items from the two forms would be retained in a single form for two updates (Terman & Merrill, 1960, 1973). From sales of test record forms, R. L. Thorndike (1975) estimated that the Stanford–Binet was administered to an average of about 150,000 persons a year from 1916 to 1937, to about 500,000 persons a year from 1937 to 1960, and to about 800,000 a year from 1960 to 1972. The fourth edition would make radical changes, including conversion to a point scale format and assessment of discrete factors of ability according to extended Gf-Gc theory (Thorndike, Hagen, & Sattler, 1986), but the fifth edition would endeavor to restore some of the features that distinguished the Stanford–Binet from its start (Roid, 2003).

Terman was also responsible, more than any other psychologist, for the rapid growth of intelligence and achievement tests in schools. The "Oakland experiment" of 1917–1918 was one of the first systematic attempts to use intelligence/ability tests as a basis for grouping students—a movement that is well documented in Chapman's *Schools as Sorters* (1988). Beginning in 1917, one of Terman's students, Virgil E. Dickson, became director of research for the Oakland Public Schools and organized the testing of 6,500 schoolchildren with the Stanford–Binet, the Otis Absolute Point Scale, and other tests in all of Oakland's 45 elementary schools. From his findings, Dickson concluded that many students fail because their ability levels make mastery of the ordinary curriculum impossible; furthermore, he asserted, the "mentally superior" are in need of accelerated curricula. Dickson called for segregation of students into special classes based on their ability levels. Figure 1.11 depicts the introduction of intelligence tests in

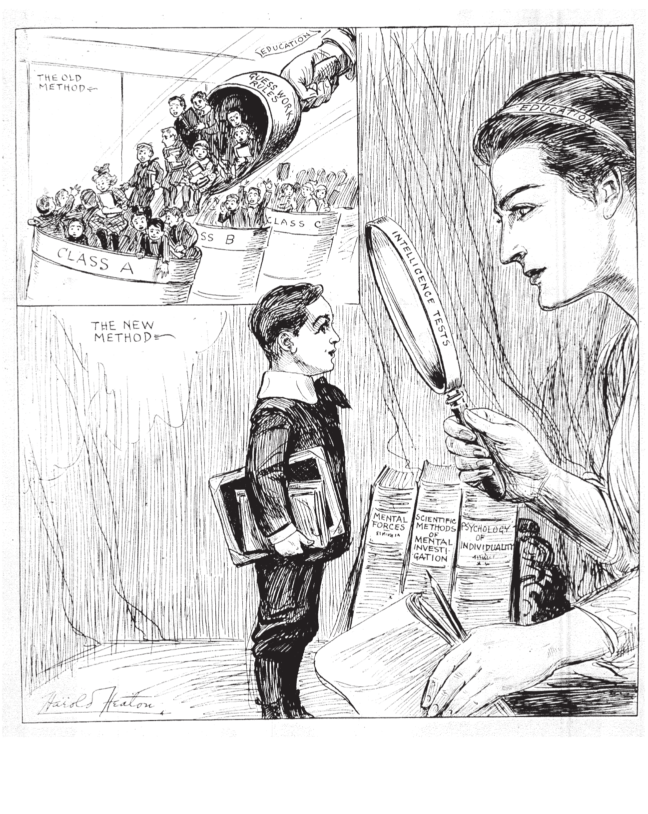


FIGURE 1.11. After the success of the "Oakland experiment" of 1917–1918, Terman and other psychologists advocated successfully for the use of intelligence tests to group students according to their ability levels. Educators recognized the value of measuring "individual differences" but were wary of the proliferating tests (e.g., Hines, 1922). From Heaton (1922). In the public domain.

the schools. Receiving enthusiastic endorsements from administrators and teachers, Dickson (1919) concluded:

> Standard tests, both psychological and pedagogical—group and individual—should be of great assistance in classification of pupils according to ability and capacity to do the work. They should inspire better teaching and better educational guidance through a more intimate knowledge of the individual child. (p. 225)

In 1923, Dickson published *Mental Tests and the Classroom Teacher*, the first in a series of "measurement and adjustment" books to be edited by Terman and published through the World Book Company. In 5 years, Terman would oversee nine additional titles, each focusing on problems of student testing and "adjustments to meet the problems of instruction and school administration arising out of individual differences" (see Chapman, 1988, p. 104, for a description of Terman's blueprint for the series).

Large-Scale Assessments and the Army Mental Tests

In retrospect, it was a remarkable accomplishment: In the span of only 18 months during World War I, a small team of psychologists developed, tried out, and then directed the administration of the first intelligence measures designed for large-scale adult testing. By the time the Armistice was signed in November 1918, an estimated 1,726,966 Army enlisted men and officers had been tested with the new group tests. More than 83,500 enlisted men were also given individual examinations. Although the military was not particularly appreciative of the testing program, psychologists used the perceived success of the Army mental tests to sell the general public on the value of mental testing; large-scale assessment thus found its way into American education system, where it remains prominent today. Accounts of World War I Army mental testing are available in official narratives from the psychologist directing the process (e.g., Yerkes, 1919, 1921; Yoakum & Yerkes, 1920), as well as a number of independent scholars (e.g., Camfield, 1969; Carson, 1993; Kevles, 1968; Napoli, 1981; Samelson, 1977; Sokal, 1987; Spring, 1972; von Mayrhauser, 1986, 1987, 1989). I draw on these sources and others for the following history.

The story of the Army mental tests begins with the United States' lack of preparation for the war. World War I, which started in 1914, was fought mainly in Europe between the Allied Powers (the Russian and British Empires, France, and later Italy and the United States) and the Central Powers (the Austro-Hungarian, German, and Ottoman Empires and Bulgaria). An isolationist United States, under the leadership of President Woodrow Wilson, sought neutrality in what was perceived as a European conflict, leaving the U.S. military unprepared to enter the war. As of April 1917, the strength of the U.S. Army was below 200,000 men, the smallest number since the Civil War (e.g., Yockelson, 1998).

Wilson finally asked Congress for a declaration of war against Germany on April 2, 1917 (Wilson, 1917). Congress declared war 4 days later. President Wilson signed the Selective Service Act into law on May 18, 1917; within a few months, 10 million men had registered for the draft, with almost 2.8 million men actually being drafted by the U.S. Army (Baker, 1918). Under General John J. Pershing, troops of the American Expeditionary Forces (AEF) began arriving in Europe in June 1917.

The draft, however, had no established procedures to identify and exclude men who were unfit for service. There was also no way to identify large numbers of potential officers, since the existing officers had been selected and trained through the U.S. Military Academy at West Point and fell far short of needs. Secretary of War Newton Baker (1918, p. 15) wrote that "one of the most serious problems confronting the War Department in April 1917, was the procurement of sufficient officers to fill the requirements of the divisions that were to be formed for overseas duty." Moreover, there was no systematic way to assign men to specialized military jobs similar to those they had held in civilian life (e.g., assigning a practicing accountant to military requisitions tracking or record keeping). The massive draft provided an opportunity for the young scientific discipline of psychology to demonstrate the value of its still-new technologies—the intelligence test and personnel selection procedures—to efficiently screen large numbers of enlisted men.

Yerkes and the Army Mental Tests

The involvement of psychologists in the war effort formally began on April 6, 1917, at the annual meeting of Edward B. Titchener's Society of Experimental Psychologists at Harvard University. When war was officially declared by the U.S. Congress on that day, Robert M. Yerkes, the president of the APA, asked the assembled psychologists how psychologists could assist the government in time of war. A committee was proposed under Yerkes's chairmanship, "to gather information concerning the possible relations of psychology to military problems" (Yerkes, 1921, p. 7). Almost 2 weeks later, on April 21, the executive council of the APA met in the Hotel Walton in Philadelphia. In preparation, Yerkes had been busy behind the scenes, touring Canadian hospitals, interviewing military doctors, and soliciting support from APA council members and members of the National Research Council. According to historian Richard T. von Mayrhauser (1987), Yerkes would use the military crisis to assert "near-dictatorial power within the profession" of psychology (p. 135).

The meeting at the Hotel Walton was misguided from the start because it involved a discussion among academic psychologists about what the military needed, rather than a request to the military as to how psychology might serve military needs. Moreover, a heavy-handed Yerkes sought to impose his narrow vision of mental testing on psychology, while suppressing input from another council member, Walter Dill Scott, who had more applied experience in personnel selection than anyone else at the meeting. With simultaneous authorization from the APA council and the National Research Council Psychology Committee, Yerkes appointed a dozen war-related psychology committees and chairs, dealing with areas such as aviation, recreation, propaganda, vision, acoustics, shellshock, emotional stability, and deception. Yerkes appointed himself chair of the "Committee on the Psychological Examining of Recruits," which was charged with preparation and standardization of testing methods and the demonstration of their effectiveness. Yerkes's initial testing plan—10-minute individual mental testing of at least 20% of "exceptional or unsatisfactory" recruits (von Mayrhauser, 1987, p. 141) by psychologists working under the supervision of military physicians—was in part a recapitulation of his own experiences working half-time directing research in the Psychopathic Department at Boston State Hospital under the supervision of Harvard psychiatrist Elmer Ernest Southard. At the same hospital, Yerkes had developed his own point scale adaptation of the Binet–Simon (Yerkes, Bridges, & Hardwick, 1915), which he probably hoped would be prominent in any testing program.

APA council member Walter V. Bingham later described his (and colleague Walter Dill Scott's) revulsion at events in Yerkes's meeting: "Meeting of the council in the smoke-filled room of a Philadelphia hotel. Midnight. Scott's utter disgust with the shortsighted self-interest revealed. His insurrection not previously told" (cited by von Mayrhauser, 1987, p. 139). Elsewhere in Bingham's papers appears the following disclosure:

> As the meeting proceeded it became clear to Scott and Bingham that Yerkes and the others were interested primarily in going into the army in order to acquire new psychological knowledge. They seemed to be more concerned with what the army could do for them than with what they could do for the army. Angrily, Scott and Bingham walked out in a huff. (cited by von Mayrhauser, 1987, p. 139)

With this divisive start, Yerkes alienated Scott, who had experience and skills he sorely needed. There was much at stake, as George Ellery Hale, famed astronomer and organizer of the National Research Council, warned Yerkes in May 1917:

> In the case of psychology, it is obvious that the first thing to do is to prove conclusively that the psychologists can perform service of unquestioned value to the government. . . . It is of fundamental importance that no tests be adopted which are not absolutely conclusive because if they were, the science of psychology would suffer an injury from which it would not recover for many years. (G. E. Hale to R. M. Yerkes, 1917; cited by Camfield, 1992, p. 107)

Yerkes's Committee and Arthur Otis

The Committee on the Psychological Examining of Recruits, made up of Robert M. Yerkes, Walter V. Bingham, Henry H. Goddard, Thomas H. Haines, Lewis M. Terman, F. Lyman Wells, and Guy M. Whipple, met at the Vineland Training School in New Jersey from May 28 to June 9 to develop the Army mental tests (see Figure 1.12). After reaching agreement that the tests had the goals of eliminating "unfit" recruits and identifying those with "exceptionally superior ability" (who might become officers), discussion turned to the merits of brief individually administered tests versus group-administered tests. Deciding that efforts should be made to test all recruits, the committee concluded that brief individual tests were problematic in terms of reliability and uniformity of method and interpretation, opting instead for group administration (Yerkes, 1921, p. 299). At this point, Lewis Terman presented the group-administered tests developed by his Stanford graduate student Arthur S. Otis. According to Yerkes (1921, p. 299), 4

FIGURE 1.12. Robert M. Yerkes's Committee on the Psychological Examination of Recruits at a 1917 meeting, during development of the Army mental tests at the Vineland Training School in New Jersey. Back row, left to right: Frederic Lyman Wells, Guy M. Whipple, Yerkes (Chair), Walter V. Bingham, and Lewis M. Terman. Front row, left to right: Edgar A. Doll (not a committee member), Henry H. Goddard, and Thomas H. Haines. Henry H. Goddard Papers, Archives of the History of American Psychology, Center for the History of Psychology, University of Akron. Reprinted by permission.

of the 10 tests in the original Army scale for group testing were accepted with little change from the Otis scale, and certain other tests were shaped in part by the content and format of the Otis series.

Committee members identified a dozen criteria to use for selection of additional tests: suitability for group use; interest and appeal; economy of administration time; score range and variability; scoring objectivity; scoring ease and rapidity; minimal writing requirements; resistance to coaching; resistance to malingering; resistance to cheating; independence from educational influences; and convergent validity with independent measures of intelligence. Each test was to consist of 10–40 items, with a time limit not to exceed 3 minutes. Moreover, oral directions needed to be simple, and written instructions easy to read. All tests needed to be accompanied by two or three completed sample items to ensure that examinees understood task requirements.

Psychologists around the country were recruited to write additional items to create 10 parallel equivalent forms of the Army mental tests. The tests underwent a series of pilot studies with 400 examinees drawn from different settings across the country. After revisions were made, a larger trial with the 10 forms was conducted on 4,000 recruits in Army and Navy settings during July and August 1917. The final test occurred in the fall of 1917, when 80,000 men in four national Army cantonments were tested, along with 7,000 college, high school, and elementary school students to check the Army results. All processing of record forms and statistical analyses were conducted by a small group working out of Columbia University, directed by Edward L. Thorndike with assistance from Arthur Otis and Louis L. Thurstone. Thorndike and his statistical analysis group endorsed the psychometric properties of the group tests, although clearly not all forms were equivalent, and some had to be dropped in the end.

Examination Beta was developed after Alpha, when it became evident that a different approach was needed for valid assessment of recruits who were either illiterate or limited in their English proficiency. It included ideas from Otis, Terman, and others and was tested at several training camps and at the Vineland Training School. After some 15 tests were reduced to 8 tests, the Beta was completed in April 1918. It was designed to correlate well with Examination Alpha, to differentiate average from very low levels of ability, and to be easily understood and administered, yielding few zero scores.

In December 1917, the Surgeon General recommended to the Secretary of War the continuance and extension of psychological examining to the entire Army. In January 1918, the Secretary of War authorized creation of a division of psychology in the Sanitary Corps out of the Surgeon General's office and expansion of the psychological examining program. A school for military psychology was organized with the Medical Officers Training Camp in Fort Oglethorpe, Georgia. While the school was active in 1918, approximately 100 officers of the Sanitary Corps and 300 enlisted men were given special training in military psychology. By the end of the war, psychological examining occurred at 35 army training camps and several army hospitals (Yerkes, 1919, 1920). Figure 1.13 shows a group administration of the Army mental tests.

Examinations Alpha and Beta

The Army Alpha was intended for fluent and literate English-language speakers. Alpha was typically administered to men who could read newspapers and write letters home in English, with at least a fourth-grade education and five years of residency in the United States (Yerkes, 1921, p. 76). The Army Beta was a largely nonverbal scale intended for examinees with inadequate English-language proficiency or illiteracy (Yoakum & Yerkes, 1920). Beta was also given to low scorers on the Alpha. Men who had difficulty reading or writing in English were to be given both Alpha *and* Beta. E. G. Boring (1961) described the informal process of separating recruits into those suitable for Alpha or Beta: "You went down the line saying 'You read American newspaper? No read American newspaper?'—separating them in that crude manner into those who could read English and take the Alpha examination and those who must rely for instructions on the pantomime of the Beta examination" (p. 30).

Examination Alpha consisted of eight tests, required approximately 40–50 minutes to administer, and could be given to groups as large as 500. A sample test from Alpha appears in Figure 1.14. Examinees were provided with the test form and a pencil. Responses were scored with stencils based upon examinee responses (which usually involved writing numbers, underlining, crossing out, or checking a selected answer). After illiterate and non-English-speaking examinees were removed, and all recruits were seated with pencils and test forms, the examiner said:

> *Attention! The purpose of this examination is to see how well you can remember, think, and carry out what you are told to do. We are not looking for crazy people. The aim is to help find out what you are best fitted to do in the Army. The grade you make in this examination will be put on your qualification card and will also go to your company commander. Some of the things you are told to do will be very easy. Some you may find hard. You are not expected to make a perfect grade, but do the very best you can.* (Yoakum & Yerkes, 1920, p. 53; emphasis added)

Beta was typically administered with task performance modeled through pantomimed demonstrations and some brief verbal directions (e.g., "Fix it!" while pointing to the incomplete pictures on Pictorial Completion; Yoakum & Yerkes, 1920, p. 87). A sample test from Beta appears in Figure

FIGURE 1.13. Group testing with the Army mental tests in a hospital ward, Camp Lee, Petersburg, Virginia, October 1917. Reproduced as Plate 5 (immediately after p. 90) in Yerkes (1921). Reprinted by permission of Time Life Pictures/Getty Images.

TEST 3

This is a test of common sense. Below are sixteen questions. Three answers are given to each question. You are to look at the answers carefully; then make a cross in the square before the best answer to each question, as in the sample:

SAMPLE { Why do we use stoves? Because
☐ they look well
☒ they keep us warm
☐ they are black

Here the second answer is the best one and is marked with a cross. Begin with No. 1 and keep on until time is called.

1 If plants are dying for lack of rain, you should
☐ water them
☐ ask a florist's advice
☐ put fertilizer around them

2 A house is better than a tent, because
☐ it costs more
☐ it is more comfortable
☐ it is made of wood

3 Why does it pay to get a good education? Because
☐ it makes a man more useful and happy
☐ it makes work for teachers
☐ it makes demand for buildings for schools and colleges

4 If the grocer should give you too much money in making change, what is the right thing to do?
☐ buy some candy of him with it
☐ give it to the first poor man you meet
☐ tell him of his mistake

5 Why should food be chewed before swallowing?
☐ it is better for the health
☐ it is bad manners to swallow without chewing
☐ chewing keeps the teeth in condition

6 If you saw a train approaching a broken track you should
☐ telephone for an ambulance
☐ signal the engineer to stop the train
☐ look for a piece of rail to fit in

7 If you are lost in a forest in the daytime, what is the thing to do?
☐ hurry to the nearest house you know of
☐ look for something to eat
☐ use the sun or a compass for a guide

8 It is better to fight than to run, because
☐ cowards are shot
☐ it is more honorable
☐ if you run you may get shot in the back

Go to No. 9 above

9 Why are warships painted gray? Because gray paint
☐ is cheaper than other colors
☐ is more durable than other colors
☐ makes the ships harder to see

10 Why should all parents be made to send their children to school? Because
☐ it prepares them for adult life
☐ it keeps them out of mischief
☐ they are too young to work

11 The reason that many birds sing in the spring is
☐ to let us know spring is here
☐ to attract their mates
☐ to exercise their voices

12 Gold is more suitable than iron for making money because
☐ gold is pretty
☐ iron rusts easily
☐ gold is scarcer and more valuable

13 The cause of echoes is
☐ the reflection of sound waves
☐ the presence of electricity in the air
☐ the presence of moisture in the air

14 We see no stars at noon because
☐ they have moved around to the other side of the earth
☐ they are so much fainter than the sun
☐ they are hidden behind the sky

15 Some men lose their breath on high mountains because
☐ the wind blows their breath away
☐ the air is too rare
☐ it is always cold there

16 Why do some men who could afford to own a house live in a rented one? Because
☐ they don't have to pay taxes
☐ they don't have to buy a rented house
☐ they can make more by investing the money the house would cost

FIGURE 1.14. The Army Examination Alpha Practical Judgment Test. Soldiers were allowed 1½ minutes for this test. From Yerkes (1921). In the public domain.

1.15. Administered to groups as large as 60, it was typically completed in about 50–60 minutes and required a blackboard with chalk, eraser, curtain, and chart (on a roller to show 27 feet of pictorial instructions in panels). The examiner gave the brief instructions, while a demonstrator pantomimed how to complete tasks correctly on the blackboard panels corresponding to the test response form. There were seven final tests in Beta.

Reports of intelligence ratings derived from test scores were typically made within 24 hours and entered on service records and qualification cards that were delivered to commanding officers and personnel officers. Individual examinations with the Yerkes–Bridges Point Scale, the Stanford–Binet Intelligence Scale, or the Army Performance Scale were usually reserved as checks on questionable or problematic group examination results.

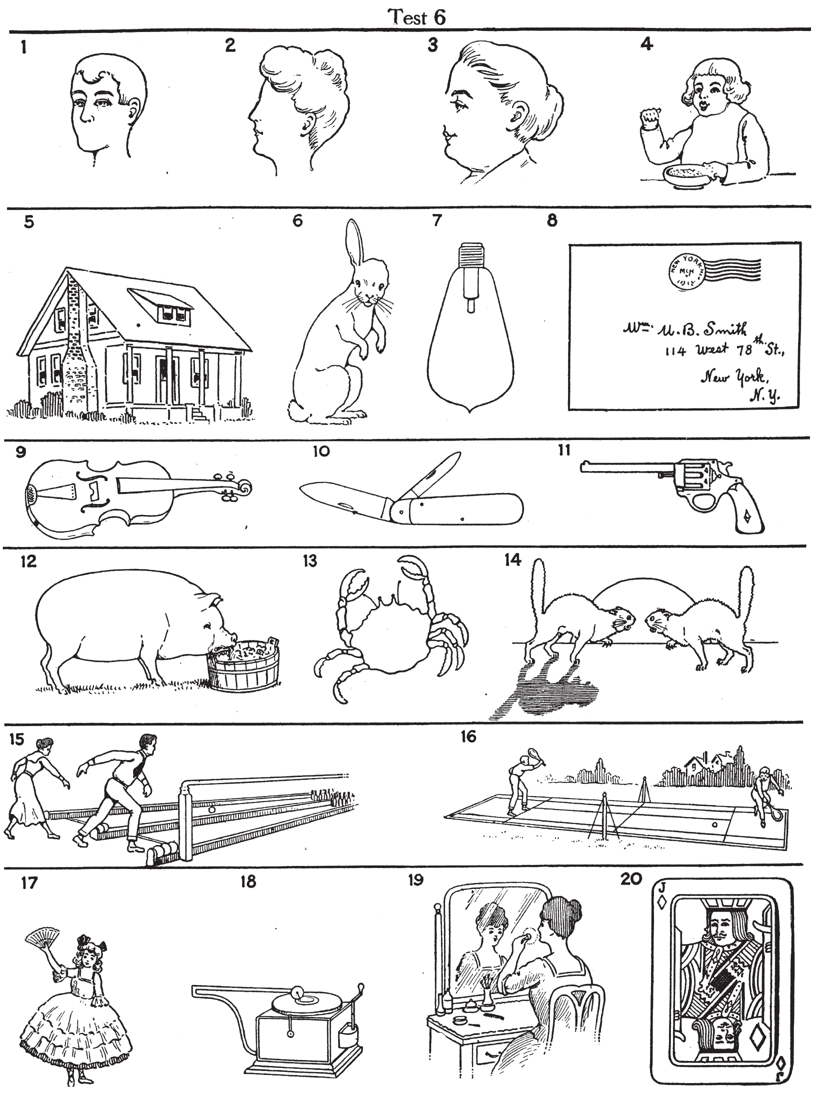


FIGURE 1.15. The Army Examination Beta Picture Completion Test. Instructions: "This is Test 6 here. Look. A lot of pictures . . . Now watch." Examiner points to separate sample at front of room and says to Demonstrator, "Fix it." After pausing, the Demonstrator draws in the missing part. Examiner says, "That's right." The demonstration is repeated with another sample item. Then Examiner points to remaining drawings and says, "Fix them all." Demonstrator completes the remaining problems. When the samples are finished, Examiner says to all examinees, "All right. Go ahead. Hurry up!" At the end of 3 minutes, Examiner says, "Stop!" From Yerkes (1921). In the public domain.

The test scores yielded grade ratings from A to E, with the following descriptions drawn from Yoakum and Yerkes (1920):

- *A (Very Superior).* An A grade was earned by only 4–5% of drafted men. These men were considered to have high officer characteristics when they were also endowed with leadership and other necessary qualities. They were shown to have the ability to make a superior record in college or university.
- *B (Superior).* A B grade was obtained by 8–10% of examinees. This group typically contained many commissioned officers, as well as a large

number of noncommissioned officers. A man with B-level intelligence was capable of making an average record in college.

- C+ *(High Average)*. The C+ group included 15–18% of all soldiers and contained a large number of recruits with noncommissioned officer potential and occasionally commissioned officer potential, when leadership and power to command were rated as being high.
- C *(Average)*. The C group included about 25% of soldiers who made excellent privates, with a certain amount of noncommissioned officer potential.
- C– *(Low Average)*. The C– group included about 20% of soldiers; these men usually made good privates and were satisfactory in routine work, although they were below average in intelligence.
- *D (Inferior)*. Men in the D group were likely to be fair soldiers, but they were usually slow in learning and rarely went above the rank of private. They were considered short on initiative and required more than the usual amount of supervision.
- *D– and E (Very Inferior)*. The last group was divided into two classes: D–, consisting of men who were very inferior in intelligence but who were considered fit for regular service, and E, consisting of men whose mental inferiority justified a recommendation for development battalion, special service organization, rejection, or discharge. The majority of men receiving these two grades had a mental age below 10 years. Those in the D– group were thought only rarely able to go beyond the third or fourth grade in primary school, however long they attended.

To his chagrin, Yerkes's division of psychology was appointed to the Sanitary Corps instead of the Medical Corps (where he had hoped psychologists would be classified), but he was still a member of the Surgeon General's staff. Yerkes encountered near-continual resistance to the testing program from the military establishment, and the Army mental examiners often had inadequate testing facilities or faced a deeply entrenched military establishment that did not see the value in intelligence tests. In response to queries about testing from Army officers, Yerkes gave the psychological examiners standard responses to provide as needed—specifying the potential value of the Alpha and Beta in military decision making, but also emphasizing that test scores alone should not constitute the sole basis for making military service decisions:

> The rating a man earns furnishes a fairly reliable index of his ability to learn, to think quickly and accurately, to analyze a situation, to maintain a state of mental alertness, and to comprehend and follow instructions. The score is little influenced by schooling. Some of the highest records have been made by men who had not completed the eighth grade. . . . The mental tests are not intended to replace other methods of judging a man's value to the service. It would be a mistake to assume that they tell us infallibly what kind of soldier a man will make. They merely help to do this by measuring one important element in a soldier's equipment, namely, intelligence. They do not measure loyalty, bravery, power to command, or the emotional traits that make a man "carry on." (Yoakum & Yerkes, 1920, pp. 22–24)

According to Yerkes (1918a, 1918b), the Army testing program was tasked with four military objectives: (1) aiding in the identification and elimination of "mentally defective" men who were unfit for service; (2) identifying men of exceptional intelligence for special responsibilities or possible officer training; (3) balancing military units in terms of intelligence; and (4) assisting personnel officers in the camps with the classification of men. The tests appear to have functioned well in identifying recruits of very high intelligence and very low intelligence, although research findings showed a disproportionate number of minority, foreign-born, and illiterate recruits as having very low intelligence, in spite of efforts to correct for the language and literacy demands of the Alpha with the Beta (Yerkes, 1921). There is little evidence that the Alpha and Beta were effectively used to balance the intellectual composition of military units. Although Army battalions ideally should be comparable and interchangeable in terms of effectiveness, individual battalion commanders no doubt wanted the best available recruits and held onto the recruit who received A and B grades. Yoakum and Yerkes (1920) described the challenge:

> In making assignments from the Depot Brigade to permanent organizations it is important to give each unit its proportion of superior, average, and inferior men. If this is left to chance there will inevitably be "weak links" in the army chain. Exception to this rule should be made in favor of certain arms of the service which require more than the ordinary number of mentally superior men; for example, Signal Corps, Machine Gun, Field Artillery and Engineers. These organizations ordinarily have about twice the

> usual proportion of "A" and "B" men and very much less than the usual proportion of "D" and "D–" men. (p. 25)

With respect to the final objective, of assisting with personnel decisions, the Army mental tests provided a single piece of information—intelligence level—that was of considerable value. It would be Walter Dill Scott's Army Committee on Classification of Personnel that provided a context for the Army mental test scores in making military personnel decisions.

Scott's System of Personnel Selection

Walter Dill Scott (1869–1955) was a pioneering industrial psychologist, applying principles of experimental methodology to practical business problems (Ferguson, 1962, 1963a; Lynch, 1968; Strong, 1955). An interest in identifying and selecting successful salesmen led Scott (1916) to develop a multimethod quantitative personnel selection approach consisting of historical information from former employers (i.e., a model letter soliciting information and ratings, which was included in a *personal history record*); performance on tests of intellectual ability devised by Scott; performance on tests of technical ability (written calculation and clerical transcription) scored for accuracy, speed, and legibility; and multiple ratings based on a series of "interviews" with trained raters (in which the examinee was to introduce himself and try to sell merchandise to a series of interviewers posing as merchants). In 1916, Walter V. Bingham, the head of the division of applied psychology at Carnegie Institute of Technology, offered Scott the opportunity to direct the newly formed Bureau of Salesmanship Research and to become the first professor of applied psychology in the United States (Ferguson, 1964a). In a remarkable partnership between the Carnegie Bureau and 30 large national businesses, Scott had the opportunity to test his personnel selection methods with the hiring of 30,000 new salesmen each year, and comparison of personnel decisions against actual sales performances. It was a highly successful arrangement, possibly unprecedented in the history of psychology, and Scott's work was well regarded by the business community.

The history of Scott's personnel selection system in the military may be found in several resources, including official accounts from the Army (Committee on Classification of Personnel in the Army, 1919a, 1919b) and contemporary accounts from von Mayrhauser (1987, 1989); the most in-depth accounts are available from Ferguson (1963b, 1963c, 1964b, 1964c). When war was declared in 1917, Scott realized that his existing methods could readily be applied to personnel selection in the military. At the Hotel Walton meeting on April 21, Scott objected to Yerkes's positions on the war as an opportunity to advance the prominence of psychology. Scott and Bingham were the only psychologists at the meeting with experience in personnel selection, and they knew that Scott's system already had demonstrated effectiveness. In Scott's system, the mental test results had value, but Scott and Bingham were certain that interview ratings would be more important in the selection of officers. Moreover, Scott did not want to subordinate psychologists to psychiatrists, but instead thought they should report to a high official such as the Secretary of War. Offended by Yerkes's self-serving agenda, Scott and Bingham walked out.

Scott decided to launch his own initiative, independent of Yerkes. Scott revised his existing salesman rating scale, completing A Rating Scale for Selecting Captains by May 4, 1917. He shared with it several psychologists and asked Edward L. Thorndike to write a letter of support to Frederick P. Keppel, who had been a dean at Columbia and was now Third Assistant Secretary of War. Keppel invited Scott to Washington, D.C., where Scott presented his scale, did some testing with it, made some improvements, and overcame institutional resistance (including having his scale ripped "to tatters" by officers in Plattsburg who had been invited to suggest improvements [Committee on Classification of Personnel in the Army, 1919a, p. 50]). When he finally met directly with Secretary of War Newton D. Baker, Scott suggested that a group of psychologists and experienced employment managers be appointed to advise the Army on personnel selection, volunteering to assemble such a group. On August 5, 1917, Scott received approval of a plan to include scientific staff, a group of civilian experts for research and planning, and a board of military representatives to bring problems to the Committee on Classification of Personnel in the Army and help implement its recommendations. Within 6 weeks, the committee created and implemented a classification and assignment system for the Army where none had existed before. It was the largest program of personnel selection ever attempted to that time. Scott was the committee's director, Bingham was its executive secretary, and they answered directly to the Adjutant

General of the Army. They began with a single office that grew to 11 rooms in the War Building (then the central hub of military decision making, housing the offices of the Secretary of War, Chief of Staff, and Adjutant General).

Scott's personnel system for the army included a Soldier's (or Officer's) Qualification Card, grades on the Army mental tests and proficiency on specialized trade tests, and the Officers' Rating Scale in various forms for noncommissioned and commissioned officers. The Qualification Card relied on interviews to obtain occupational history, education, leadership experience, and military history. Test scores on the Army mental tests ranged from A through E and were provided by Yerkes's examiners. For recruits claiming experience in specific trades of value to the military, Scott's committee oversaw development of special trade tests that measured specific proficiencies, generating a range of scores from "Expert" through "Novice." Finally, the Officers' Rating Scale became the main tool used for the selection and promotion of officers, with all officers receiving quarterly ratings by the end of the war. This scale involved ratings in five areas: physical qualities, intelligence, leadership, personal qualities, and general value to the service.

If the Army mental tests were intended to measure general intelligence, the trade tests measured specific ability and knowledge related to the performance of several hundred specific occupations needed by the military. Vocational training was impractical, and men were frequently found to have misrepresented their civilian jobs and skills on the Soldier's Qualification Card. In order to identify personnel requirements for specific jobs, occupational titles were compiled and detailed personnel specifications were prepared by Scott's team. With the criteria of covering all trades rapidly and objectively by examiners who did not have to be knowledgeable about each individual trade, a series of oral, picture, and/or performance trade tests were administered and scored so that the number of questions correctly answered predicted status as a novice, apprentice, journeyman, or expert. There were 84 trade tests for jobs as varied as butchers, electricians, pipefitters, and most other specialties needed by the military. For example, the trade test officer issued driver's licenses for all drivers of touring cars, motorcycles, and trucks (Committee on Classification of Personnel in the Army, 1919a, p. 135). Examples of trade tests appear in the committee's personnel manual (1919b), and after the war compilations of trade tests were published in Chapman (1921) and Toops (1921).

From the military's perspective, it is clear that Scott's personnel selection procedures were much more valued than Yerkes's testing program. At the end of the war, Yerkes's Division of Military Psychology was summarily and completely shut down. Scott's Committee on Classification of Personnel in the Army was transferred to the General Staff and merged with the newly created Central Personnel Branch, in effect institutionalizing Scott's personnel selection procedures within the Army (Yerkes, 1919). The War Department awarded Scott the Distinguished Service Medal when he left the service in 1919, and asked *him* to convey its appreciation to Major Yerkes. Scott became the highest-ranking psychologist in the Army, having been commissioned as a colonel in the Adjutant General's Department in November 1918.

Undoubtedly multiple factors explained the military's different responses to Scott and to Yerkes. Scott adapted his system to military needs, while Yerkes sought to impose academic knowhow on an unreceptive army. Scott partnered with military personnel, while Yerkes's examiners were seen as unwelcome, externally imposed "pests" and "mental meddlers" by camp commanders (cited by Kevles, 1968, p. 574). No less than three independent investigations were launched by Army personnel, suspicious of Yerkes and his men (Zeidner & Drucker, 1988, p. 11). Scott worked initially in an advisory capacity, while Yerkes continually sought authority. Scott had considerable personal skills in persuasion and salesmanship (Strong, 1955), whereas Yerkes was a strong planner but a poor manager (Dewsbury, 1996). Scott's system had substantial and understandable face validity for military performance, while Yerke's Examinations Alpha and Beta did not have obvious relevance for soldiering. From the perspective of the history of intelligence testing, however, a broader argument should be considered: Yerkes's committee created the tests and his examiners generated scores, but Scott's committee provided a systematic context (including recruits' history and specific skills) within which the test scores made sense and could be used to make practical decisions by teams of military personnel not schooled in psychology.

World War II Assessment Procedures

In World War II, the plan developed by Scott was streamlined and implemented again, this time with Walter V. Bingham in charge of the personnel system and mental tests (Bingham, 1942, 1944). Bingham served as chair of the Committee on Classification of Military Personnel, the committee having been appointed in 1940 by the National Research Council at the request of the Adjutant General, months before passage of the Selective Service and Training Act (Bingham, 1944). In contrast to the unwelcome reception Yerkes had received in World War I, Bingham and the infrastructure he established were valued by the Army (Zeidner & Drucker, 1988). The Army Alpha was replaced by the Army General Classification Test, a shorter version of the Alpha; initial versions were administered in spiral omnibus form in about 40 minutes, and there were four parallel forms (Bittner, 1947; Staff, Personnel Research Section, 1945). Conceptualized as a test of "general learning ability," it consisted of vocabulary items (intended to tap verbal comprehension), arithmetic problems (thought to tap quantitative reasoning), and block-counting problems (intended to measure spatial thinking), all endeavoring to deemphasize speed somewhat. Grades of A to E were replaced with five levels of learning readiness, I to V. Terms like *mental age* and *IQ* were largely eliminated from group tests.

As in World War I, specialized Non-Language Tests were developed and standardized to test illiterate and non-English-speaking recruits (Sisson, 1948). In 1944, the Wechsler Mental Ability Scale, also known as the Army Wechsler, was replaced by the Army Individual Test, which included three verbal subtests (Story Memory, Similarities–Differences, and Digit Span) and three nonverbal subtests (Shoulder Patches, Trail Making, and Cube Assembly) (Staff, Personnel Research Section, 1944). Rapaport (1945) praised the Army Individual Test, noting that it was "admirably well-constructed" (p. 107) as a measure of general mental ability, but he also raised cautions about its diagnostic limitations. Numerous specialized trade tests and aptitude tests (e.g., mechanical aptitude, clerical aptitude) were developed as well. Most importantly, the Personnel Research Section of the Adjutant General's Office that Bingham established quickly earned the military's trust, leading to the creation of the Army Research Institute (which still exists). One of Bingham's charges was to put the "brakes on projects . . . of great scientific interest" if they did not help the Army "toward early victory" (Zeidner & Drucker, 1988, p. 24). It was a lesson in military priorities that Yerkes, whose agenda included advancing psychology as a science, may not have learned in World War I.

David Wechsler: The Practical Clinician

The practice of contemporary applied intelligence assessment in the second half of the 20th century may arguably be said to have been most strongly and directly influenced by the measurement instruments developed by David Wechsler (1896–1981). Beginning in the 1960s, the Wechsler intelligence scales supplanted the Stanford–Binet as the leading intelligence tests (Lubin, Wallis, & Paine, 1971). Surveys of psychological test usage decades after his death show that Wechsler's intelligence tests continue to dominate intelligence assessment among school psychologists, clinical psychologists, and neuropsychologists (Camara, Nathan, & Puente, 2000; Wilson & Reschly, 1996). The Wechsler scales for adults, children, and preschoolers are taught at much higher frequencies than any other intelligence tests in North American clinical and school psychology training programs (Cody & Prieto, 2000).

In many ways, David Wechsler was an unexpected success—coming to the United States as a child amid a flood of Eastern European immigrants, losing both parents by the age of 10, compiling a relatively ordinary academic record in high school and college (while graduating early), registering as a conscientious objector to the 1917 World War I draft (a risky decision at the time, when "slackers" were universally condemned), and not having become a naturalized citizen by the time of the war. Even so, these risk factors may have been somewhat ameliorated by the guidance of an accomplished older brother (pioneering neurologist Israel S. Wechsler), who became his caretaker and role model; by the opportunity to provide military service as an Army mental test examiner, thereby quickly learning about assessment and making key professional contacts; and by receiving his graduate education and professional psychology training at an opportune time and place in the development of what eventually would become "clinical" psychology.

Wechsler's Early Life and Education

David Wechsler was the youngest of three boys and four girls born in Romania to Moses Wechsler, a merchant, and Leah (Pascal) Wechsler, a shopkeeper (see, e.g., Matarazzo, 1972). At the time the Wechsler family emigrated in 1902, poor harvests in 1899 and 1900 had produced famine and an economic downturn in Romania, worsening the scapegoating of Jews and resulting in severe applications of existing anti-Jewish decrees (e.g., Kissman, 1948). The family's new life on the Lower East Side of New York City was marked by tragedy. Within 5 years of their arrival both Moses and Leah Wechsler passed away from malignancies ("Deaths reported Aug. 23," 1903; Wechsler, 1903; Wexler, 1906). The effects of these losses upon the family, particularly David as the youngest, are likely to have been profound. By 1910, David's older brother Israel S. Wechsler, then a physician in general practice, appears to have taken over as the head of the family.

Wechsler was educated in the public schools on the Lower East Side (see Wechsler, 1925, p. 181). After high school graduation, Wechsler attended the College of the City of New York (now known as City College) from 1913 to 1916, graduating with an AB degree but without honors at the age of 20 ("206 get degrees at City College," 1916). Following his graduation, Wechsler enrolled in graduate studies in psychology at Columbia University, where he would complete his master's degree in 1917 and his doctorate in 1925. His decision to continue his education beyond college had family precedent; his older brother, Israel, had graduated from New York University and Bellevue Medical College in 1907 at the age of 21. Israel would take a position in neurology at Mount Sinai Hospital in 1916 and begin teaching neurology at the outpatient clinic of the Columbia University College of Physicians and Surgeons in 1917 (Stein, 2004; see also "Israel Wechsler, Neurologist, Dies," 1962). Israel initially intended to become a psychiatrist and was self-taught in psychoanalysis, but personally identified as a neurologist because, as he later explained, "If the brain is the organ of thought and disturbance of its functions expresses itself in disorders which are called neuroses and psychoses, it seemed reasonable and necessary to know neurology" (Wechsler, 1957, pp. 1113–1114). Israel Wechsler was involved in the early-20th-century struggles between psychiatry and neurology, when each medical discipline was vying for control of the care of those with mental illness. David instead pursued psychology, but he would follow his brother's lead in becoming a practicing clinician, a hospital-based academic, and the author of professional textbooks.

Columbia was one of the few major universities that provided graduate experimental psychology training with a willingness to address applied problems, termed *experimental abnormal psychology* by Woodworth (1942, p. 11)—an educational orientation that would eventually evolve into clinical psychology (Routh, 2000). The Columbia graduate psychology department in this era was made up primarily of commuter students "who emerged from the subway for their classes and research and departed immediately into the subway afterwards" (Thorne, 1976, p. 164). Columbia University was the academic home of faculty J. McKeen Cattell (until his dismissal in October 1917), Robert S. Woodworth, and Edward L. Thorndike, three of the most influential psychologists of the early 20th century. "Cattell, Woodworth, and Thorndike were the trio at Columbia," said F. L. Wells, who had worked as an assistant to Cattell and Woodworth, adding, "Cattell might inspire awe, Thorndike admiration, and Woodworth affection. Affection toward Cattell and Thorndike was not possible" (quoted in Burnham, 2003, p. 34).

For his master's thesis, completed (and published as a journal article) in 1917, Wechsler (1917a) patched together a clinical memory battery from existing published and unpublished tests, closely following the memory framework suggested by Whipple (1915). Wechsler spent 2½ months conducting in-depth assessment of six patients with Korsakoff psychosis at the Manhattan State Hospital on Ward's Island. He saw each patient as many as 20 times, and he also had the opportunity to observe psychiatric assessment on the wards. Wechsler's master's thesis represented his first known attempt to build a test battery. He established a pattern he was later to follow with intelligence tests, memory tests, and a failed personality test: that of appropriating practical and clinically useful procedures from other authors, making slight improvements and modifications, and synthesizing them into a battery of his own.

World War I Service

After the U.S. Congress declared war on Germany in April 1917, Wechsler (1917b) completed his required registration for the draft, listing himself as a "Consciencious [*sic*] Objector" and as an alien who was a citizen of Romania, who was disabled by "Near Sightness [*sic*]" and "Physical Unfitness." To the item asking about his occupation, he wrote "Am student in school of philosophy." Wechsler's draft registration thus used multiple methods to avoid being drafted—claiming status as a conscientious objector, claiming exemption from military service by reason of alien (noncitizen) status, and claiming physical deficiencies that would disqualify him for military service. We do not know what motivated David Wechsler to try to avoid military service at age 21, but the public press treated conscientious objectors with contempt, and some were arrested and imprisoned. Even Army mental examiners considered conscientious objector status to be a form of psychopathology (May 1920). Wechsler's status as a noncitizen native of Romania, some 15 years after his arrival in the United States, also put him at risk. As an alien, he could not be drafted, but he could be deported. The U.S. Congress tried to close this draft loophole in response to the perceived "alien slacker" problem, but treaty obligations circumvented final passage ("Alien Slackers May Not Escape Service," 1917; "Pass Alien Slacker Bill," 1918). Within military training camps, however, some officers considered all aliens who had not become naturalized citizens as suspect.

Becoming an Army mental test examiner represented a way by which Wechsler could avoid seeing combat, and it was probably through back-channel communications from his professor Robert S. Woodworth to Robert M. Yerkes that Wechsler was identified as a prospective tester. In May 1918, Yerkes requested in writing that Wechsler and 13 others who had "qualifications for psychological service" be sent authorization for military induction and be assigned for course instruction in military psychology at Camp Greenleaf, Chickamauga Park, Georgia. Shown in Figure 1.16 at the time of his military service, Wechsler reported to the School for Military Psychology, where he was taught the Army Alpha, Army Beta, Stanford–Binet, Yerkes Point Scale, and other tests. Trainees also received instruction in military practices, including military law, field service, honors and courtesies, equipment, and gas attack defense instructions and drills. E. G. Boring, who reported to Camp Greenleaf as a captain in February 1918, described what may have also been Wechsler's experience:

> We lived in barracks, piled out for reveillé, stood inspection, drilled and were drilled, studied testing procedures, and were ordered to many irrelevant lectures. As soon as I discovered that everyone else resembled me in never accomplishing the impossible, my neuroses left me, and I had a grand time, with new health created by new exercise and many good friendships formed with colleagues under these intimate conditions of living. (Boring, 1961, p. 30)

In May 1918, Congress enacted legislation that allowed aliens serving in the U.S. armed forces to file a petition for naturalization without having made a declaration of intent or proving 5 years' residence (e.g., Scott, 1918). Under this new law, Wechsler became a naturalized citizen in June 1918, with Captain John E. Anderson and Lt. Carl A. Murchison, two psychologists who would have noteworthy careers, serving as his witnesses (Wechsler, 1918b). Wechsler completed his training at Camp Greenleaf in July, was promoted to the rank of corporal, and was assigned to Camp Logan in Houston, Texas, in early August 1918. There he would give individual psychological assessments to recruits who had failed the Alpha and/or the Beta, largely because of limited English proficiency or illiteracy. Conditions at Camp Logan were poor, with inadequate space and support, but the Army examiners administered over 300 individual assessments (Yerkes, 1921, p. 80).

FIGURE 1.16. David Wechsler at the age of 23, from his 1918 passport application. Wechsler used a program designed to educate World War I veterans in Europe to pursue educational opportunities in France and London, including time with Charles E. Spearman and Karl Pearson. From Wechsler (1918b). National Archives and Records Administration, Washington, D.C. In the public domain.

It was during his time as an Army examiner that many of Wechsler's core ideas about assessment were born, especially his idea to construct an intelligence scale combining verbal and nonverbal tests, paralleling the Army Alpha and Army Beta/performance exams (Wechsler, 1981). Most of the assessment procedures appropriated by Wechsler for his intelligence scales appear in Yerkes (1921). Matarazzo (1981) relates that Wechsler realized the value of individual assessment when group tests yielded misleading results, as many of his examinees functioned adequately in civilian life in spite of their low group test scores. Wechsler also reportedly learned the value of nonverbal assessment and the limitations of the Stanford–Binet with adults. He even (Wechsler, 1932), described an approach to profile analysis of Army Alpha subtests—a clear antecedent to the intraindividual (ipsative) profile analyses still used in interpreting the Wechsler intelligence scales.

With the signing of the armistice, Wechsler participated in AEF University, a program created by order of General John J. Pershing and other military leaders to serve the 2 million idle (and bored) American servicemen who remained stationed in Europe, waiting to be shipped home (Cornebise, 1997; "Education for American Soldiers in France," 1919). Although Wechsler had never served overseas, he arranged to spend time in France (December 1918 to March 1919) and then in London (March 1919 through July 1919) as part of this program. Some 2,000 soldiers attended the Sorbonne, while about 2,000 soldier-students attended British universities, with 725 going to University College London ("U.S. Maintains Great Schools on Foreign Soil," 1919). At University College London, Wechsler had the opportunity to work for 3 months with Charles E. Spearman and to meet Karl Pearson, becoming familiar with Spearman's work on the general intelligence factor and Pearson's correlation statistic, as well as to note their professional rivalry (Wechsler, Doppelt, & Lennon, 1975). Wechsler was honorably discharged from the military in July 1919. Given his efforts to avoid military service in 1917, it might be considered ironic that the skills he acquired and contacts he made during his military service would shape his career in assessment and test development.

From 1919 to 1921, Wechsler studied and conducted research at the University of Montpelier and principally at the Sorbonne, under the supervision of Henri Pieron and Louis Lapicque (Rock, 1956, p. 675; Wechsler, 1925, p. 8). Wechsler used the research to complete his doctorate at Columbia, under the guidance of Robert Woodworth (Wechsler, 1925, p. 8). The opportunity to study at the Sorbonne came through Wechsler's application for an American Field Service fellowship from the Society for American Fellowships in French Universities (Wechsler, 1918b). In its first year (1919–1920), there were eight fellows, one of which was Wechsler.

Bellevue Psychiatric Hospital and Other Clinical Experiences

After completing his fellowship at the Sorbonne, Wechsler traveled through France, Switzerland, and Italy before reluctantly returning to the United States (Wechsler, 1921). Once he was settled in New York, he began practicing psychology, mostly conducting assessments, in a variety of clinical and industrial settings. His ambivalence about returning, as disclosed to Edwards (1974), was reflected in his 1922 paper on the psychopathology of indecision.

Wechsler spent the summer of 1922 working with F. L. Wells at the Psychopathic Hospital in Boston, followed by 2 years as a psychologist with the New York Bureau of Children's Guidance. The Bureau of Children's Guidance was a psychiatric clinic, operating under the aegis of the New York School of Social Work and reflecting the values of the popular child guidance movement. Directed by Bernard Glueck, the bureau served troubled children referred by school principals or selected teachers for problems in the areas of scholarship, attendance, behavior, or general welfare. It was staffed by social workers, psychiatrists, and psychologists. The bureau emphasized problems with delinquency, with the objective of "a keener understanding of the child as an individual, and assistance to the school in working out needed readjustments, whether they be physical, social or educational" ("Crime Clinics Growing," 1922).

From 1925 to 1927, Wechsler worked with J. McKeen Cattell as acting secretary and research associate of The Psychological Corporation (Wasserman & Maccubbin, 2002). Created by Cattell, The Psychological Corporation did not directly employ any psychologists at the time; instead, consulting psychologists worked in nonsalaried, commission-based arrangements, undertaking projects for businesses and dividing the payment between themselves and the corporation. A 29-year-old David Wechsler, having completed his dissertation, had difficulty finding a job and contacted his old professor, Cattell, who hired him; according to Wechsler, Cattell told him, "You can get the *pro tem* acting secretary here. You have to get your own business and whatever business you get, the company will get half of your remunerations" (Wechsler et al., 1975). Wechsler undertook two known projects at The Psychological Corporation: the development of an automobile driving simulator and psychometric tests for taxicab drivers (Wechsler, 1926), and a tabloid newspaper study with a *New York World* reporter to test the intelligence of Ziegfeld chorus girls with the Army Alpha.

In 1932, following the tragic death of two Bellevue Hospital staff psychologists in a boating accident ("*Sea Fox* Wreckage," 1931), Wechsler was hired as a psychologist by the Psychiatric Division of Bellevue Hospital, New York. Bellevue was the oldest public hospital in the United States, but its psychopathic wing was scheduled for replacement by the Bellevue Psychiatric Hospital, described at its groundbreaking as the "chief battle-ground in the war against diseases of the mind" ("Old Bellevue and New," 1930). When the new unit finally opened in 1933, its capacity was planned at 600 patients to "give wide scope and facility for modern methods of investigating and treating mental disorders" ("A Bellevue Unit Formally Opened," 1933). By 1941, Wechsler had become chief psychologist and a clinical faculty member at the New York University College of Medicine, supervising more than 15 clinical psychologists, five interns, and two research psychologists on grants (Weider, 2006). Wechsler would retire from Bellevue in 1967, after having pioneered the role of the psychologist in a psychiatric hospital (Wechsler, 1944), and his clinical experiences would help him remain oriented to the use of psychological testing as it relates to practical patient care.

Concept of Intelligence

In his earliest scholarly statement on intelligence in his brother's neurology book, Wechsler (1927) ventured a definition: "All definitions of intelligence refer essentially to ability to learn and adapt oneself to new conditions; that is, not knowledge and practical success, but ability to acquire knowledge and ability to cope with experience in a successful way" (p. 105). It is Wechsler's (1939) definition, which built on his previous efforts and borrowed elements from his predecessors, that remains best known among definitions of intelligence:

> Intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment. It is global because it characterizes the individual's behavior as a whole; it is an aggregate because it is composed of elements or abilities which, though not entirely independent, are qualitatively differentiable. By measurement of these abilities, we ultimately evaluate intelligence. But intelligence is not identical with the mere sum of these abilities, however inclusive. (p. 3)

The long-standing popularity of this definition is probably due to the enduring popularity of the Wechsler intelligence scales with which it is associated. The definition reflects Wechsler's generally cautious writing style; it was exceptionally rare that he made any bold statement in writing that might alienate any colleagues. The phrase "aggregate or global capacity" appears to encompass Spearman's general factor, *g*—but Wechsler included an accommodation for the group factors, which, "though not entirely independent, are qualitatively differentiable." According to Wechsler (Wechsler et al., 1975), this definition also subsumes Binet's emphasis on adaptation. The phrase "to deal effectively with his environment" recapitulates Binet's (1911/1916) observation that "Intelligence marks itself by the best possible adaptation of the individual to his environment" (p. 301), as well as the use of adaptation in the definition of intelligence by others. In one of his final publications, Binet (1910) also took the position that intelligence is a dynamic synthesis, more than the different "pieces of the machine" that comprise it; this may have influenced Wechsler's statement that intelligence is more than the sum of its constituent abilities.

Creation and Development of the Wechsler Intelligence Scales

Of course, it is for his intelligence tests that David Wechsler is best remembered. Wechsler's gifts in the area of test development lay in his ability to synthesize the work of others—that is, to recognize clinically useful measurement procedures and to streamline and package them so as to be maximally useful for the practicing psychologist. His test work was unoriginal, and his intelligence tests consist entirely of tests (sometimes incrementally improved) that were originally devised by other psychologists. Several researchers have sought to trace the origins of the specific Wechsler intelligence subtests (e.g., Boake, 2002; Frank, 1983), a historically important endeavor, but it is notable that from the start Wechsler (1939) openly disclosed the sources he drew upon. As Boake (2002) suggested, it is most unfortunate that the names of the original innovators who created the Wechsler subtest procedures have been forgotten, omitted from mention in contemporary test manuals.

The Bellevue Intelligence Scale was originally subsidized by a Works Progress Administration grant during the Great Depression (Wechsler, 1981; Wechsler et al., 1975). Wechsler (1939, p. 137) reported that the test took 7 years to develop, and it first underwent trials in 1937 and 1938 at the Bellevue Psychiatric Hospital, the Court of General Sessions of New York City, and the Queens General Hospital. The need for a new adult test stemmed largely from the inadequacy of the Stanford–Binet, particularly its poor normative sample for adults, and the poor fit of the Army mental tests for clinical decision making. As the chief psychologist in a large public hospital, Wechsler had the opportunity to appreciate the needs and applications for an adult intelligence test. After careful review, Wechsler essentially cherry-picked his subtests from the most clinically useful and psychometrically adequate tests of his era; he thus provided practitioners with an easy transition to make from using many separate, independently normed tests with a variety of instructions and scoring rules to a single battery of co-normed tests, with streamlined administration and fairly uniform scoring rules. He acknowledged, "Our aim was not to produce a set of brand new tests but to select, from whatever source available, such a combination of them as would best meet the requirements of an effective adult scale" (Wechsler, 1939, p. 78). Most of the standardization sample of 1,586 participants was collected in the city and state of New York; the sample was stratified by age, sex, education, and occupation, but was limited to English-speaking white examinees.

The Bellevue consisted of 10 subtests, with the Vocabulary subtest serving as an alternate. With the exception of a single speeded subtest (Digit Symbol), items on each subtest were sequenced in approximate order of difficulty, from easiest to hardest. Performance on the first five subtests contributed to the Verbal IQ, and performance on the second five subtests contributed to the Performance IQ. Full Scale IQ scores ranged from 28 to 195. Subtest raw scores were converted to a mean of 10 and standard deviation of 3, while IQ scores approximated a mean of 100 and standard deviation of 15. Wechsler's subtests dichotomized the composition of his test battery into Verbal and Performance/nonverbal, just as the Army mental tests had distinguished between the Alpha and the Beta/performance tests. This dichotomy remained of value for the same reasons it was helpful with Army mental testing: It permitted valid assessment of individuals whose intelligence was likely to be underestimated by verbal intelligence tests alone (i.e., those who were poorly educated, from non-English-language origins, or otherwise disadvantaged by language-dependent tests). Moreover, Wechsler considered distinctive Verbal and Performance intelligence tasks to sample behaviors in multiple areas of interest, generating important diagnostic information rather than representing different forms of intelligence (Wechsler, 1939). He considered the Verbal and Performance tests to be equally adequate measures of general intelligence, but he emphasized the importance of appraising people "in as many different modalities as possible" (Wechsler et al., 1975, p. 55).

The 1939 test battery (and all subsequent Wechsler intelligence scales) also offered a deviation IQ, the index of intelligence based on statistical distance from the normative mean in standardized units, as Arthur Otis (1917) had proposed. Wechsler deserves credit for popularizing the deviation IQ, although the Otis Self-Administering Tests and the Otis Group Intelligence Scale had already used similar deviation-based composite scores in the 1920s. Inexplicably, Terman and Merrill made the mistake of retaining a ratio IQ (i.e., mental age/chronological age) on the 1937 Stanford–Binet, even though the method had long been recognized as producing distorted IQ estimates for adolescents and adults (e.g., Otis, 1917). Terman and Merrill (1937, pp. 27–28) justified their decision on the dubious ground that it would have been too difficult to reeducate teachers and other test users familiar with the ratio IQ.

Wechsler first introduced the Bellevue Intelligence Scale at a meeting at the New York Academy of Medicine in 1937, and the first edition of *The Measurement of Adult Intelligence*—which would include the manual for the test soon known as the Wechsler–Bellevue Form I—was published in 1939. Early after its publication, Wechsler was approached by George K. Bennett, director of the Tests Division of The Psychological Corporation, who was impressed by the test and asked to produce the test materials (Edwards, 1974). Critics generally praised the "organization of well-known tests into a composite scale" with "considerable diagnostic as well as measurement value" (Lorge, 1943, p. 167), but Wechsler was faulted on technical errors (Anastasi, 1942; Cureton, 1941; McNemar, 1945) and theoretical shortcomings (e.g., Anastasi, 1942; Cronbach, 1949). Figure 1.17 shows Wechsler in the 1940s, after his test had become a success.

Among practicing psychologists and researchers working with adults, the Wechsler–Bellevue was a resounding success. In his review of research on the Wechsler–Bellevue in its first 5 years, Rabin (1945, p. 419) concluded:

> The Wechsler–Bellevue Scales have stimulated considerable psychometric research and have supplanted some time-honored diagnostic tools. The reliability and validity of Wechsler's scales, as a whole and in part, have been proved in several studies. The consensus of opinion is that the test correlates highly with some of the best measures of intellect and that it tends to differentiate better than other measures between the dull and feebleminded. (p. 419)

In an update 6 years later, Rabin and Guertin (1951) noted the "vast popularity and wide usage of the test" (p. 239) and a "veritable flood" of research (p. 211), making the Wechsler–Bellevue "a commonly used measuring rod for comparison and validation, if not actual calibration of newer and more recent techniques" (p. 239).

From 1941 to 1945, Wechsler served as an expert civilian consultant to the Adjutant General's Office, preparing the Wechsler Mental Ability Scale, Form B (Wechsler, 1942, cited by Altus, 1945), also known as the Army Wechsler, and the Wechsler Self-Administering Test. These tests appear to have been of limited use for the military, in large part because they were too difficult for many Army recruits. The Wechsler Mental Ability Scale, Form B is of interest because it consisted of seven Verbal and nine Performance subtests, including Mazes and Series Completion (Altus, 1945), signaling possible additions to the battery. Wechsler also taught in the Army Psychological Training Program (Seidenfeld, 1942).

FIGURE 1.17. David Wechsler was chief psychologist at New York's Bellevue Psychiatric Hospital when he published his Bellevue Intelligence Scale (later known as the Wechsler–Bellevue), which quickly became the intelligence test of choice for adults. Reprinted by courtesy of Arthur Weider.

In the years and decades after the war, Wechsler developed the Wechsler–Bellevue Form II (Wechsler, 1946), the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949), the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955), and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI; Wechsler, 1967). Although David Wechsler died in 1981, most of these tests have gone through multiple editions, with staff test development specialists and external expert advisors substituting for a living author in recent years. In 1975, Wechsler expressed support for measuring intelligence in individuals older than age 65 "without exposing the older person to tests involving speed, perception, and so forth." He proposed to call this test the Wechsler Intelligence Scale for the Elderly, or the WISE (Wechsler et al., 1975; D. O. Herman, personal communication, November 9, 1993). Wechsler never proposed or wrote about achievement tests or nonverbal tests like those that currently carry his name.

In creating his intelligence scales, Wechsler combined popular and clinically useful existing tests into a streamlined, well-organized, and psychometrically innovative battery. Although his tests have become established as industry standards over many decades, Chattin and Bracken (1989) surveyed practicing school psychologists and reported that efficiency and practicality remain the central reasons why the Wechsler intelligence scales remain popular.

LOOSE THREADS: RESOLVED AND UNRESOLVED ISSUES IN INTELLIGENCE TESTING

Students of history are likely to find intelligence and its assessment a fascinating and frustrating subject—full of remarkable characters and events like those I have described—but also with many problems that surface over and over again. Because intelligence testing is a young science, it should be no surprise that so many strands in its story remain loose and unresolved, and there is sufficient diversity in thought among psychologists that even the most scientifically proven ideas will have dissenters. At the same time, it does not seem scientifically unreasonable to expect at some point a consensus-based definition of *intelligence*, agreement on the existence of a general factor of intelligence, and establishment of a uniform framework for understanding the structure of human cognitive abilities (all of which are discussed below). The historical association of intelligence testing with eugenics, however, is an ideological problem that may be harder to resolve; it may forever taint the tests with the appearance of social inequity and racism, in spite of many efforts to enhance the fairness of intelligence tests. In this section, I describe a few of many loose thematic threads that have contributed to breaks in the fabric of applied intelligence testing from its early days.

Before I begin describing long-standing unresolved issues in intelligence, it may be helpful first to note areas that appear to be resolved. In response to the public controversy associated with Herrnstein and Murray's (1994) book *The Bell Curve*, Linda S. Gottfredson of the University of Delaware contacted an editor at the *Wall Street Journal*, who agreed to publish a statement signed by experts about mainstream scientific thinking on intelligence. Gottfredson drafted the statement, had it reviewed by several authorities, and solicited signatures of agreement from experts across psychology and other disciplines. The resulting statement with 25 conclusions, "Mainstream Science on Intelligence," was published in late 1994 with 52 signatories (Gottfredson, 1994); it was later reprinted with supplemental information as an editorial in the journal *Intelligence* (Gottfredson, 1997). In another response to Herrnstein and Murray's book, the APA Board of Scientific Affairs created a task force to issue an authoritative scientific statement about intelligence and its assessment, entitled "Intelligence: Knowns and Unknowns" (Neisser et al., 1996). These two statements represent relatively rare scientific consensus statements about intelligence in the history of psychology. Ironically, there are many areas in which they appear to disagree.

The Definition of *Intelligence*

An initial step in any scholarly endeavor is to define one's terms, but the term *intelligence* still has no consensus-based definition. Efforts to arrive at a consensus date back about a century, as do criticisms that "psychologists have never agreed on a definition" (Lippmann, 1922c, p. 213). In a frequently quoted but much reviled definition, E. G. Boring (1923) wrote:

> Intelligence as a measurable capacity must at the start be defined as the capacity to do well in an intelligence test. Intelligence is what the tests test. This is a narrow definition, but it is the only point of departure for a rigorous discussion of the tests . . . no harm need result if we but remember that measurable intelligence is simply what the tests of intelligence test, until further scientific observation allows us to extend the definition. (p 35)

The failure to arrive at a consensus on defining *intelligence* after a century of research constitutes one of the most surprising loose threads in the history of psychology. Terman (1916) demurred, essentially arguing that we can work with the construct of intelligence without arriving at a definition:

> To demand, as critics of the Binet method have sometimes done, that one who would measure intelligence should first present a complete definition of it, is quite unreasonable. As Stern points out, electrical currents were measured long before their nature was well understood. Similar illustrations could be drawn from the processes involved in chemistry physiology, and other sciences. In the case of intelligence it may be truthfully said that no adequate definition can possibly be framed which is not based primarily on the symptoms empirically brought to light by the test method. (p. 36)

As demonstrated in the statements above, Boring and Terman expected that research would eventually lead to a definition of *intelligence*. How much longer must we wait?

As we have reported, the association of intelligence with evolutionary *adaptation* dates back to Spencer (1855), who described intelligence as "an adjustment of inner to outer relations" (p. 486). This definition may be understood as suggesting that intelligence confers a capacity to adapt to environmental change, but principles of neo-Darwinian evolution hold that natural selection favors adaptations that enhance survival and reproductive fitness. In order to validate a definition of intelligence featuring adaptation, then, the logical and empirical question is whether intelligence confers any advantages in terms of longer lifespans, fecundity, or other aspects of reproductive fitness. Studies relating intelligence to evolutionary fitness (e.g., family size, number of children) date back to the 1930s, and clearly a meta-analysis is needed to make sense of the many contradictory findings. Gottfredson (2007) recently reported evidence that higher intelligence may improve overall survival rate, and that lower intelligence may be associated with a disproportionately elevated risk of accidental death. Together with colleagues, she has also reported findings of a fitness factor that is related to intelligence (Arden, Gottfredson, Miller, & Pierce, 2009).

Several formal meetings or print symposia have sought a definition of *intelligence*, and the clear-cut conclusion from these efforts is that the experts do not agree on a definition. A list of proposed definitions for the term appears in Table 1.2. The earliest symposium I can identify, entitled "Instinct and Intelligence" (e.g., Myers, 1910), was held in London in July 1910, at a joint meeting of the Aristotelian and British Psychological Societies and the Mind Association, with resulting papers appearing in the *British Journal of Psychology*. The best-known print symposium is "Intelligence and Its Measurement: A Symposium," appearing in the *Journal of Educational Psychology* (Peterson, 1921; Pintner, 1921; Thorndike, 1921). The symposium asked 17 leading investigators explicitly what they conceived intelligence to be. Another symposium, "The Nature of General Intelligence and Ability," was conducted at the Seventh International Congress of Psychology, held at Oxford University in 1923 (e.g., Langfeld, 1924). In a follow-up to the 1921 *Journal of Educational Psychology* symposium, Sternberg and Detterman (1986) asked 25 authorities to write essays conveying what they believe intelligence to be. Sternberg and Berg (1986) tabulated facets of the definitions provided: In descending order, the most frequent attributes in definitions of intelligence were higher-level cognitive functions (50%), that which is valued by culture (29%), executive processes (25%), elementary processes (perception, sensation, and/or attention; 21%), knowledge (21%), and overt behavioral manifestations of intelligence (such as effective or successful responses; 21%). By comparison, the most frequent attributes in definitions from the 1921 symposium were higher-level cognitive functions (57%), adaptation (29%), ability to learn (29%), physiological mechanisms (29%), elementary processes (21%), and overt behavioral manifestations of intelligence (21%). Even efforts to seek definitions of intelligence among laypeople have found that definitions vary; moreover, people can be self-serving and seem to offer definitions that also capture some quality readily found in themselves (e.g., Gay, 1948).

Never one to embrace diverse perspectives, Charles E. Spearman (1927) disparaged "repeated recourse to symposia" (p. 8) and surveys of expert opinion in efforts to define intelligence:

> Chaos itself can go no further! The disagreement between different testers—indeed, even the doctrine and the practice of the selfsame tester—has reached its apogee. If they still tolerate each other's proceedings, this is only rendered possible by the ostrich-like policy of not looking facts in the face. In truth, "intelligence" has become a mere vocal sound, a word with so many meanings that it finally has none. (p. 14)

Jensen (1998) echoed Spearman's sentiment, recommending that psychologists "drop the ill-fated word from our scientific vocabulary, or use it only in quotes, to remind ourselves that it is not only scientifically unsatisfactory but wholly unnecessary" (p. 49).

The argument has also been made that a structural/statistical understanding of intelligence may serve as an adequate substitute for a verbal/descriptive definition. Gottfredson and Saklofske (2009) suggest that definitional issues of intelligence are "now moot because the various empirical referents to which the term is commonly applied can be distinguished empirically and related within a common conceptual structure [i.e., the Cattell–Horn–Carroll model of human cognitive abilities]" (p. 188).

To *g* or Not to *g*?

Another long-standing unresolved thread in the history of intelligence testing has to do with the general factor of intelligence, psychometric *g*. General intelligence was affirmed in the 1994 "Mainstream Science on Intelligence" statement (Gottfredson, 1997), but the 1996 "Intelligence: Knowns and Unknowns" statement hedged on *g*, stating that "while the *g*-based factor hierarchy is the most widely accepted current view of the structure of abilities, some theorists regard it as misleading" (Neisser et al., 1996, p. 81). Here I describe some history for *g*.

In 1904, Charles E. Spearman (1863–1945) published a groundbreaking paper reporting the discovery of a factor of "general intelligence," derived from positive intercorrelations between individual scores on tests of sensory discrimination, musical talent, academic performance, and common sense. Although the correlation coefficient statistic was still relatively new, Spearman realized that previous studies (e.g., those by Gilbert and by Wissler) had failed to account for measurement error—that is, reduced score reliability, which invariably reduces the magnitude of correlations. He devised a method to correct the correlation coefficient for attenu-

TABLE 1.2. Selected Definitions of Intelligence (Arranged Chronologically)

Herbert Spencer (1855): "Instinct, Reason, Perception, Conception, Memory, Imagination, Feeling, Will, &c., &c., can be nothing more than either conventional groupings of the correspondences; or subordinate divisions among the various operations which are instrumental in effecting the correspondences. However widely contrasted they may seem, these various forms of intelligence cannot be anything else than either particular modes in which the adjustment of inner to outer relations is achieved; or particular parts of the process of adjustment" (p. 486).

Alexander Bain (1868): "The functions of Intellect, Intelligence, or Thought, are known by such names as Memory, Judgment, Abstraction, Reason, Imagination" (p. 82).

Hermann Ebbinghaus (1908): "Intelligence means organization of ideas, manifold interconnection of all those ideas which ought to enter into a unitary group because of the natural relations of the objective facts represented by them. The discovery of a physical law in a multitude of phenomena apparently unrelated, the interpretation of an historical event of which only a few details are directly known, are examples of intelligence thought which takes into consideration innumerable experiences neglected by the less intelligent mind. Neither memory alone nor attention alone is the foundation of intelligence, but a union of memory and attention" (pp. 150–151).

Charles S. Myers (1910): "As the organism becomes endowed with an increasingly larger number of mutually incompatible modes of reaction, the intelligent aspect apparently comes more and more to the fore while the instinctive aspect apparently recedes *pari passu* into the background" (p. 214).

C. Lloyd Morgan (1910): "I regard the presence of implicit expectation (in the lower forms) or explicit anticipation (in the higher forms) as distinguishing marks or criteria of intelligence. In other words for the intelligent organism the present experience at any given moment comprises more or less 'meaning' in terms of previously-gotten experience" (p. 220).

H. Wildon Carr (1910): "Intelligence is the power of using categories, it is knowledge of the relations of things. It is a knowledge that gives us the representation of a world of objects externally related to one another, a world of objects in space, or measurable actions and reactions. . . . Intelligence is an outward view of things, never reaching the actual reality it seeks to know" (pp. 232–233).

Alfred Binet and Théodore Simon (Binet, 1911/1916): "Intelligence serves in the discovery of truth. But the conception is still too narrow; and we return to our favorite theory; the intelligence marks itself by the best possible adaptation of the individual to his environment" (pp. 300–301).

William Stern (1914): "Intelligence is a general capacity of an individual consciously to adjust his thinking to new requirements: it is general mental adaptability to new problems and conditions of life" (p. 3).

M. E. Haggerty (1921). "In my thinking the word intelligence does not denote a single mental process capable of exact analytic definition. It is a practical concept of connoting a group of complex mental processes traditionally defined in systematic psychologies as sensation, perception, association, memory, imagination, discrimination, judgment and reasoning "(p. 212).

V. A. C. Henmon (1921): "Intelligence . . . involves two factors—the capacity for knowledge and knowledge possessed" (p. 195).

Joseph Peterson (1921): "Intelligence seems to be a biological mechanism by which the effects of a complexity of stimuli are brought together and given a somewhat unified effect in behavior. It is a mechanism for adjustment and control, and is operated by internal as well as by external stimuli. The degree of a person's intelligence increases with his range of receptivity to stimuli and the consistency of his organization of responses to them" (p. 198).

Rudolf Pintner (1921): "I have always thought of intelligence as the ability of the individual to adapt himself adequately to relatively new situations in life. It seems to include the capacity for getting along well in all sorts of situations. This implies ease and rapidity in making adjustments and, hence, ease in breaking old habits and in forming new ones" (p. 139).

Lewis M. Terman (1921): "The essential difference, therefore, is in the capacity to form concepts to relate in diverse ways, and to grasp their significance: *An individual is intelligent in proportion as he is able to carry on abstract thinking*" (p. 128; emphasis in original).

Edward L. Thorndike (1921): "Realizing that definitions and distinctions are pragmatic, we may then define intellect in general as *the power of good responses from the point of view of truth or fact*, and may separate it according as the situation is taken in gross or abstractly and also according as it is experienced directly or thought of" (p. 124; emphasis in original).

L. L. Thurstone (1921): "Intelligence as judged in everyday life contains at least three psychologically differentiable components: a) the capacity to inhibit an instinctive adjustment, b) the capacity to redefine the inhibited instinctive adjustment in the light of imaginally experienced trial and error, c) the volitional capacity to realize the modified instinctive adjustment into overt behavior to the advantage of the individual as a social animal" (pp. 201–202).

Herbert Woodrow (1921): "Intelligence . . . is the capacity to acquire capacity" (p. 208).

(cont.)

TABLE 1.2. *(cont.)*

E. G. Boring (1923): "Intelligence as a measurable capacity must at the start be defined as the capacity to do well in an intelligence test. Intelligence is what the tests test" (p. 35).

Édouard Claparède (1924): "[Intelligence is] the ability to solve new problems" (quoted by Langfeld, 1924, p. 149).

Godfrey H. Thomson (1924): "[Intelligence is] the ability to meet new situations with old responses and to discard those responses which prove unsuccessful" (quoted by Langfeld, 1924, p. 149).

David Wechsler (1939): "Intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment" (p. 3).

Anne Anastasi (1986): "Intelligence is not an entity within the organism but a quality of behavior. Intelligent behavior is essentially adaptive, insofar as it represents effective ways of meeting the demands of a changing environment" (pp. 19–20).

Jonathan Baron (1986): "I define intelligence as the set of whatever abilities make people successful at achieving their rationally chosen goals, whatever those goals might be, and whatever environment they are in. . . . To say that a person has a certain level of ability is to say that he or she can meet a certain standard of speed, accuracy, or appropriateness in a component process defined by the theory in question" (p. 29).

J. W. Berry (1986): "At the present time intelligence is a construct which refers to the end product of individual development in the cognitive-psychological domain (as distinct from the affective and conative domains); this includes sensory and perceptual functioning but excludes motor, motivational, emotional, and social functioning . . . it is also adaptive for the individual, permitting people to operate in their particular cultural and ecological contexts" (p. 35).

J. P. Das (1986): "Intelligence, as the sum total of all cognitive processes, entails planning, coding of information and attention arousal. Of these, the cognitive processes required for planning have a relatively higher status in intelligence. Planning is a broad term which includes among other things, the generation of plans and strategies, selection from among available plans, and the execution of those plans. . . . Coding refers to two modes of processing information, simultaneous and successive. . . . The remaining process (attention arousal) is a function basic to all other higher cognitive activities" (pp. 55–56).

Douglas K. Detterman (1986): "In my opinion, intelligence can best be defined as a finite set of independent abilities operating as a complex system" (p 57).

John Horn (1986): "'What do I conceive intelligence to be?' This is rather like asking me: 'What do I conceive invisible green spiders to be?' For current knowledge suggests to me that intelligence is not a unitary entity of any kind. Attempts to describe it are bound to be futile" (p. 91).

Earl Hunt (1986): "'Intelligence' is solely a shorthand term for the variation in competence on cognitive tasks that is statistically associated with personal variables. . . . Intelligence is used as a collective term for 'demonstrated individual differences in mental competence'" (p. 102).

James W. Pellegrino (1986): "The term intelligence denotes the general concept that individuals' responses to situations vary in quality and value as judged by their culture" (p. 113).

Sandra Scarr (1986): "To be an effective, intelligent human being requires a broader form of personal adaptation and life strategy, one that has been described in 'invulnerable' children and adults: They are copers, movers, and shapers of their own environments" (p. 120).

Richard E. Snow (1986): "[Intelligence can be defined in several ways:] . . . [1] the incorporation of concisely organized prior knowledge into purposive thinking—for short, call it *knowledge-based thinking*. . . . [2] *apprehension* captures the second aspect of my definition—it refers to Spearman's (1923, 1927) principle that persons (including psychologists) not only feel, strive, and know, but also *know* that they feel, strive, and know, and can anticipate further feeling, striving, and knowing; they monitor and reflect upon their own experience, knowledge, and mental functioning in the past, present, and future tenses. . . . [3] *adaptive purposeful striving*. It includes the notion that one can adopt or shift strategies in performance to use what strengths one has in order to compensate for one's weaknesses. . . . [4] agile, analytic reasoning of the sort that enables significant features and dimensions of problems, circumstances, and goals to be decontextualized, abstracted, and interrelated rationally . . . *fluid-analytic reasoning*. . . . [5] *mental playfulness* . . . able to find or create interesting problems to solve and interesting goals toward which to strive. This involves both tolerance of ambiguity and pursuit of novelty. . . . [6] *idiosyncratic learning* . . . Persons differ from one another in the way they assemble their learning and problem-solving performance, though they may achieve the same score. Persons differ *within* themselves in how they solve parts of a problem, or different problems in a series" (pp. 133–134; emphasis in original).

Robert J. Sternberg (1986): "Intelligence is mental self-government. . . . The essence of intelligence is that it provides a means to govern ourselves so that our thoughts and actions are organized, coherent, and responsive to both our internally driven needs and to the needs of the environment" (p. 141).

ation, reporting subsequently that his correlational analyses showed "all branches of intellectual activity have in common one fundamental function (or group of functions)" (p. 284), which he later described using concepts from physics such as "the amount of a general mental energy" (Spearman, 1927, p. 137). The *g* factor, or psychometric *g*, was a mathematically derived general factor, stemming from the shared variance that saturates batteries of cognitive/intelligence tests. Jensen (1998) has summarized the literature showing that correlates of *g* include scholastic performance, reaction time, success in training programs, job performance in a wide range of occupations, occupational status, earned income, and creativity, among others.

Critics of general intelligence appeared quickly. Edward L. Thorndike, who challenged Spearman's work for decades, reported no support for *g* on a set of measures similar to those originally used by Spearman, finding a weak correlation between sensory discrimination and general intelligence, and stating that "one is almost tempted to replace Spearman's statement by the equally extravagant one that there is *nothing whatever* common to all mental functions, or to any half of them" (Thorndike, Lay, & Dean, 1909, p. 368; emphasis in original).

Until Spearman's death, Thorndike; a Scotsman, Godfrey Thomson; and two Americans, Truman L. Kelley and Louis L. Thurstone, participated in an ongoing scholarly debate with him on the existence and nature of *g*, as well as other aspects of the structure of intelligence. Spearman devoted the rest of his career to elaboration and defense of his theory, authoring *The Nature of "Intelligence" and the Principles of Cognition* (Spearman, 1923), *The Abilities of Man: Their Nature and Measurement* (Spearman, 1927), and *Human Ability: A Continuation of "The Abilities of Man"* (Spearman & Wynn Jones, 1950). A good account of this debate may be found in R. M. Thorndike and Lohman (1990). Newly discovered exchanges among Thorndike, Thomson, and Spearman in the 1930s serve to highlight Spearman's dogmatism (Deary, Lawn, & Bartholomew, 2008).

The leading intelligence test developers generally accepted the existence of a psychometric *g* factor. After initial reticence, Alfred Binet eventually embraced a general factor; in *Les Idées Modernes sur les Enfants*, Binet (1909/1975) wrote that "the mind is unitary, despite the multiplicity of its faculties . . . it possesses one essential function to which all the others are subordinated" (p. 117). In the 1916 Stanford–Binet, Lewis M. Terman accepted the concept of general intelligence and conceded that the IQ score provided a good estimate of *g*:

> It is true that more than one mental function is brought into play by the test. The same may be said of every other test in the Binet scale and for that matter of any test that could be devised. It is impossible to isolate any function for separate testing. In fact, the functions called memory, attention, perception, judgment, etc., never operate in isolation. There are no separate and special "faculties" corresponding to such terms, which are merely convenient names for characterizing mental processes of various types. In any test it is "general ability" which is operative, perhaps now *chiefly* in remembering, at another time *chiefly* in sensory discrimination, again in reasoning, etc. (p. 194; emphasis in original)

David Wechsler, who had been deeply impressed with Spearman during his few months at University College London in 1919, wrote that Spearman's theory and its proofs constitute "one of the great discoveries of psychology" (Wechsler, 1939, p. 6). He further noted that "the only thing we can ask of an intelligence scale is that it measures sufficient portions of intelligence to enable us to use it as a fairly reliable index of the individual's global capacity" (p. 11).

What is the current status of *g*? When Reeve and Charles (2008) surveyed 36 experts in intelligence, they found a consensus that *g* is an important, nontrivial determinant (or at least predictor) of important real-world outcomes, and that there is no substitute for *g* even if performance is determined by more than *g* alone. With the leading authors of intelligence tests accepting psychometric *g*, and with authorities in intelligence research consensually accepting its importance, the thread of general intelligence would appear to be well secured in our metaphorical tapestry of the history intelligence.

Yet the concept of general intelligence continues to be challenged, most often on theoretical grounds but also on statistical grounds. Stephen J. Gould (1996) forcefully challenged *g*, associating it with many of the historically negative (and shameful) applications of intelligence testing. Several intelligence theorists, including Raymond B. Cattell, J. P. Das, Howard Gardner, and Robert J. Sternberg, have also rejected the concept of general intelligence. The most cogent challenges to *g* have come from John L. Horn (Horn & Noll, 1994, 1997), who pointed out fallacies of extracting *g* from the *positive manifold* (i.e., the finding that almost all tests that reliably measure a cognitive ability correlate positively with all other such tests).

The Structure of Intelligence

The struggle to construct a complex model of intelligence probably began with the phrenologists, who specified individual faculties (each corresponding to an "organ" of the brain) that together constituted intelligence. For example, Combe (1830) described faculties of perception (e.g., form, size, weight, eventuality, language) and faculties of reflection (e.g., comparison, causality) that altogether constituted intellectual faculties; he also described a separate set of affective faculties. With the discovery of *g* by Spearman (1904), the notion of a unitary intelligence gained traction, but by the end of the 1930s, psychologists and educators were again embracing the complexity of the mind (e.g., Ackerman, 1995). Current hierarchical models of intelligence feature broad ability factors, which have grown steadily in number: from the two factors enumerated by Cattell (1941) and Vernon (1950) to the eight specified by Carroll (1993) to about 10 factors specified by Carroll (2003) to about 15 or 16 broad factors in 2010 (e.g., McGrew, 2009; Newton & McGrew, 2010). The question that appears to be unresolved in this thread is this: Just how many group factors constitute the structure of intelligence?

For much of the 20th century and into the 21st, the complex structure of intelligence has been revealed through statistical methodologies that discover and define sources of test performance variance, usually through factor analyses. Factor analysis is a statistical technique capable of reducing many variables into a few underlying dimensions. The foundation for use of factor analysis in understanding the structure of cognition was laid with Spearman (1904). Spearman's theory encompassing general intelligence was originally called *two-factor theory* because it partitioned performance variance into a *general factor* shared across tasks, and *specific factors* that were unique to individual tasks. Following the contributions of Kelley, Thorndike, and Thurstone (among others), Spearman (1927) reluctantly came to acknowledge the existence of *group factors* formed by clusters of tests that yielded higher-than-expected intercorrelations by virtue of similarities in their content, format, or response requirements: "Any element whatever in the specific factor of an ability will be turned into a group factor, if this ability is included in the same set with some other ability which also contains this element" (p. 82). The extraction of a general factor and group factors (now called *broad ability factors*) contributed to the development of *hierarchical* structural analyses of intelligence. In hierarchical factor analyses, a general factor is first extracted; the residual variance is factored to extract any group factors; and the remaining variance is often said to be specific.

Although there have been well over 1,000 factor-analytic investigations in the literature of intelligence and cognitive abilities (see Carroll, 1993), many of which remain important in understanding the structure of cognitive abilities, space only permits coverage of a few prototypal models with distinctive characteristics.

Thurstone's Primary Mental Abilities

Louis L. Thurstone (1887–1955) developed the statistical technique of multiple factor analysis and is best remembered for his theory of primary mental abilities, a factor-analysis-derived model of multiple cognitive abilities that effectively challenged Spearman's single general factor of intelligence. Thurstone developed factor analysis techniques permitting the extraction of factors that are orthogonal to each other (i.e., separate, independent, and unrelated). From a battery of 56 paper-and-pencil tests administered in about 15 hours to each of 240 superior, college-level students, Thurstone (1938) extracted seven primary factors: spatial/visual, perception of visual detail, numerical, two verbal factors (logic and words), memory, and induction. From a study of over 700 students age 14, who were given 60 tests in 11 sessions lasting 1 hour each, Thurstone and Thurstone (1941) extracted six factors: verbal comprehension, word fluency, space, number, memorizing, and reasoning/induction. By 1945, Thurstone had settled on eight primary mental abilities, each denoted by a letter: Verbal Comprehension (V), Word Fluency (W), Number Facility (N), Memory (M), Visualizing or Space Thinking (S), Perceptual Speed (P), Induction (I), and Speed of Judgment (J). Although Thurstone (1947) eventually accepted the existence of a general factor, he considered the use of a single score such as the IQ to be inadequate, and urged the use of cognitive profiles describing strengths and weaknesses among the fundamental abilities (Thurstone, 1945).

Vernon's Hierarchical Model

In what has been called the first truly hierarchical model of intelligence, Philip E. Vernon (1905–1987) proposed that a higher-order *g* factor dominates two lower-order factors, v:ed (verbal:educational) and k:m (spatial:mechanical); in turn, v:ed and k:m subsume various minor group factors, which in turn dominate very narrow and specific factors. Based on his review of factor-analytic investigations through 1950, Vernon (1950, 1961) considered v:ed to dominate verbal, number, reasoning, attention, and fluency factors, while k:m dominates spatial ability, mechanical ability, psychomotor coordination, reaction time, drawing, handwork, and various technical abilities. He considered it a likely oversimplification to assume that there are just two factors at the level below *g*, although his simple dichotomy may be seen as having supported the verbal–performance dichotomy traditionally associated with the Wechsler intelligence scales.

Cattell, Horn, and Carroll's Model of Fluid and Crystallized Intelligence

Arguably the most important contemporary structural and hierarchical model of intelligence is based upon extensions of the theory of fluid (Gf) and crystallized (Gc) intelligence first proposed by Raymond B. Cattell (1905–1998) in a 1941 APA convention presentation. Cattell, who completed his doctorate in 1929 at University College London with Spearman, joined E. L. Thorndike's research staff at Columbia University in 1937, where he worked closely with proponents of multifactor models of intelligence. He authored over 500 articles and 43 books during his career. In his 1941 APA presentation, Cattell asserted the existence of two separate general factors: g_f (fluid ability or fluid intelligence) and g_c (crystallized ability or crystallized intelligence). The convention was later adopted that these factors would be represented by uppercase G, whereas a single general factor would be represented by lowercase *g*.

Fluid ability was described by Cattell (1963, 1971) and Horn (1976) as a facility in reasoning, particularly where adaptation to new situations is required and crystallized learning assemblies are of little use. Ability is considered to be fluid when it takes different forms or utilizes different cognitive skill sets according to the demands of the problem requiring solution. For Cattell, fluid ability is the most essential general-capacity factor, setting an upper limit on the possible acquisition of knowledge and crystallized skills. In contrast, *crystallized* intelligence refers to accessible stores of knowledge and the ability to acquire further knowledge via familiar learning strategies. It is typically measured by recitation of factual information, word knowledge, quantitative skills, and language comprehension tasks because these include the domains of knowledge that are culturally valued and educationally relevant in the Western world (Cattell, 1941, 1963, 1971, 1987; Horn & Cattell, 1966).

Cattell's model of fluid and crystallized intelligence was energized by the contribution of John L. Horn (1928–2006). Not only was Horn's (1965) dissertation the first empirical study of the theory since 1941; it also showed that fluid and crystallized abilities have different developmental trajectories over the lifespan (McArdle, 2007). Cattell and Horn expanded the number of ability factors from two to five (adding visualization, retrieval capacity, and cognitive speed; Horn & Cattell, 1966). In the next 25 years or so, Horn had arrived at nine ability factors (Horn & Noll, 1994, 1997), while Cattell's list had grown to six ability factors (adding distant memory and retrieval) plus three smaller provincial factors (visual, auditory, and kinesthetic; Cattell, 1998). The growth of the number of factors in this model continues, and a 2001 symposium at the University of Sydney enumerated even more potential ability factors (Kyllonen, Roberts, & Stankov, 2008). As noted earlier, McGrew (2009; see also Newton & McGrew, 2010) now lists 15 or 16 broad ability factors.

In 1993, John B. Carroll (1916–2003) built upon the work of Cattell and Horn by proposing a hierarchical, multiple-stratum model of human cognitive abilities with the general intelligence factor, *g*, at the apex (or highest stratum); eight broad factors of intelligence at the second stratum; and at least 69 narrow factors at the first (or lowest) stratum. Carroll was the author of nearly 500 books and journal articles over the span of 60 years; he had been mentored early in his career by L. L. Thurstone, and some years later after Thurstone's death he became director of the Thurstone Psychometric Laboratory at the University of North Carolina, Chapel Hill (Jensen, 2004). For a dozen years after his retirement, Carroll (1983, 1993, 1994) accumulated over a thousand archival datasets related to human cognitive test performance; 461 of the datasets were ultimately judged adequate for his analyses. He then conducted iterative principal-factor analyses requiring convergence to a strict criterion, followed by varimax rotation of the principal-factor matrix, with the requirement that each extracted factor contain salient loadings on at least two variables. If necessary, promax or other rotational procedures were used. Factorization was then carried up to the highest viable order. The data were subjected to the Schmid–Leiman orthogonalized hierarchical-factor procedure, and factor interpretations were based on the resulting hierarchical-factor matrix. Carroll's results showed general intelligence (*g*) as appearing in the highest stratum; the second stratum, listed in descending strength of association with *g*, consisted of fluid intelligence (Gf), crystallized intelligence (Gc), general memory and learning (Gsm), broad visual perception (Gv), broad auditory perception (Ga), broad retrieval ability (Gr), broad cognitive speediness (Gs), and processing speed (reaction time decision speed); finally, very narrow and specific factors were placed in the lowest stratum. Although Carroll's three-stratum model is historically young, its early reception suggests that it has quickly become a landmark study. The following samples from reviews are fairly representative:

- "Further research may alter details of the map, although it is unlikely that any research for some years to come will lead to a dramatic alteration in Carroll's taxonomy." (Brody, 1994, p. 65)
- "It is simply the finest work of research and scholarship I have read and is destined to be the classic study and reference work of human abilities for decades to come." (Burns, 1994, p. 35)
- "[It is] a truly monumental work." (Jensen, 2004, p. 3)
- "Carroll's work represents what may well be the most extensive, indeed, exhaustive analysis of a data case that has ever been attempted in the field of intelligence. The theory deserves to be taken seriously." (Sternberg, 1994, p. 65)

A note of caution for applied practitioners, however, comes from Carroll himself (1993): He indicated that his survey of cognitive abilities "paid very little attention to the importance, validity, or ultimate usefulness of the ability factors that have been identified" (p. 693). Carroll's three stratum theory has been integrated with extended Gf-Gc theory to form the Cattell–Horn–Carroll (CHC) framework, a name to which Horn and Carroll both agreed a few years after Cattell's death (Newton & McGrew, 2010). The CHC framework already appears to have exerted a strong influence upon the development of contemporary intelligence tests (e.g., Keith & Reynolds, 2010). Shortly before his death, Carroll (2003) expanded his model to include 10 second-stratum factors, indicating that even this definitive model may be expanded.

Intelligence and Eugenics

> We have seen more than once that the public welfare may call upon the best citizens for their lives. It would be strange if it could not call upon those who already sap the strength of the State for these lesser sacrifices, often not felt to be such by those concerned, in order to prevent our being swamped with incompetence. It is better for all the world, if instead of waiting to execute degenerate offspring for crime, or to let them starve for their imbecility, society can prevent those who are manifestly unfit from continuing their kind. The principle that sustains compulsory vaccination is broad enough to cover cutting the Fallopian tubes. Three generations of imbeciles are enough.
>
> —OLIVER WENDELL HOLMES (*Buck v. Bell*, 1927)

So wrote Justice Oliver Wendell Holmes, Jr., for the majority opinion of the U.S. Supreme Court in 1927, in the case of Carrie Buck versus James Hendren Bell, Superintendent of the Virginia State Colony for Epileptics and Feeble Minded. Carrie Buck was an 18-year-old woman with the mental age equivalent of 9 when the superintendent of the Virginia State Colony petitioned to have her sterilized. She was reported to be the daughter of a feeble-minded mother in the same institution and the mother of a feeble-minded child—hence Holmes's statement that "Three generations of imbeciles are enough." By an 8-to-1 margin, the court upheld the 1924 Virginia statute, the Eugenical Sterilization Act of 1924, authorizing the compulsory sterilization of "mental defectives," including individuals who were "feeble-minded" (i.e., intellectually disabled). On October 19, 1927, Carrie Buck was sterilized. Although the Supreme Court ruling has never been challenged or reversed, the Commonwealth of Virginia repealed the 1924 sterilization law in 1974. Historian Paul A. Lombardo (2008) recently reexamined this case, finding that there was insufficient evidence ever to assert cognitive impairment in Buck or her daughter, based on their school records.

For our purposes, it may be enough to cite compulsory sterilization laws for those with intellectual disabilities as a historical illustration of how intelligence test results may be (mis)used. In the broader context are questions about the wisdom of making legal, political, and public policy decisions on the basis of intelligence test research. Scholars in intelligence are at risk when they stray too far from psychological science into the realm of social engineering.

Francis Galton coined the term *eugenics* in 1883, describing it as "the science of improving stock" and defining it as "all influences that tend in however remote a degree to give to the more suitable races or strains of blood a better chance of prevailing speedily over the less suitable than they otherwise would have had" (p. 25). In *Hereditary Genius* (1869), he had already presented evidence that superior abilities are found more often among eminent families (i.e., those of judges, statesmen, premiers, commanders, scientists, scholars, etc.), and he proposed to increase the proportion of individuals with superior genetic endowments and thereby benefit the national intelligence through selective early marriages. "A man's natural abilities are derived by inheritance," Galton (1869) wrote, "under exactly the same limitations as are the form and physical features of the whole organic world" (p. 1). He related his vision of a eugenics-practicing society in an unpublished fictional tale entitled "Kantsaywhere" (Galton, 1930). In this utopia, an individual's hereditary worth was measured by anthropometric tests, genetic failures were placed in labor colonies, enforced celibacy was the rule, and childbirth for the "unfit" was a crime. Karl Pearson noted in a footnote to this tale (p. 416 in Galton, 1930) that Galton's fictional laboratory in Kantsaywhere bears an uncanny resemblance to his anthropometric laboratory at South Kensington, one of the places where intelligence testing began. A photograph of Galton toward the end of his life, with his friend, colleague, and biographer, Karl Pearson, appears in Figure 1.18. Pearson, the renowned statistician, was also a dedicated eugenicist.

FIGURE 1.18. Francis Galton at age 87 with his biographer, Karl Pearson. Both were dedicated eugenicists. Photo from Pearson (1930, Plate 36). Reprinted by permission of The Pearson Papers, UCL Library Services, Special Collections.

Almost all of the early authorities in the field of intelligence either wrote favorably about eugenics or belonged to organizations advocating eugenics. Some of the authorities on record as favoring eugenics in one form or another include J. McKeen Cattell, Raymond B. Cattell, Henry H. Goddard, Lewis M. Terman, Edward L. Thorndike, and Robert M. Yerkes. Until the horrors of Nazi genocide were exposed, including euthanasia of individuals with intellectual disabilities or mental disorders, eugenics was commonly seen as a contribution of science to human (and national) improvement. Lewis M. Terman, author of the Stanford–Binet, took a particularly active role in advocating for eugenics. For example, in a report to the California state legislature, Terman (1917) saw those with intellectual disabilities as having only negative impacts on society:

> Feeble-mindedness has always existed; but only recently have we begun to recognize how serious a menace it is to the social, economic, and moral welfare of the state. Extensive and careful investigations, in large numbers and in diverse parts of the United States, have furnished indisputable evidence that it is responsible for at least one-fourth of the commitments to state penitentiaries and reform schools, for the majority of cases of chronic and semi-chronic pauperism, and for much of our alcoholism, prostitution, and venereal diseases. (p. 45)

Terman's solutions were to segregate "feeble-minded" students in special classes so as not to "interfere with instruction" or "be a source of moral contagion" for other students (p. 51). He did not overtly recommend sterilization, but he implied that some action was necessary to prevent reproduction: "Three-fourths of the cases of feeble-mindedness are due to a single cause, heredity; and the one hopeful method of curtailing the increasing spawn of degeneracy is to provide additional care for our higher-grade defectives during the reproductive period" (p. 52).

Two of the most important 20th-century figures in applied intelligence testing, however, Alfred Binet and David Wechsler, are on record as having rejected perspectives associated with eugenics. Binet argued that intelligence can be changed, and he even developed a program of "mental orthopedics" to make educational interventions:

> I have often observed, to my regret, that a widespread prejudice exists with regard to the educability of intelligence. . . . A few modern philosophers seem to lend their moral support to these deplorable verdicts when they assert that an individual's intelligence is a fixed quantity, a quantity which cannot be increased. We must protest and react against this brutal pessimism. We shall attempt to prove that it is without foundation. (Binet, 1909/1975, pp. 105–106)

David Wechsler found a more oblique way to criticize the eugenicists—by associating them with totalitarianism. In a 1961 paper, shortly after defining the terms *eugenes* and *apartheid*, he wrote, "The belief in class distinctions, whether considered innate or acquired, is . . . an essential tenet of all groups who are afraid of being ousted or displaced, and in particular of totalitarian governments" (Wechsler, 1961, p. 421). As Wechsler was a member of an oppressed immigrant group (Eastern European Jews) threatened with genocide in Romania and Germany, his condemnation of eugenics should not be surprising.

How does eugenics constitute a loose thread in the history of intelligence? Although no mainstream authorities today advocate for eugenics of the type that led to tragedies in the past, scholarship in the biology and heredity of intelligence remains extremely controversial, with recent accounts including threats to the academic freedom and even loss of lifetime recognition of achievements in psychology for those who conduct research in associated areas, including heritability (e.g., APA, 1997; Gottfredson, 2010; Horn, 2001; Tucker, 2009). Moreover, it might be argued that the history of intelligence and eugenics has contributed to a public perception that intelligence is about elitism, racism, and exclusion. In spite of over a century of research, the study of intelligence remains controversial for its social applications and implications.

REFERENCES

Ackerman, M. (1995). Mental testing and the expansion of educational opportunity. *History of Education Quarterly*, *35*, 279–300.

Alien slackers may not escape service. (1917, April 22). *New York Times*, p. E3.

Altus, W. D. (1945). The differential validity and difficulty of subtests of the Wechsler Mental Ability Scale. *Psychological Bulletin*, *42*, 238–249.

American Psychological Association (APA). (1997). Gold Medal Award for Life Achievement in Psychological Science: Raymond B. Cattell. *American Psychologist*, *52*, 797–799.

Anastasi, A. (1942). Review: *The Measurement of Adult Intelligence* by David Wechsler. *American Journal of Psychology*, *55*, 608–609.

Anastasi, A. (1986). Intelligence as a quality of behavior. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 19–21). Norwood, NJ: Ablex.

Arden, R., Gottfredson, L. S., Miller, G., & Pierce, A. (2009). Intelligence and semen quality are positively correlated. *Intelligence*, *37*, 277–282.

Bain, A. (1868). *Mental science: A compendium of psychology, and the history of philosophy*. New York: Appleton.

Baker, N. D. (1918). *Annual report of the Secretary of War, 1918* (Vol. 1, pp. 5–65). Washington, DC: U.S. Government Printing Office.

Baron, J. (1986). Capacities, dispositions, and rational thinking. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 29–33). Norwood, NJ: Ablex.

A Bellevue Unit formally opened. (1933, November 3). *New York Times*, p. 20.

Berry, J. W. (1986). A cross-cultural view of intelligence. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 35–38). Norwood, NJ: Ablex.

Binet, A. (1890). The perception of lengths and numbers in some small children. *Revue Philosophique*, *30*, 68–81.

Binet, A. (1894). *Introduction à la psychologie expérimentale*. Paris: Baillière et Cie.

Binet, A. (1903). *L'étude expérimentale de l'intelligence*. Paris: Schleicher Frères.

Binet, A. (1910). Avant-propos: Le bilan de la psychologie en 1910. *L'Année Psychologique*, *17*, v–xi.

Binet, A. (1916). New investigations upon the measure of the intellectual level among school children. In H. H. Goddard (Ed.), *The development of intelligence in children (the Binet–Simon Scale)* (E. S. Kite, Trans.;

pp. 274–329). Baltimore: Williams & Wilkins. (Original work published 1911)

Binet, A. (1975). *Modern ideas about children* (S. Heisler, Trans.). Menlo Park, CA: Suzanne Heisler. (Original work published 1909)

Binet, A., & Henri, V. (1894). Le développement de la mémoire visuelle chez les enfants. *Revue Générale des Sciences Pures et Appliquées*, *5*, 162–169.

Binet, A., & Henri, V. (1895). La psychologie individuelle. *L'Année Psychologique*, *2*, 411–465.

Binet, A., & Simon, T. (1914). *Mentally defective children* (W. B. Drummond, Trans). New York: Longmans, Green. (Original work published 1907)

Binet, A., & Simon, T. (1916a). Application of the new methods to the diagnosis of the intellectual level among normal and subnormal children in institutions and in the primary schools. In H. H. Goddard (Ed.), *The development of intelligence in children (the Binet–Simon Scale)* (E. S. Kite, Trans; pp. 91–181). Baltimore: Williams & Wilkins. (Original work published 1905)

Binet, A., & Simon, T. (1916b). The development of intelligence in the child. In H. H. Goddard (Ed.), *The development of intelligence in children (the Binet–Simon Scale)* (E. S. Kite, Trans.; pp. 182–273). Baltimore: Williams & Wilkins. (Original work published 1908)

Binet, A., & Simon, T. (1916c). New methods for the diagnosis of the intellectual level of subnormals. In H. H. Goddard (Ed.), *The development of intelligence in children (the Binet–Simon Scale)* (E. S. Kite, Trans.; pp. 37–90). Baltimore: Williams & Wilkins. (Original work published 1905)

Bingham, W. V. (1942). The Army Personnel Classification System. *Annals of the American Academy of Political and Social Science*, *220*, 18–28.

Bingham, W. V. (1944). Personnel classification testing in the Army. *Science (New Series)*, *100*, 275–280.

Bittner, R. H. (1947). The Army General Classification Test. In G. A. Kelly (Ed.), *New methods in applied psychology* (pp. 45–55). College Park: University of Maryland.

Boake, C. (2002). From the Binet–Simon to the Wechsler–Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology*, *24*, 383–405.

Bolton, T. L. (1892). The growth of memory in school children. *American Journal of Psychology*, *4*, 362–380.

Boring, E. G. (1923, June 6). Intelligence as the tests test it. *The New Republic*, *35*, 35–37.

Boring, E. G. (1961). *Psychologist at large: An autobiography and selected essays*. New York: Basic Books.

Brody, N. (1994). Cognitive abilities. Reviewed work(s): *Human Cognitive Abilities: A Survey of Factor-Analytic Studies* by J. B. Carroll. *Psychological Science*, *5*, 63, 65–68.

Buck v. Bell, 274 U.S. 200 (1927).

Buckhalt, J. A. (1991). Reaction time and intelligence: Contemporary research supports Cattell's abandoned hypothesis. *School Psychology International*, *12*, 355–360.

Burnham, J. C. (2003). Interviewing as a tool of the trade: A not-very-satisfactory bottom line. In D. B. Baker (Ed.), *Thick description and fine texture: Studies in the history of psychology* (pp. 19–37). Akron, OH: University of Akron Press.

Burns, R. B. (1994). Surveying the cognitive terrain. Reviewed work(s): *Human Cognitive Abilities: A Survey of Factor-Analytic Studies* by John B. Carroll. *Educational Researcher*, *23*, 35–37.

Buros, O. K. (1977). Fifty years in testing: Some reminiscences, criticisms, and suggestions. *Educational Research*, 6(7), 9–15.

Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice*, *31*, 141–154.

Camfield, T. M. (1969). *Psychologists at war: The history of American psychology and the First World War*. Unpublished doctoral dissertation, University of Texas at Austin. (AAT No. 7010766)

Camfield, T. M. (1992). The American Psychological Association and World War I: 1914 to 1919. In R. B. Evans, V. S. Sexton, & T. C. Cadwallader (Eds.), *The American Psychological Association: A historical perspective* (pp. 91–118). Washington, DC: American Psychological Association.

Carr, H. W. (1910). Instinct and intelligence. *British Journal of Psychology*, *3*, 230–236.

Carroll, J. B. (1978). On the theory-practice interface in the measurement of intellectual abilities. In P. Suppes (Ed.), *Impact of research on education: Some case studies* (pp. 1–105). Washington, DC: National Academy of Education.

Carroll, J. B. (1983). Studying individual differences in cognitive abilities: Through and beyond factor analysis. In R. F. Dillon & R. R. Schmeck (Eds.), *Individual differences in cognition* (pp. 1–33). New York: Academic Press.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.

Carroll, J. B. (1994). Cognitive abilities: Constructing a theory from data. In D. K. Detterman (Ed.), *Current topics in human intelligence: Vol. 4. Theories of intelligence* (pp. 43–63). Norwood, NJ: Ablex.

Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports *g* and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–21). Boston: Pergamon Press.

Carson, J. (1993). Army Alpha, Army brass, and the search for Army intelligence. *Isis*, *84*, 278–309.

Carson, J. (2007). *The measure of merit: Talents, intelligence, and inequality in the French and American republics, 1750–1940.* Princeton, NJ: Princeton University Press.

Carter, M. H. (1909). The conservation of the defective child. *McClure's Magazine*, *33*(2), 160–171.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, *11*, 63–65.

Cattell, J. M. (1890). Mental tests and measurements. *Mind*, *15*, 373–381.

Cattell, J. M. (1924). The interpretation of intelligence tests. *Scientific Monthly*, *18*, 508–516.

Cattell, J. M. (1930). Psychology in America. *Scientific Monthly*, *30*, 114–126.

Cattell, J. M., & Farrand, L. (1896). Physical and mental measurements of the students of Columbia University. *Psychological Review*, *3*, 618–648.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing [Abstract]. *Psychological Bulletin*, *38*, 592.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*, 1–22.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.

Cattell, R. B. (Ed.). (1987). *Intelligence: Its structure, growth and action.* New York: Elsevier Science.

Cattell, R. B. (1998). Where is intelligence?: Some answers from the triadic theory. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 29–38). Mahwah, NJ: Erlbaum.

Chapman, J. C. (1921). *Trade tests: The scientific measurement of trade proficiency.* New York: Holt.

Chapman, P. D. (1988). *Schools as sorters: Lewis M. Terman, applied psychology, and the intelligence testing movement, 1890–1930.* New York: New York University Press.

Chattin, S. H., & Bracken, B. A. (1989). School psychologists' evaluation of the K-ABC, McCarthy scales, Stanford–Binet IV, and WISC-R. *Journal of Psychoeducational Assessment*, *7*, 112–130.

Cody, M. S., & Prieto, L. R. (2000). Teaching intelligence testing in APA-accredited programs: A national survey. *Teaching of Psychology*, *27*, 190–194.

Combe, G. (1830). *A system of phrenology* (3rd ed.). Edinburgh: Anderson.

Combe, G. (1853). *A system of phrenology* (5th ed.). Edinburgh: Maclachlan & Stewart.

Committee on Classification of Personnel in the Army. (1919a). *The personnel system of the United States Army: Vol. 1. History of the personnel system.* Washington, DC: U.S. Government Printing Office.

Committee on Classification of Personnel in the Army. (1919b). *The personnel system of the United States Army: Vol. 2. The personnel manual.* Washington, DC: U.S. Government Printing Office.

Cornebise, A. E. (1997). *Soldier-scholars: Higher education in the AEF, 1917–1919.* Philadelphia: American Philosophical Society.

Crime clinics growing. (1922, June 18). *New York Times*, p. 75.

Cronbach, L. J. (1949). *Essentials of psychological testing.* New York: Harper & Brothers.

Cureton, E. (1941). Reviewed work(s): *The Measurement of Adult Intelligence* by David Wechsler. *American Journal of Psychology*, *54*, 154.

Dahlstrom, W. G. (1985). The development of psychological testing. In G. A. Kimble & K. Schlesinger (Eds.), *Topics in the history of psychology* (Vol. 2, pp. 63–113). Hillsdale, NJ: Erlbaum.

Dallenbach, K. M. (1915). The history and derivation of the word "function" as a systematic term in psychology. *American Journal of Psychology*, *26*, 473–484.

Darwin, C. R. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: Murray.

Darwin, C. R. (1871). *The descent of man, and selection in relation to sex* (Vol. 1). London: Murray.

Das, J. P. (1986). On definition of intelligence. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 55–56). Norwood, NJ: Ablex.

Deary, I. J. (1994). Sensory discrimination and intelligence: Postmortem or resurrection? *American Journal of Psychology*, *107*, 95–115.

Deary, I. J., Bell, P. J., Bell, A. J., Campbell, M. L., & Fazal, N. D. (2004). Sensory discrimination and intelligence: Testing Spearman's other hypothesis. *American Journal of Psychology*, *117*, 1–18.

Deary, I. J., Lawn, M., & Bartholomew, D. J. (2008). A conversation between Charles Spearman, Godfrey Thomson, and Edward L. Thorndike: The International Examinations Inquiry Meetings 1931–1938. *History of Psychology*, *11*, 122–142.

Deaths reported Aug. 23. (1903, August 24). *New York Times*, p. 7.

Detterman, D. K. (1986). Human intelligence is a complex system of separate processes. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?:*

Contemporary viewpoints on its nature and definition (pp. 57–61). Norwood, NJ: Ablex.

Dewey, J. (1904). The philosophical work of Herbert Spencer. *Philosophical Review*, *13*, 159–175.

Dewsbury, D. A. (1996). Robert M. Yerkes: A psychobiologist with a plan. In G. A. Kimble, C. A. Boneau, & M. Wertheimer (Eds.), *Portraits of pioneers in psychology* (Vol. 2, pp. 87–105). Washington, DC: American Psychological Association.

Diamond, S. (1977). Francis Galton and American psychology. *Annals of the New York Academy of Sciences*, *291*, 47–55.

Dickson, V. E. (1919). Report of the Department of Research. In *Superintendent's annual report, Oakland Public Schools, 1917–1918* (pp. 174–247). Oakland, CA: Tribune.

Dickson, V. E. (1923). *Mental tests and the classroom teacher.* Yonkers-on-Hudson, NY: World Book.

DuBois, P. H. (1970). Varieties of psychological test homogeneity. *American Psychologist*, *25*, 532–536.

Ebbinghaus, H. (1908). *Psychology: An elementary textbook* (M. Meyer, Trans.). Boston: Heath.

Education for American soldiers in France. (1919). *Scientific Monthly*, *8*, 475–477.

Edwards, A. E. (Ed.). (1974). *Selected papers of David Wechsler.* New York: Academic Press.

Fancher, R. E. (1998). Alfred Binet, general psychologist. In G. A. Kimble & M. Wertheimer (Eds.), *Portraits of pioneers in psychology* (Vol. 3, pp. 67–83). Washington, DC: American Psychological Association.

Ferguson, L. W. (1962). Walter Dill Scott: First industrial psychologist. *Heritage of Industrial Psychology*, *1*, 1–10.

Ferguson, L. W. (1963a). Bureau of Salesmanship Research: Walter Dill Scott, Director. *Heritage of Industrial Psychology*, *5*, 53–66.

Ferguson, L. W. (1963b). Psychology and the Army: Classification of personnel (1). *Heritage of Industrial Psychology*, *10*, 141–167.

Ferguson, L. W. (1963c). Psychology and the Army: Introduction of the rating scale. *Heritage of Industrial Psychology*, *9*, 125–139.

Ferguson, L. W. (1964a). Division of applied psychology: Carnegie Institute of Technology. *Heritage of Industrial Psychology*, *3*, 25–34.

Ferguson, L. W. (1964b). Psychology and the Army: Classification of personnel (2). *Heritage of Industrial Psychology*, *11*, 169–205.

Ferguson, L. W. (1964c). Psychology and the Army: Classification of personnel (3). *Heritage of Industrial Psychology*, *12*, 209–246.

Frank, G. (1983). *The Wechsler enterprise: An assessment of the development, structure, and use of the Wechsler tests of intelligence*. New York: Pergamon Press.

Franz, S. I. (1912). New phrenology. *Science (New Series)*, *35*, 321–328.

Friedlander, H. (1995). *The origins of Nazi genocide: From euthanasia to the final solution.* Chapel Hill: University of North Carolina Press.

Fuster, J. M. (2008). *The prefrontal cortex* (4th ed.). Amsterdam: Elsevier.

Gall, F. J. (1857). Letter to Joseph Fr[eiherr] von Retzer, upon the functions of the brain, in man and animals. In D. G. Goyder (Ed. & Trans.), *My battle for life: The autobiography of a phrenologist* (pp. 143–152). London: Simpkin, Marshall. (Original work published October 1, 1798)

Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences*. London: Macmillan.

Galton, F. (1883). *Inquiries into human faculty and its development.* London: Macmillan.

Galton, F. (1884, June 5). Notes. "Rede" Lecture: The measurement of human faculty. *Nature*, *30*, 129–131.

Galton, F. (1890a). Remarks [after J. M. Cattell's "Mental tests and measurements"]. *Mind*, *15*, 380–381.

Galton, F. (1890b). Why do we measure mankind? *Lippincott's Monthly Magazine*, *45*, 236–241.

Galton, F. (1891). Useful anthropometry. In *Proceedings of the American Association for the Advancement of Physical Education at its Sixth Annual Meeting, April 3 and 4, 1891* (pp. 51–58). Ithaca, NY: Andrus & Church.

Galton, F. (1892). Retrospect of work done at my anthropometric laboratory at South Kensington. *Journal of the Anthropological Institute*, *21*, 32–35.

Galton, F. (1907). *Inquiries into human faculty and its development* (2nd ed.). London: Dent.

Galton, F. (1930). Kantsaywhere. In K. Pearson (Ed.), *The life, letters and labours of Francis Galton* (Vol. 3A, pp. 414–424). London: Cambridge University Press.

Gay, C. J. (1948). The blind look at an elephant: 10 experts, 25 "people" define "intelligence." *The Clearing House*, *22*, 263–266.

Gilbert, J. A. (1894). Researches on the mental and physical development of school children. In E. W. Scripture (Ed.), *Studies from the Yale Psychological Laboratory* (Vol. 2, pp. 40–100). New Haven, CT: Yale University.

Goddard, H. H. (1916). Editor's introduction. In H. H. Goddard (Ed.), *The development of intelligence in children (the Binet–Simon Scale)* (pp. 5–8). Baltimore: Williams & Wilkins.

Gottfredson, L. S. (1994, December 13). Mainstream science on intelligence. *Wall Street Journal*, p. A18.

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, *24*, 13–23.

Gottfredson, L. S. (2007). Innovation, fatal accidents, and the evolution of general intelligence. In M. J. Roberts (Ed.), *Integrating the mind: Domain general versus domain specific processes in higher cognition* (pp. 387–425). Hove, UK: Psychology Press.

Gottfredson, L. S. (2010). Lessons in academic freedom as lived experience. *Personality and Individual Differences*, *49*, 272–280.

Gottfredson, L. S., & Saklofske, D. H. (2009). Intelligence: Foundation and issues in assessment. *Canadian Psychology*, *50*, 183–195.

Gould, S. J. (1996). *The mismeasure of man* (rev. ed.). New York: Norton.

Haggerty, M. E. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology*, *12*, 212–216.

Heaton, H. [Artist] (1922). The pupil becomes an individual [Illustration]. *American School Board Journal*, *64*(4), cover page.

Henmon, V. A. C. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology*, *12*, 195–198.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Hildreth, G. H. (1939). *A bibliography of mental tests and rating scales* (2nd ed.). New York: Psychological Corporation.

Hines, H. C. (1922). Measuring the intelligence of school pupils. *American School Board Journal*, *64*(4), 35–37, 135.

Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities*. Unpublished doctoral dissertation, University of Illinois at Urbana–Champaign. (AAT 6507113)

Horn, J. L. (1976). Human abilities: A review of research and theory in the early 1970s. *Annual Review of Psychology*, *27*, 437–485.

Horn, J. L. (1986). Some thoughts about intelligence. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 91–96). Norwood, NJ: Ablex.

Horn, J. L. (2001). Raymond Bernard Cattell (1905–1998). *American Psychologist*, *56*, 71–72.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology*, *57*, 253–270.

Horn, J. L., & Noll, J. (1994). A system for understanding cognitive capabilities: A theory and the evidence on which it is based. In D. K. Detterman (Ed.), *Current topics in human intelligence: Vol. 4. Theories of intelligence* (pp. 151–203). Norwood, NJ: Ablex.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). New York: Guilford Press.

Hunt, E. (1986). The Heffalump of intelligence. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 101–107). Norwood, NJ: Ablex.

Israel Wechsler, neurologist, dies. (1962, December 7). *New York Times*, p. 39.

Jastrow, J. (1893). The section of psychology. In M. P. Hardy (Ed.), *Official catalogue—World's Columbian Exposition* (Part 7, pp. 50–60). Chicago: Conkey.

Jastrow, J. (1901). Some currents and undercurrents in psychology. *Psychological Review*, *8*, 1–26.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Jensen, A. R. (2004). Obituary: John Bissell Carroll. *Intelligence*, *32*, 1–5.

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Amsterdam: Elsevier.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools*, *47*, 635–650.

Kevles, D. J. (1968). Testing the Army's intelligence: Psychologists and the military in World War I. *Journal of American History*, *55*, 565–581.

Kissman, J. (1948). The immigration of Romanian Jews up to 1914. In N. Shlomo (Ed.), *YIVO annual of Jewish social science* (Vols. 2–3, pp. 160–179). New York: Yiddish Scientific Institute.

Kyllonen, P. C., Roberts, R. D., & Stankov, L. (Eds.). (2008). *Extending intelligence: Enhancement and new constructs*. New York: Erlbaum.

Langfeld, H. S. (1924). The Seventh International Congress of Psychology. *American Journal of Psychology*, *35*, 148–153.

Lewes, G. H. (1867). *The history of philosophy: From Thales to Comte* (Vol. 2, 3rd ed.). London: Longmans, Green.

Lippmann, W. (1922a). The abuse of the tests. *The New Republic*, *32*, 297–298.

Lippmann, W. (1922b). A future for the tests. *The New Republic*, *33*, 9–10.

Lippmann, W. (1922c). The mental age of Americans. *The New Republic*, *32*, 213–215.

Lippmann, W. (1922d). The mystery of the "A" men. *The New Republic*, *32*, 246–248.

Lippmann, W. (1922e). The reliability of intelligence tests. *The New Republic*, *32*, 275–277.

Lippmann, W. (1922f). Tests of hereditary intelligence. *The New Republic*, *32*, 328–330.

Lippmann, W. (1923). The great confusion: A reply to Mr. Terman. *The New Republic*, *33*, 145–146.

Littman, R. A. (2004). Mental tests and fossils. *Journal of the History of the Behavioral Sciences*, *40*, 423–431.

Lombardo, P. A. (2008). *Three generations, no imbeciles: Eugenics, the Supreme Court, and Buck v. Bell.* Baltimore: Johns Hopkins University Press.

Lorge, I. (1943). Review of *The measurement of adult intelligence*. *Journal of Consulting Psychology*, *7*, 167–168.

Louttit, C. M., & Browne, C. G. (1947). The use of psychometric instruments in psychological clinics. *Journal of Consulting Psychology*, *11*, 49–54.

Lubin, B., Wallis, R. R., & Paine, C. (1971). Patterns of psychological test usage in the United States: 1935–1969. *Professional Psychology*, *2*, 70–74.

Lynch, E. C. (1968). Walter Dill Scott: Pioneer industrial psychologist. *Business History Review*, *42*, 149–170.

Matarazzo, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). New York: Oxford University Press.

Matarazzo, J. D. (1981). Obituary: David Wechsler (1896–1981). *American Psychologist*, *36*, 1542–1543.

May, M. A. (1920). The psychological examination of conscientious objectors. *American Journal of Psychology*, *31*, 152–165.

McArdle, J. J. (2007). Obituary: John Leonard Horn 1928–2006. *Intelligence*, *35*, 517–518.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1–10.

McNemar, Q. (1945). Reviewed work(s): *The Measurement of Adult Intelligence* by David Wechsler. *American Journal of Psychology*, *58*, 420–422.

Minton, H. L. (1988). *Lewis M. Terman: Pioneer in psychological testing.* New York: New York University Press.

Morgan, C. L. (1910). Instinct and intelligence. *British Journal of Psychology*, *3*, 219–229.

Myers, C. S. (1910). Instinct and intelligence. *British Journal of Psychology*, *3*, 209–218.

Napoli, D. S. (1981). *Architects of adjustment; The history of the psychological profession in the United States.* Port Washington, NY: Kennikat Press.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*, 77–101.

Newton, J. H., & McGrew, K. S. (2010). Introduction to the special issue: Current research in Cattell–Horn–Carroll-based assessment. *Psychology in the Schools*, *47*, 621–634.

Nicolas, S., & Ferrand, L. (2002). Alfred Binet and higher education. *History of Psychology*, *5*, 264–283.

Oehrn, A. (1896). Experimentelle Studien zur Individualpsychologie. *Psychologische Arbeiten*, *1*, 92–151.

Old Bellevue and new. (1930, June 20). *New York Times*, p. 19.

Otis, A. S. (1917). A criticism of the Yerkes–Bridges Point Scale, with alternative suggestions. *Journal of Educational Psychology*, *8*, 129–150.

Pass alien slacker bill. (1918, February 28). *New York Times*, p. 5.

Payne, A. F. (1920). The scientific selection of men. *Scientific Monthly*, *11*, 544–547.

Pearson, K. (Ed.). (1924). *The life, letters and labours of Francis Galton: Vol. 2. Researchers of middle life*. London: Cambridge University Press.

Pearson, K. (Ed.). (1930). *The life, letters and labours of Francis Galton: Vol. 3A. Correlation, personal identification and eugenics.* London: Cambridge University Press.

Pellegrino, J. W. (1986). Intelligence: The interaction of culture and cognitive processes. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 113–116). Norwood, NJ: Ablex.

Peterson, J. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology*, *12*, 198–201.

Pintner, R. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology*, *12*, 139–143.

Prendergast, P. (1844). A "page" of phrenology. *The Illuminated Magazine*, *11*, 17–20.

Rabin, A. I. (1945). The use of the Wechsler–Bellevue scales with normal and abnormal persons. *Psychological Bulletin*, *42*, 410–422.

Rabin, A. I., & Guertin, W. H. (1951). Research with the Wechsler–Bellevue test: 1945–1950. *Psychological Bulletin*, *48*, 211–248.

Rapaport, D. (1945). The new Army Individual Test of general mental ability. *Bulletin of the Menninger Clinic*, *9*, 107–110.

Reeve, C. L., & Charles, J. E. (2008). Survey of opinions on the primacy of *g* and social consequences of ability testing: A comparison of expert and non-expert views. *Intelligence*, *36*, 681–688.

Rock, G. (1956). *The history of the American Field Service, 1920–1955.* New York: American Field Service.

Roid, G. H. (2003). *Stanford–Binet Intelligence Scales, Fifth Edition*. Itasca, IL: Riverside.

Routh, D. K. (2000). Clinical psychology training: A history of ideas and practices prior to 1946. *American Psychologist*, *55*, 236–241.

Samelson, F. (1977). World War I intelligence testing and the development of psychology. *Journal of the History of the Behavioral Sciences*, *13*, 274–282.

Scarr, S. (1986). Intelligence: Revisited. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 117–120). Norwood, NJ: Ablex.

Schneider, W. H. (1992). After Binet: French intelligence testing, 1900–1950. *Journal of the History of the Behavioral Sciences*, *28*, 111–132.

Scott, J. B. (1918). The amendment of the Naturalization and Citizenship Acts with respect to military service. *American Journal of International Law, 12*, 613–619.

Scott, W. D. (1916). Selection of employees by means of quantitative determinations. *Annals of the American Academy of Political and Social Science*, *65*, 182–193.

Sea Fox wreckage. (1931, May 28). *New York Times*, p. 1.

Seagoe, M. V. (1975). *Terman and the gifted.* Los Altos, CA: Kaufmann.

Seidenfeld, M. A. (1942). The Adjutant General's School and the training of psychological personnel for the Army. *Psychological Bulletin*, *39*, 381–384.

Sharp, S. E. (1899). Individual psychology: A study in psychological method. *American Journal of Psychology*, *10*, 329–391.

Siegler, R. S. (1992). The other Alfred Binet. *Developmental Psychology*, *28*, 179–190.

Simpson, D. (2005). Phrenology and the neurosciences: Contributions of F. J. Gall and J. G. Spurzheim. *Australian and New Zealand Journal of Surgery*, *75*, 475–482.

Sisson, E. D. (1948). The Personnel Research Program of the Adjutant General's Office of the United States Army. *Review of Educational Research*, *18*, 575–614.

Smith, M. L. (2000). By way of Canada: U.S. records of Immigration across the U.S.–Canadian border, 1895–1954 (St. Albans Lists). *Prologue*, *32*, 192–199.

Snow, R. E. (1986). On intelligence. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 133–139). Norwood, NJ: Ablex.

Sokal, M. M. (Ed.). (1981). *An education in psychology: James McKeen Cattell's journal and letters from Germany and England, 1880–1888.* Cambridge, MA: MIT Press.

Sokal, M. M. (Ed.). (1987). *Psychological testing and American society, 1890–1930.* New Brunswick, NJ: Rutgers University Press.

Sokal, M. M. (2006). James McKeen Cattell: Achievement and alienation. In D. A. Dewsbury, L. T. Benjamin, & M. Wertheimer (Eds.), *Portraits of pioneers in psychology* (Vol. 6, pp. 18–35). Mahwah, NJ: Erlbaum.

Spanier, M., & Spanier, D. (1902, September 17). Canadian passenger manifest for the S. S. *Lake Megantic* (sailed from Liverpool 9 September, arrived 17). *Québec Ports, Immigrant Passenger Lists, 1865–1935 (RG 76): Québec 1865–1921 (includes Montreal). Microfilm Reel T-481 (1902 25 Aug to 1903 July 28).* Ottawa: Library and Archives Canada.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.

Spearman, C. (1923). *The nature of "intelligence" and principles of cognition.* London: Macmillan.

Spearman, C. (1927). *The abilities of man: Their nature and measurement.* New York: Macmillan.

Spearman, C., & Wynn Jones, L. (1950). *Human ability: A continuation of "The Abilities of Man."* London: Macmillan.

Spencer, H. (1855). *The principles of psychology.* London: Longman, Brown, Green, & Longmans.

Spencer, H. (1864). *The principles of biology* (Vol. 1). London: Williams & Norgate.

Spring, J. H. (1972). Psychologists and the war: The meaning of intelligence in the Alpha and Beta tests. *History of Education Quarterly*, *12*, 3–15.

Spurzheim, J. G. (1815). *The physiognomical system of Drs. Gall and Spurzheim; founded on an anatomical and physiological examination of the nervous system in general, and of the brain in particular; and indicating the dispositions and manifestations of the mind* (2nd ed.). London: Baldwin, Cradock, & Joy.

Staff, Personnel Research Section. (1942, March). *Analysis of Wechsler Self-Administering Test data* (PRS Report No. 286). Washington, DC: U.S. War Department, Adjutant General's Office.

Staff, Personnel Research Section. (1944). The new Army Individual Test of general mental ability. *Psychological Bulletin*, *41*, 532–538.

Staff, Personnel Research Section. (1945). The Army General Classification Test. *Psychological Bulletin*, *42*, 760–768.

Stein, M. (2004). The establishment of the Department of Psychiatry in the Mount Sinai Hospital: A conflict between neurology and psychiatry. *Journal of the History of the Behavioral Sciences*, *40*, 285–309.

Stern, W. (1914). *The psychological methods of testing intelligence* (Educational Psychology Monographs No. 13; G. M. Whipple, Trans.). Baltimore: Warwick & York. (Original work published 1912)

Sternberg, R. J. (1986). Intelligence is mental self-government. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 141–148). Norwood, NJ: Ablex.

Sternberg, R. J. (1994). 468 factor-analyzed data sets:

What they tell us and don't tell us about human intelligence. Reviewed work(s): *Human Cognitive Abilities: A Survey of Factor-Analytic Studies* by J. B. Carroll. *Psychological Science*, *5*, 63–65.

Sternberg, R. J., & Berg, C. A. (1986). Quantitative integration: Definitions of Intelligence: A comparison of the 1921 and 1986 symposia. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence?: Contemporary viewpoints on its nature and definition* (pp. 155–162). Norwood, NJ: Ablex.

Sternberg, R. J., & Detterman, D. K. (Eds.). (1986). *What is intelligence?: Contemporary viewpoints on its nature and definition*. Norwood, NJ: Ablex.

Strong, E. K. (1955). Walter Dill Scott: 1869–1955. *American Journal of Psychology*, *68*, 682–683.

Terman, L. M. (1904). A preliminary study in the psychology and pedagogy of leadership. *Pedagogical Seminary*, *11*, 413–451.

Terman, L. M. (1906). Genius and stupidity: A study of some of the intellectual processes of seven "bright" and seven "stupid" boys. *Pedagogical Seminary*, *13*, 307–373.

Terman, L. M. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the Binet–Simon Intelligence Scale.* Boston: Houghton Mifflin.

Terman, L. M. (1917). Feeble minded children in the schools. In F. C. Nelles (Chair), *Report of 1915 legislature committee on mental deficiency and the proposed institution for the care of feeble-minded and epileptic persons* (pp. 45–52). Whittier, CA: Whittier State School, Department of Printing Instruction.

Terman, L. M. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology*, *12*, 127–133.

Terman, L. M. (1922a). The great conspiracy, or the impulse imperious of intelligence testers, psychoanalyzed and exposed by Mr. Lippmann. *The New Republic*, *33*, 116–120.

Terman, L. M. (1922b). Were we born that way? *The World's Work*, *44*, 655–660.

Terman, L. M. (1932). Trails to psychology. In C. Murchison (Ed.), *A history of psychology in autobiography* (Vol. 2, pp. 297–331). Worcester, MA: Clark University Press.

Terman, L. M., & Childs, H. G. (1912a). A tentative revision and extension of the Binet–Simon measuring scale of intelligence. *Journal of Educational Psychology*, *3*, 61–74.

Terman, L. M., & Childs, H. G. (1912b). A tentative revision and extension of the Binet–Simon measuring scale of intelligence: Part II. Supplementary tests. 1. Generalization test: interpretation of fables. *Journal of Educational Psychology*, *3*, 133–143.

Terman, L. M., & Childs, H. G. (1912c). A tentative revision and extension of the Binet–Simon measuring scale of intelligence: Part II. Supplementary tests—continued. *Journal of Educational Psychology*, *3*, 198–208.

Terman, L. M., & Childs, H. G. (1912d). A tentative revision and extension of the Binet–Simon measuring scale of intelligence: Part III. Summary and criticisms. *Journal of Educational Psychology*, *3*, 277–289.

Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence: A guide to the administration of the new Revised Stanford–Binet tests of intelligence*. Boston: Houghton Mifflin.

Terman, L. M., & Merrill, M. A. (1960). *Stanford–Binet Intelligence Scale: Manual for the third revision, Form L-M*. Boston: Houghton Mifflin.

Terman, L. M., & Merrill, M. A. (1973). *Stanford–Binet Intelligence Scale: Manual for the third revision, Form L-M (1972 table of norms by R. L. Thorndike)*. Boston: Houghton Mifflin.

Thorndike, E. L. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology*, *12*, 124–127.

Thorndike, E. L., Lay, W., & Dean, P. R. (1909). The relation of accuracy in sensory discrimination to general intelligence. *American Journal of Psychology*, *20*, 364–369.

Thorndike, R. L. (1975). Mr. Binet's test 70 years later. *Educational Researcher*, *4*, 3–7.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford–Binet Intelligence Scale: Fourth Edition*. Chicago: Riverside.

Thorndike, R. M., & Lohman, D. F. (1990). *A century of ability testing.* Chicago: Riverside.

Thorne, F. C. (1976). Reflections on the golden age of Columbia psychology. *Journal of the History of the Behavioral Sciences*, *12*, 159–165.

Thurstone, L. L. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology*, *12*, 201–207.

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.

Thurstone, L. L. (1945). Testing intelligence and aptitudes. *Hygeia*, *23*, 32–36, 50, 52, 54.

Thurstone, L. L. (1947). *Multiple factor analysis: A development and expansion of the vectors of mind.* Chicago: University of Chicago Press.

Thurstone, L. L., & Thurstone, T. G. (1941). Factorial studies of intelligence. *Psychometric Monographs*, *2*.

Toops, H. A. (1921). *Trade tests in education*. New York: Teachers College, Columbia University.

Tucker, W. H. (2009). *The Cattell controversy: Race, science, and ideology*. Urbana: University of Illinois Press.

Tuddenham, R. D. (1974). Review: Fame and oblivion. *Science (New Series)*, *183*, 1071–1072.

206 get degrees at City College. (1916, June 23). *New York Times*, p. 7.

U.S. maintains great schools on foreign soil: Government trains soldiers for civilian life—thousands are enrolled. (1919, June 12). *Lima Daily News* [Lima, OH], p. 5.

Venezky, R. L. (2002). The history of reading research. In P. D. Pearson (Ed.), *Handbook of reading research* (Vol. 1, pp. 3–38). Mahwah, NJ: Erlbaum.

Vernon, P. E. (1950). *The structure of human abilities*. London: Methuen.

Vernon, P. E. (1961). *The structure of human abilities* (2nd ed.). London: Methuen.

von Mayrhauser, R. T. (1986). *The triumph of utility: The forgotten clash of American psychologies in World War I*. Unpublished doctoral dissertation, University of Chicago. (AAT T-29951)

von Mayrhauser, R. T. (1987). The manager, the medic, and the mediator: The clash of professional psychological styles and the wartime origins of group mental testing. In M. M. Sokal (Ed.), *Psychological testing and American society, 1890–1930* (pp. 128–157). New Brunswick, NJ: Rutgers University Press.

von Mayrhauser, R. T. (1989). Making intelligence functional: Walter Dill Scott and applied psychological testing in World War I. *Journal of the History of the Behavioral Sciences*, *25*, 60–72.

Walsh, A. A. (1976). The "new science of the mind" and the Philadelphia physicians in the early 1800s. *Transactions and Studies of the College of Physicians of Philadelphia*, *43*, 397–413.

Wasserman, J. D., & Maccubbin, E. M. (2002, August). *David Wechsler at The Psychological Corporation, 1925–1927: Chorus girls and taxi drivers*. Paper presented at the annual meeting of the American Psychological Association, Chicago.

Wechsler, D. (1917a). A study of retention in Korsakoff psychosis. *Psychiatric Bulletin*, *2*, 403–451.

Wechsler, D. (1917b, June 5). *World War I Selective Service System Draft Registration Cards, 1917–1918* (Administration. M1509, Roll: 1765679, Draft Board: 96). Washington, DC: National Archives and Records.

Wechsler, D. (1918a, October 4). Passport Application No. 124984. In *Passport Applications, 1906–March 31, 1925*. Washington, DC: National Archives and Records Administration.

Wechsler, D. (1918b, June 7). *Petition for naturalization, David Wechsler. United States District Court of the Northern District of Georgia, Rome Division*. Morrow, GA: NARA Southeast Region.

Wechsler, D. (1921). Report of amendment of passport, August 18, 1921. In *Passport Applications, 1906–March 31, 1925*. Washington, DC: National Archives and Records Administration.

Wechsler, D. (1922). Quelques remarques sur la psychopathologie de l'indecision [Some remarks on the psychopathology of indecision]. *Journal de Psychologie*, *19*, 47–54.

Wechsler, D. (1925). The measurement of emotional reactions. *Archives of Psychology*, *76*, 1–181.

Wechsler, D. (1926). Tests for taxicab drivers. *Journal of Personnel Research*, *5*, 24–30.

Wechsler, D. (1927). Psychometric tests. In I. S. Wechsler, *Textbook of clinical neurology* (pp. 104–116). Philadelphia: Saunders.

Wechsler, D. (1932). Analytic use of the Army Alpha examination. *Journal of Applied Psychology*, *16*, 254–256.

Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.

Wechsler, D. (1944). The psychologist in the psychiatric hospital. *Journal of Consulting Psychology*, *8*, 281–285.

Wechsler, D. (1946). *Wechsler–Bellevue Intelligence Scale, Form II. Manual for administering and scoring the test*. New York: The Psychological Corporation.

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children (WISC)*. New York: Psychological Corporation.

Wechsler, D. (1955). *Wechsler Adult Intelligence Scale—Restandardized (WAIS)*. New York: Psychological Corporation.

Wechsler, D. (1961). Conformity and the idea of being well born. In I. A. Berg & B. M. Bass (Eds.), *Conformity and deviation* (pp. 412–423). New York: Harper & Row.

Wechsler, D. (1967). *Wechsler Preschool and Primary Scale of Intelligence (WPPSI)*. New York: The Psychological Corporation.

Wechsler, D. (1981). The psychometric tradition: Developing the Wechsler Adult Intelligence Scale. *Contemporary Educational Psychology*, 6, 82–85.

Wechsler, D., Doppelt, J. E., & Lennon, R. T. (1975). *A conversation with David Wechsler* (Unpublished transcript of interview). San Antonio, TX: The Psychological Corporation.

Wechsler, I. (1957). A neurologist's point of view. *Medical Clinics of North America*, *41*, 1111–1121.

Wechsler, M. (1903, August 23). *Certificate and record of death, Certificate No. 24787. City of New York, Department of Health, State of New York*. New York: New York City Department of Records and Information Services Municipal Archives.

Weider, A. (2006). *A Weider Weltanschauung: A wider perspective of psychology*. Bloomington, IN: Xlibris.

Wexler, L. (1906, October 31). *Certificate and record of* *death, Certificate No. 33534. City of New York, Department of Health, State of New York.* New York: New York City Department of Records and Information Services Municipal Archives.

Whipple, G. M. (1915). *Manual of mental and physical tests: Part 2. Complex processes* (2nd ed.). Baltimore: Warwick & York.

Wilson, M. S., & Reschly, D. J. (1996). Assessment in school psychology training and practice. *School Psychology Review*, *25*, 9–23.

Wilson, W. (1917). Address of the President of the United States delivered at a joint session of the two houses of Congress, April 2, 1917. In *65th Congress, 1st Session, Senate* (Document No. 5, Serial No. 7264, pp. 3–8). Washington, DC: U.S. Government Printing Office.

Wissler, C. (1901). The correlation of mental and physical tests. *Psychological Monographs*, *3*(6), 1–62.

Wolf, T. (1961). An individual who made a difference. *American Psychologist*, *16*, 245–248.

Wolf, T. (1966). Intuition and experiment: Alfred Binet's first efforts in child psychology. *Journal of the History of the Behavioral Sciences*, *2*, 233–239.

Wolf, T. (1969). The emergence of Binet's conceptions and measurement of intelligence: II. A case history of the creative process. *Journal of the History of the Behavioral Sciences*, *5*, 207–237.

Wolf, T. (1973). *Alfred Binet.* Chicago: University of Chicago Press.

Woodrow, H. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology*, *12*, 207–210.

Woodworth, R. S. (1942). *The Columbia University psychological laboratory.* New York: Columbia University.

Yerkes, R. M. (1918a). Measuring the mental strength of an army. *Proceedings of the National Academy of Sciences USA*, *4*, 295–297.

Yerkes, R. M. (1918b, May 3). Memorandum for Mr. Garrett. *Records of the Office of the Surgeon General (Army) (RG 112).* College Park, MD: National Archives and Records Administration.

Yerkes, R. M. (1919). Report of the Psychology Committee of the National Research Council. *Psychological Review*, *26*, 83–149.

Yerkes, R. M. (1920). What psychology contributed to the war. In R. M. Yerkes (Ed.), *The new world of science: Its development during the war* (pp. 364–389). New York: Century.

Yerkes, R. M. (Ed.). (1921). *Memoirs of the National Academy of Sciences: Vol. 15. Psychological examining in the United States Army.* Washington, DC: U.S. Government Printing Office.

Yerkes, R. M., Bridges, J. W., & Hardwick, R. S. (1915). *A point scale for measuring mental ability.* Baltimore: Warwick & York.

Yoakum, C. S., & Yerkes, R. M. (1920). *Army mental tests.* New York: Holt.

Yockelson, M. (1998). They answered the call: Military service in the United States Army during World War I, 1917–1919. *Prologue: Quarterly of the National Archives and Records Administration*, *30*, 228–234.

Zazzo, R. (1993). Alfred Binet (1857–1911). *Prospects: The Quarterly Review of Comparative Education*, *23*, 101–112.

Zeidner, J., & Drucker, A. J. (1988). *Behavioral science in the army: A corporate history of the Army Research Institute.* Arlington, VA: United States Army Research Institute for the Behavioral and Social Sciences.

Zenderland, L. (1998). *Measuring minds: Henry Herbert Goddard and the origins of American intelligence testing.* New York: Cambridge University Press.