

Running Head: MENTAL ABILITY TESTS

**A History of Mental Ability Tests and Theories**

**John D. Wasserman**

Independent Practice, Burke, Virginia

**Alan S. Kaufman**

Yale University School of Medicine

**DO NOT QUOTE UNTIL PUBLISHED**

### **Abstract**

The concepts of measurement and theory have always been central to psychological science. This chapter reviews the history of applied mental tests and the ideas behind them, with a specific emphasis on individually-administered intellectual measures in the era of scientific psychology (i.e., after Wundt). The theoretical underpinnings associated with mental tests are summarized, and test/theory falsifications are discussed. Beginning with the contributions of Francis Galton and J. McKeen Cattell and continuing through the present day, the topics discussed include anthropometric testing, Charles Spearman's two-factor theory and general intelligence factor, Alfred Binet and David Wechsler's pragmatic approaches, Raymond B. Cattell and John L. Horn's fluid and crystallized intelligence, John B. Carroll's three stratum model of cognitive abilities, and Alexander R. Luria's conceptualization of brain-based, cognitive processing. The successes of the Binet-Simon, Stanford-Binet, and Wechsler intelligence scales in the United States are suggestive that the factors driving practitioners to use specific tests may be different than the factors driving research and theory. The chapter closes with a discussion about theory-building and falsification in mental testing and the importance of reconciling theory with clinical practice in psychological assessment.

### **Keywords**

Intelligence    mental testing    anthropometrics    cognitive processing    Binet  
Wechsler    Galton    Cattell    Spearman    Horn    Carroll    Luria

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. (J. M. Cattell, 1890, p. 373)

When James McKeen Cattell (1890) introduced the term *mental tests* (above) in the British journal *Mind*, he was already cognizant that measurement was essential to establishing the field of psychology as an emerging experimental science. In Wilhelm Wundt's experimental psychology laboratory at the University of Leipzig, widely considered the birthplace of scientific psychology, Cattell was the first American to earn a Ph.D. in 1886 (Sokal, 1981). During a

sojourn at St. Johns College at the University of Cambridge, Cattell came to know well and work closely with Francis Galton, the father of psychometrics, who launched the first large-scale program of anthropometric testing at the International Health Exhibition in London. Galton's approach emphasized accumulation and analysis of large quantities of normative data, with the capacity to identify individual and group differences (phenomena of little value to Wundt, who sought universals governing elementary mental processes). Upon his return to the United States from England, Cattell set up experimental psychology laboratories and testing programs like those he had seen with Wundt and Galton (see Sokal, 2010).

Cattell would become a forceful lifelong advocate for mental tests based on their potential applied and practical value (see e.g., Cattell, 1923). As a positivist, Cattell was dedicated to the use of quantitative and statistical methods to discover scientific laws governing the natural world. As the first psychologist elected to the National Academy of Sciences (1901) and as the long-time editor and publisher (from 1894-1944) of the journal *Science*, Cattell offered an entrée to the scientific community that was shared, at least symbolically, by all psychologists. Finally, as the founder of The Psychological Corporation, Cattell laid the groundwork for a scientific entrepreneurialism that remains prominent today in the commercial test publishing industry.

Arguably (and ironically), one of Cattell's greatest contributions to psychology as a science may have been the high profile *falsification* of his Galton-influenced anthropometric test battery as a valid measure of intelligence, and his uncharacteristic grace at accepting the apparent research-based outcome. Cattell's principal research initiative at Columbia University was to determine whether a battery of anthropometric tests supplemented by a variety of sensory, motor, and higher cognitive tasks could constitute a measure of intelligence. Beginning in 1894, the

Cattell-Columbia Tests (Cattell & Farrand, 1896; Cattell, 1924) were given to university freshman, with Cattell also having access to student academic records and athletic accomplishments. Beyond seeking to predict academic achievement, Cattell offered no substantive theoretical underpinnings for the test battery; he was skeptical about the value of most theoretical systems and never developed one himself (Gates, 1971). Two investigations effectively concluded that the Cattell-Columbia Tests did not measure intelligence: Stella E. Sharp's (1899) report about the limited value of elementary mental processes relative to more complex processes, and Clark Wissler's (1901) failure to correlate anthropometric battery test scores with academic grades. Not only did these findings discredit Galtonian anthropometric testing in general (Sokal, 2006, p. 29), but they falsified the *mental tests* of the man who coined the term *mental tests*, perhaps setting the stage for critical empirical evaluation as a defining feature of scientific psychology.

In this chapter, we recount the history of mental tests through a focused examination on selected individually-administered tests and measurements in conjunction with the associated ideas and theories (when offered) supporting their development and application. Although there are several important domains in mental testing, our focus is on intelligence tests. In a sense, we are reviewing a history of ideas instead of events, so the success of the World War I Army Mental Tests in launching the testing revolution in the military, schools, and industry is not discussed. We invite readers more interested in the narrative stories and events in the history of mental testing to consider some of our other historical reviews (e.g., Kaufman, 2009; Kaufman & Lichtenberger, 2005; Wasserman, 2012). Here our lens is focused on the history of individually-administered mental tests and the ideas behind them.

### **Of Tests, Theories, and Paradigms**

“Psychology has a long past, yet its real history is short,” wrote Hermann Ebbinghaus in 1908/1911 to describe long-standing and ancient interests in mental processes and behaviors and their relatively brief history of scientific experimental investigation. In this section, we comment on some of the qualities of mental tests and theories that permit psychology to be considered a science. We also discuss some matters from the philosophy of science relating to what constitutes scientific theory and divergent thinking as to how practice advances.

### **Mental Tests and Test Usage Surveys**

It is not always clear what defines a *mental test*, as almost any puzzle, question, or task may potentially be converted into a measure of some mental phenomena. Some of our favorite tests are exceedingly brief (e.g., repetition of the phrase “No ifs, ands, or buts” as a test of conduction aphasia [Geschwind, 1970]; and repetition of the word *Shibboleth* [שִׁבּוֹלֶת] as a way that the Gileadites identified refugee Ephraimites in the Old Testament [*Judges* 12:5-6]). On the other extreme, some systems, such as Gardner’s preschool tasks of multiple intelligences, may require as much as a year to collect (Gardner, Feldman, & Krechevsky, 1998a, 1998b, 1998c). Tests may be administered individually, in groups, online, or telephonically. Tests may be naturalistically observed (i.e., by examiner, other, or self), true-false, multiple choice, short answer, or extended response, with either qualitative or objective scoring. There are few limits to what constitutes a test.

Given the thousands of tests that have been created, it is sobering to find that a relatively small number have dominated applied practice since psychological test usage became the subject of formal surveys. Our review of the most frequently used tests among practicing clinical psychologists from the 1930s and 1940s through the year 2000 shows that a few assessment procedures—the Stanford-Binet and/or the Wechsler intelligence scales (Wechsler-Bellevue,

WAIS, WISC, and their revisions), the *Minnesota Multiphasic Personality Inventory* (MMPI), the *Bender-Gestalt*, human figure drawings, the Rorschach inkblot test, and the *Thematic Apperception Test*—consistently ranked among the top ten tests in terms of usage, indicating that clinical testing practice did not change much in the second half of the twentieth century (Brown *et al.*, 1935; Camara, Nathan, & Puente, 2000; Louttit & Browne, 1947; Lubin, Larsen, & Matarazzo, 1984; Lubin, Wallis, & Paine, 1971; Sundberg, 1961). An examination of test usage surveys over time makes clear that psychologists change their test preferences *slowly*, at best.

### **Theories: Definitions and Criteria**

Theories are essential because they have the capacity to explain empirical findings and predict future findings. According to a formal definition from the National Academy of Sciences, a *theory* in science is “a well-substantiated explanation of some aspect of the natural world that can incorporate facts, laws, inferences, and tested hypotheses” (National Academy of Sciences, 1999). Theories are conceptualized as end points in science, reflecting accumulated understanding that develops from observation, experimentation, and creative reflection.

Historian of science Thomas S. Kuhn (1977) described five characteristics of a good scientific theory: *accuracy* (empirical adequacy via experimentation and observation), *consistency* (coherence with other aspects of the theory and with other theories), *scope* (breadth, with implications extending beyond that which it was initially intended to explain), *simplicity* (economy and parsimony, similar to Occam’s razor), and *fruitfulness* (capacity to disclose new phenomena or new relations among phenomena).

“It must be possible for an empirical scientific system to be refuted by experience,” wrote Karl Popper (1935/2005, p. 18), describing the *falsifiability* criterion that qualifies a theory as empirical or scientific only if it can potentially be tested, refuted, or disproven. From Popper’s

perspective, verification or confirmation of a theory is never entirely possible because a near infinite number of predictions must be tested, while falsification may require only a single reliable finding that contradicts the theory. It is not entirely clear how the criterion of falsifiability operates in actual scientific practice. Gray (2007) points out that accepting the criterion of falsifiability would have precluded initial consideration of the theories of Darwin and Einstein because each of them were in some way at odds with some available evidence: “Applying Popper’s account of scientific method would have killed these theories at birth” (p. 22)

In assessment psychology, it can be hard to find any wholly falsified theories (or tests) or any wholly accepted theories (and tests). As Thomas Kuhn put it, “Whatever scientific truth may be, it is through-and-through relativistic” (Kuhn, 2000, p. 156). Over a century after objective “falsifications,” it is still possible to find persuasive scholarly advocates for Galton and Cattell’s emphasis on elementary processes (reaction time, sensory discrimination) as effective predictors of intelligence (e.g., Deary, 1994; Deary, Bell, Bell, Campbell, & Fazal, 2004; Jensen, 2006). Conversely, over a century after its introduction, Spearman’s (1904a) theory of general intelligence (psychometric *g*) continues to be challenged by intelligence theorists, including the late Raymond B. Cattell and John L. Horn, Howard Gardner, and Robert J. Sternberg, although it appears to represent consensus thought in intelligence (e.g., Gottfredson, 1997a; Reeve & Charles, 2008). The 1996 *Intelligence: Knowns and Unknowns* statement by an APA task force was curiously equivocal on psychometric *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” (emphasis added; Neisser *et al.*, 1996, p. 81).

## **Paradigm Shifts**



A *paradigm* generally refers to the prevailing theoretical framework shared by members of a scientific community at any given time (although any framework may be generically designated as a paradigm). Kuhn (1962/2012) asserted that when an existing paradigm can no longer accommodate inconsistent and anomalous empirical findings, a scientific crisis is precipitated during which time new arguments are delivered, new empirical findings reported, new predictions made, and new theories generated. If a new theory has greater quantitative precision while predicting new phenomena as well as explaining old phenomena, it will have a substantial competitive advantage over the old theory. From this intellectual battle between followers of the new and old paradigms may potentially come a *paradigm shift* and scientific revolution. Kuhn notes that paradigm shifts commonly see some scientists clinging tenaciously to the prevailing paradigm, some undergoing a scientific conversion to the new paradigm, and some who are less entrenched (especially younger scientists) adopting the new paradigm. Kuhn quotes Max Planck as explaining that scientific truths do not necessarily convince their opponents to see the light; instead attrition occurs as the older generation passes and a younger generation grows up familiar with the new paradigm.

Of course, science does not conform neatly to the expectations of Kuhn or Popper. Gray (2007) argues that the history of science consistently shows scientists flouting the rules of scientific method in advancing their fields. In an argument that epistemological anarchy advances scientific progress more than any fixed and universal methodological rules, Feyerabend (1975) advocates for “an anything goes” approach to scientific practice and theory. In contrast, Sokal and Bricmont (1998) hold that theories come to be accepted when they work successfully in practicality.

We now turn our attention to the father of psychometrics, large scale testing, and individual differences, Francis Galton, who formulated the methods and aspirations for the first generations of mental testers, just a few years after Wundt founded the first experimental psychology laboratory in Leipzig, Germany in 1879.

**“Wonderful Form of Cosmic Order”: Galton and Anthropometric Testing**

*Order in Apparent Chaos.*—I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the “Law of Frequency of Error.” The law would be personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along.

So wrote Francis Galton (1889b, p. 66), describing what he saw as the transcendent quality of the normal probability distribution (i.e., the bell curve). Galton (1822-1911) is today remembered as the father of contemporary psychometrics, and he was among the first to realize that human attributes including both physical and mental traits could be quantified and represented in a population distribution. Galton (1885a, 1885c) understood the relative position of individual scores in this probability distribution through the use of percentile ranks, probable error (an antecedent to standard deviation), and deviations from the normative mean or median score. On the basis of his quantitative and psychometric work alone, which proved foundational for the field of psychology, Galton would be remembered as an important innovator.

Galton is also remembered for his unfortunate advocacy of *eugenics*, a term he coined in 1883 to signify “the science of improving stock” (Galton, 1883a, p. 25). Galton saw eugenics as an alternative to natural selection and as a boon to humankind, never anticipating the genocidal horrors that would be seen in Nazi Germany and other countries. A half-cousin of Charles Darwin, Galton was a polymath with a remarkably diverse range of interests, unified only with the publication of Darwin’s (1859) *On the Origin of Species* which Galton (1908) described as creating “a marked epoch in my own development, as it did in that of human thought generally” (p. 287). “I was encouraged by the new [evolutionary] views,” wrote Galton in his autobiography, “to pursue many inquiries which had long interested me, and which clustered round the central topics of Heredity and the possible improvement of the Human Race” (Galton, 1908, p. 288).

Galton’s studies in measurable differences between people (sometimes called *differential psychology*; e.g., Minton & Schneider, 1980), is a clear antecedent to contemporary assessment practice and a major break from Wilhelm Wundt’s emphasis on universal laws of mental processing. The idea of a testing program first appeared in 1865, when Galton thought about a system of competitive examinations for youth “to embrace every important quality of mind and body” (p. 165). Borne out of his interest in the heritability of exceptional qualities, Galton was interested in identifying individuals who were high in intelligence, character, physique, and motivation (Galton, 1884b, 1910). Galton was among the first to see the value in measuring the test performance of special populations, and his work features individuals now considered to be gifted as well as individuals we would now describe as intellectually disabled.

### **Theoretical Underpinnings**

Galton never attempted to formulate a distinctive theory of intelligence, seeing efforts to establish the parameters of intelligence as analogous to plumbing the depth of a lake or sinking shafts in the earth (Galton, 1884b, 1890c). His conviction that sensory discrimination capacity is a good estimate of mental ability was presented in *Inquiries into Human Faculty and its Development* (1883a):

The only information that reaches us concerning outward events appears to pass through the avenue of our senses; and the more perceptive the senses are of difference, the larger is the field upon which our judgment and intelligence can act. (Galton, 1883a, p. 27)

Galton's hypothesis was based on his observations that "the intellectually ablest" (Galton, 1883b, p. 474) have the highest capacity for sensory discrimination, such as judging small differences in weights, while the intellectually disabled have relatively poor discriminative faculties (see also Galton, 1883a). Galton was influenced by the numerous accounts coming out of Wundt's laboratory and others in Germany about the "discriminative powers of the various senses" which he interpreted as signaling the dawn of a new anthropometric science (Galton, 1884b, p. 79). Florence Goodenough (1949, p. 41) later explained that Galton and his followers conceptualized lower- and higher-order mental faculties as coextensive "rungs of the same ladder," believing that a dependable estimate of higher-order faculties could be had by measuring those considered lower order faculties, such as sensory and motor functions.

Galton's anthropometric laboratory focused primarily on physical measurements and sensory-motor performance, but he clearly saw "natural ability" as much more: a multiplicity of faculties (lower order and higher order); ability that encompasses cognition, motivation, and persistence; and, for individuals of high ability, the capacity to work through adversity. Galton was interested in a range of faculties, including memory (Galton, 1869, p. 23; 1879a) mental

imagery/visualization (Galton, 1880), span of apprehension (Galton, 1886), and reaction time (Galton, 1889a, 1890b). In *Inquiries into Human Faculty and its Development* (1883a), he associated intelligence with abstract thinking: “Nothing is a surer sign of high intellectual capacity than the power of quickly seizing and easily manipulating ideas of a very abstract nature” (p. 201).

Galton’s (1879b, 1879c) introspective analyses of his own mental operations reminded him of “the complex system of drains and gas- and water-pipes, flues, bell-wires, and so forth” found in the basement of a house (Galton, 1879b, p. 151). In his daily walk of some 450 yards, he counted his discernible mental processes, perceiving 300 separate objects and generating associations at the rate of about 50 per minute. Rather than trying to isolate elementary processes like Wundt, he saw his anthropometric methods as being concerned with products of thought and action (Galton, 1885b, p. 208).

### **Tests and Measures**

The field of anthropometry encompasses measurements of the human body, including weight, height, and size, but it is clear from that start that Galton hoped to do more. In a March, 1883 privately printed pamphlet, he solicited input about simple instruments for “testing the measuring the efficiency of the various mental of bodily powers” (p. 1); he stated the intent to first measure the “delicacy of the several Senses” and then consider the “Higher Mental Processes” (Galton, 1883c, p. 1).

The battery for which Galton is best known appeared at the Anthropometric Laboratory in the International Health Exhibition beginning in 1884. For a cost of three pence, an individual could undergo the examination and receive a summary card of results at the close. The laboratory was in a space 36 feet long by 6 feet wide, fenced off from the side of a gallery by open lattice-

work and a door at each end. As described by Galton (1884a), the successive stations were in order:

- *Reception desk*
- *Standard color for eyes and hair*
- *Sight: (a) its keenness; (b) the color-sense; (c) judgment in estimating length and squareness*
- *Hearing: (a) its keenness; (b) highest audible note*
- *Touch*
- *Breathing capacity*
- *Swiftmess of blow with fist*
- *Strength: (a) of pull; (b) of squeeze with right and left hands*
- *Height: (a) when sitting; (b) standing in shoes against a wall; (c) the thickness of the heel of the shoe*
- *Span of the arms*
- *Weight*

With the close of the International Health Exhibition, the laboratory moved to the South Kensington Museum, where Galton (1887) added two measures more oriented toward psychology along with one craniology measure:

- *Reaction time to sound and sight*
- *Appreciation of slight differences in weight*
- *Head measurement: horizontal head spanner; vertical head spanner*

Galton's interest in measurement of the head (*craniology*) stemmed from his conviction that a large head houses a large brain, which he presumed had greater mental power than a small brain.

In his studies of eminence and reputation as evidence of natural ability, Galton (1869) wrote: “I look upon social and professional life as a continuous examination” (p. 6). With lifelong achievement being impractical for assessment purposes, Galton considered performance on formal academic literary examinations or mathematical honors as effective proxies for natural ability and intelligence (e.g., Galton, 1869, 1884c, 1891). Fancher (1989, p. 450) documented an unpublished Galton study from 1883 entitled “Notes by FG on the connexion between high examination marks and success in life,” in which Galton tried to develop a correlation tool (before the correlation coefficient had been invented) to describe the relation between Civil Service examination scores and actual merit ratings. Ultimately Galton’s objective was to discover efficient ways by which he could predict likely eminence and life success:

The leading ideas of such a laboratory as I had in view, were that its measurements should effectually “sample” a man with reasonable completeness. It should measure *absolutely* where it was possible, otherwise *relatively* among his class fellows, the quality of each selected faculty. The next step would be to estimate the combined effect of these separately measured faculties in any given proportion, and ultimately to ascertain the degree with which the measurement of sample faculties in youth justifies a prophecy of future success in life, using the word “success” in its most liberal meaning. (Galton, 1908, p. 267)

Galton set up several anthropometric laboratories including one at the University of Cambridge (1884c), but he also proposed to collect anthropometric statistics for children from “schools of all description” (Galton, 1874, p. 311). He envisioned a society with ready access to anthropometric laboratories where a parent could easily have his or her children weighed, measured, photographed, and tested (Galton, 1882). Consequently, Galton may be credited with

the ideas of testing for natural ability, interpreting test battery results according to normative reference, and testing in schools, although it would remain for Alfred Binet to develop a test battery that effectively predicted academic performance.

### **Ultimate Outcomes**

The earliest known scientific study to cast serious doubt upon anthropometry as a measure of mental ability was conducted by John Venn (best remembered for developing the method of Venn diagrams) and was published in 1889 (partial sample) and 1890 (with a complete sample). Venn tested over 2,134 undergraduate students at the University of Cambridge with anthropometric tests provided by Galton, acknowledging that the sample was representative of the upper professional class and performed better, on average, than the participants measured by Galton at South Kensington. On the basis of information provided by their tutors, students were classified into one of three groups: *A* (college scholars and high honors students;  $n=487$ ), *B* (honors students;  $n=913$ ), or *C* (students taking an ordinary degree;  $n=734$ ). Venn (1889) considered this classification “about as good as any such intellectual test can be” (p. 144), inferring that the three groups respectively could be said to have “very good”, “good”, and “indifferent” intellectual capacity.

After examining his findings, Venn (1890) concluded “there is no difference whatever . . . between the physical characteristics of the different intellectual grades. Whether in respect of height, weight, power of squeeze, eyesight, breathing capacity, or head-dimensions, there is no perceptible distinction” (p. 451). In summary, Venn (1890) asserted, “Mental superiority may be considered as perfectly ‘independent’ (in the mathematical sense) of physical . . . It does not appear that intellectual superiority is in the slightest significant degree either correlated with any kind of natural physical superiority or inferiority, or that it tends to incidentally produce any



*general* superiority or inferiority” (pp. 452-453). Galton (1890a) generally accepted Venn’s findings and conclusions, and his interest in anthropometric testing waned after this study. In his memoirs, Galton (1908) dedicated a mere six pages to his anthropometric work; by comparison, while his research on fingerprints constituted seven pages.

There are many other aspects of Galton’s work that could be described, but the testing methods he developed inspired some two dozen mental testing programs throughout Europe and the United States before his death (Spearman, 1904a). The earliest American efforts with mental testing came through large scale anthropometric studies at Columbia University (Cattell & Farrand, 1896), Clark University (see Bolton, 1892), Yale University (Gilbert, 1894), and the University of Wisconsin, Madison (Jastrow, 1893). Authors of the leading intelligence tests in the twentieth century praised Galton (see Terman, 1932, p. 331; Wechsler, 1939, p. 44), and many of his ideas are still embedded in contemporary assessment such as the concept of a normally distributed intelligence and qualitative description according to deviations from the normative mean.

### **“Mental Energy”: Charles Spearman and Two-Factor Theory**

As the dominant theory of intelligence in the twentieth century, the *general factor* of intelligence has been described as “one of the most replicated results in psychology” (Deary, 2012, p. 146), as saturating “almost all human performance (work competence) dispositions” (Meehl, 1990, p. 124), as sitting at “the heart of the prediction of real-life performances that is possible from tests” (R. L. Thorndike, 1994, p. 150), and as “one of the most central phenomena in all of behavioral science, with broad explanatory powers”<sup>1</sup> (Jensen, 1998, p. xii). The general factor, also known as psychometric *g* or Spearman’s *g*, also comes closest to achieving scientific consensus—Reeve and Charles (2008) surveyed 36 experts in intelligence and found agreement

that  $g$  is an important, non-trivial determinant of important real world outcomes and that there is no substitute for  $g$ , even if performance is determined by more than  $g$  alone.

The groundbreaking paper reporting the discovery of the general factor of intelligence was published in 1904 as one of a pair of papers authored by a 40-year old British graduate student, Charles Spearman (1863-1945), in the *American Journal of Psychology*. Spearman came late to psychology, resigning his military commission at the age of 34 after having served 14 years as an army engineer and earning the rank of Captain in the Royal Munster Fusiliers. Spearman described his years of military service as “almost wasted years” (Spearman, 1930b, p. 304), but there can be little doubt that Spearman’s discipline and authority were shaped by his military background. In 1906, Spearman would complete his doctorate on spatial localization in Wundt’s laboratory at Leipzig, then remaining in Germany to study with Oswald Külpe, G. E. Müller, and Edmund Husserl. Impressed by his 1904 publications, William McDougall arranged for Spearman in 1907 to take charge of Galton’s anthropometric survey of the British population and to become a reader in experimental psychology at University College, London. By 1911, Spearman had been designated the head of psychology and secured the Grote Professorship of Mind and Logic, his title changing to Professor of Psychology in 1928. His major books, *The Nature of ‘Intelligence’ and the Principles of Cognition* (1923) and *The Abilities of Man* (1927), emphasized theory and remain accessible today. Spearman retired in 1931 and began an international collaboration on the Unitary Traits Committee (e.g., Holzinger, 1936), a project that continued until the onset of the second world war. Spearman’s life and career have been most thoroughly investigated by Pat and Sandy Lovie (e.g., Lovie & Lovie, 1996).

### **Theoretical Underpinnings**

Historians credit Spearman with identifying the *general factor of intelligence*, as well as originating factor analysis and establishing the foundations of the classical test theory in psychometrics (e.g., Cudeck & MacCullum, 2007; Williams, Zimmerman, Ross, & Zumbo, 2006). Of these, our focus is on the general factor.

**The general ability factor.** Spearman's most important (1904a) discovery was that "all branches of intellectual activity have in common one fundamental function (or group of functions), whereas the remaining or specific elements of the activity seem in every case to be wholly different from that in all the others" (p. 284). The division of intellectual performance variance into a general ability factor and task-specific factors, respectively denoted as the *g* factor and *s* factors, led Spearman to call his theory *two-factor theory*. In his 1904 paper Spearman was inspired by Galton's (1883a) *Inquiries into Human Faculty and its Development* to determine whether intellectual abilities had any correlation with each other or with sensory discrimination (Spearman, 1930b). Spearman collected academic rankings, ratings, and sensory discrimination performances on students and adults from a village school and a university preparatory school. After establishing global estimates for intellectual status (estimated from average academic standing in various courses) and ascertaining measurement error, he computed correlation coefficients with corrections for attenuation. Spearman observed that when he organized the corrected correlations in rows and columns, placing them in such an order that the corrected correlation coefficients formed a hierarchy, each correlation ended being greater than any to the right of it (in the same row or below it in the same column). Moreover, the ratio between adjacent elements was fairly constant for all pairs of elements in any two columns or any two rows. Spearman interpreted these findings as indicating the presence of some common, central intellectual factor as being involved in all mental activities and characteristically

saturation of academic performance. Essentially, Spearman (1904a; see also Hart & Spearman, 1912) argued that the hierarchy of correlations could only be produced by a general factor without the interference of other intervening factors. Spearman's proof of the general factor lay in the "tetrad difference" formula, in which given four tests each measuring a different mental function, if the difference between the product of any two correlations and the product of the other two correlations approaches zero, the functions tested are considered to tap a single general factor and four specific factors.

In trying to describe the general factor, Spearman used a variety of energetic terms (e.g., "mind power," "mental energy," "intellectual energy"), perhaps hoping that neural underpinnings of *g* would be forthcoming or that the energy metaphor might enhance the scientific accessibility of what was essentially a mathematically-derived construct. Norton (1979) notes that William MacDougall, Spearman's predecessor at the University College, had hypothesized the existence of a form of liquid energy in the brain, perhaps explaining Spearman's conceptualization of *g*. Washburn (1929) was critical of Spearman's extended use of the metaphor, which in various places Spearman used to encompass an individual's quantity of mental energy (i.e., the *g* factor), the engines consuming the energy (i.e., the relative efficiency of various mental processes), and even the engineer pulling the levers (i.e., Spearman's acceptance of Edward Webb's conative factor). Spearman's service as a military engineer also provides a biographical context for the energy metaphor.

**Noëgenetic context.** Spearman's two-factor theory may be understood as occurring in the context of a more complex theory of human cognition. Spearman (1923; see also Carr, Wolf, & Spearman, 1925) conceptualized intelligence and cognition as consisting of three elementary cognitive processes that he considered *noëgenetic* (capable of creating or generating new

cognitions and behaviors): *apprehension of experience*, *eduction of relations*, and *eduction of correlates*. “Apprehension of experience” refers to immediate awareness of one’s own actual experiences, as well as the capacity to consciously observe one’s own thought processes. “Eduction” is reasoning that draws abstract meaning from the chaotic information of the world around us and includes *eduction of relations* (understanding the relation between two or more things) and *eduction of correlates* (understanding the relation between things and relations, or between pairs of relations). Spearman (1930c) asserted that *g* is manifested by every one of the 10 different classes of eduction (space, time, likeness, evidence, identity, conjunction, attribution, causation, constitution, and objectivity) and is by no means confined to obvious instances of reasoning alone; he argued that even tasks like paragraph comprehension, the Ebbinghaus sentence completions, and vocabulary tests all involve educative relations (Spearman, 1930c).

Spearman (1923, 1930a,c) contrasted these so-called creative and productive processes with *anoëgenetic* elementary cognitive processes, the most important of which are *reproductive* mental processes that require retention, recall, and repetition of “old content” (Spearman, 1930a, p. 32) and are dependent upon the same content having “been mentally presented before” (Spearman, 1923, p. 347). He described processes of reproduction as “the very antithesis to creation” (Spearman, 1930a, p. 32). Reproductive processes range from retrieving acquired knowledge to responding to a problem in a well-practiced, previously learned manner. The amount of *g* in educative processes was thought to be far greater than the amount of *g* in reproductive processes (Spearman in Carr, Wolf, & Spearman, 1925, p. 26), but Spearman (1923) acknowledged that there are many compound cognitive operations involving eduction blended with reproduction.

**Group ability factors.** Spearman's (1904a, 1927) original two-factor theory dichotomized variation in test performance as being attributable to either the general factor *g* or task-specific *s* factors, and Spearman only reluctantly acknowledged the existence of *group* factors, a third source of variance explaining abilities common to clusters of performance tasks but not to all of them:

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. (Spearman, 1927, p. 82)

Spearman (1930c) cautioned that to have scientific significance, group factors must possess a sufficiently *broad* range (i.e., avoiding obvious sharing of method variance, such as counting or cancellation tasks) and must avoid a reversion to faculty psychology. In 1927, he acknowledged four “broad group factors” of logical, mechanical, psychological, and arithmetical abilities, each having “sufficient breadth and degree to possess serious practical consequences, educational, industrial, and vocational” (p. 242). He also acknowledged additional general factors independent of *g*, including inertia, perseveration, oscillation, and will (Spearman, 1927, 1930c). But always he returned to the core elements of two-factor theory, and these important alterations to the theory were never consistently reported.

After his retirement from the University of London in 1931, Spearman frequently travelled to the United States to work with former student Karl J. Holzinger on what became known as the Spearman-Holzinger Unitary Trait Study, a series of factor analytic studies for the Unitary Traits Committee of the American Council of Education. Although the ultimate completion of the study was interrupted by World War II and Spearman's passing, a total of nine preliminary reports were released between 1933 and 1936 with some additional analyses released

later. In Holzinger and Swineford (1939), factor analyses were reported from a test battery consisting of 24 tests tapping abilities in five broad areas: spatial, verbal, memory, speed, and mathematical deduction. Some 300 students in seventh and eighth grades completed the battery. Results supported a hierarchical arrangement with one general factor and four group factors, the general factor appearing redundant with the mathematical deduction factor. Holzinger and Swineford (1939) explained, “Indeed, the general factor may be just such a deductive [reasoning] factor as these tests were expected to measure” (p. 8). In an obituary written for Spearman, Holzinger (1946) explained Spearman’s usual dismissal of group factors as being due to the kinds of tests he worked with, but he acknowledged that Spearman was “well aware of the existence” of group factors, also acknowledging that ivory tower conceptualizations had given way to practical considerations: “In the Unitary Traits Committee of 1931-1934 we talked about ‘pure’ factors like chemical elements, but since the atom bomb perhaps this foolish idea may also be exploded” (p. 234).

### **Tests and Measures**

Spearman personally authored at least two intelligence tests: the *Measure of Intelligence* (Spearman, 1925), a highly verbal measure administered orally with written answers, and the *Spearman Visual Perception Test* (Spearman, 1933), a multiple subtest nonverbal battery that had oral directions and later pantomimed directions. One of Spearman’s students, S. A. Hamid (1925), recounted that altogether Spearman created hundreds of tests for research, but he made little attempt to commercialize and promote the tests for applied use.

*A Measure of “Intelligence” for Use in Schools* (Spearman, 1925) was intended for group administration to students, ages 10 through 14 years. It was an entirely verbal measure, administered orally, with written answers. It contained seven (sub)tests: Same or Opposite,

Synonyms, Classification, Questions, Completion, Analogies, and Inferences. It was typically administered in about 75 minutes. In a sample of  $n=761$  students, Hardie (1928) reported a corrected correlation of .72 with English and Arithmetic achievement test performance and a correlation of  $r=.42$  with teacher ratings of academic skills. The Completion test was found to have the highest association with general ability. Walters and Thomas (1929) detailed the test's standardization.

Spearman's (1925) design for the *Measure of "Intelligence"* is surprising because it appears, at least superficially, to be more a measure of acquired knowledge (*reproduction* in Spearman's anoëgenetic terminology) than intellectual reasoning ability (*education*, in Spearman's noëgenetic terminology). In terms of the amount of  $g$  measured by a test, however, Spearman (1927, p. 213) was clear that verbal or pictorial content made little difference, so long as other aspects of test content were carefully considered, specifically the fineness of distinctions to be made across successive acts of education, the novelty of either test content or mental problem-solving processes required, item complexity in terms of attentional requirements, and item abstractness (see also Hamid, 1925, for an excellent discussion). In the introduction to his *Measure of "Intelligence"*, Spearman (1925) emphasized the importance of novelty and abstractness:

Particularly important is the fact that some mental processes need a large supply of the  $g$ , whilst others only require a small supply . . . Certain processes are picked out as needing so much of it, that the person's success depends predominantly upon his supply of this  $g$  and only to a slight extent upon his  $s$ 's. Such processes always consist in novel perceiving of relations between abstract ideas already known in themselves, or in novel proceeding from one such idea and a relation to the correlative idea" (Spearman, 1925, p. 2).



In stark contrast to his *Measure of Intelligence*, Spearman also authored an early nonverbal measure of general ability, the *Spearman Visual Perception Test* (SVPT; Spearman, 1933). The SVPT was a multiple subtest experimental battery that initially had oral directions although pantomimed directions were later developed, along with norms. The SVPT sought to provide the most direct measure possible of the education of relations (see e.g., Zubin, 1933) through its requirements that examinees derive and understand relations between and among geometric figures. The SVPT was also designed to minimize any semantic or pictorial meanings. The SVPT may be considered an immediate antecedent to the Raven's Progressive Matrices which Spearman praised:

Originally . . . this test [the SVPT] was invented by the present writer for the very purpose of measuring 'g'. Subsequently it was successfully employed for the same purpose by many authors, notably Zubin and Lorge. An avowed development of it has been perhaps the best of all non-verbal tests of 'g', namely, the 'Matrices' of Penrose & Raven. (Spearman, 1946, p. 127)

This passage is repeated almost verbatim in Spearman and Wynn-Jones (1950, p. 70), with the significant addition that of some one hundred diverse tests each homogeneous in content, the SVPT (listed as "Spatial Relations" on p. 72) yielded the highest single loading on the general factor, based on findings from the Unitary Traits studies. This finding is the basis for considering these types of tasks nearly "pure" measures of *g*.

Spearman (1939a) described the Progressive Matrices as the "most fully developed of all such tests based on the theory of Two Factors" (p. 252). Spearman's SVPT and the Progressive Matrices were also influences on the development of Raymond B. Cattell's (1940) *Culture-Free*

*Intelligence Test*, which contained three variants of matrix relations tasks depicting both pictorial objects and geometric forms.

### **Ultimate Outcomes**

Previously we have described something of a contemporary consensus on Spearman's general factor of intelligence, so it may be surprising to add that several significant aspects of Spearman's work may be considered to have been falsified. Hearnshaw (1964) and Meredith (1948-1949) both describe Spearman's noëgenetic system as having been inadequately formulated (with large theoretical and methodological gaps), Meredith providing evidence of how it "perished of neglect" (p. 93) and was largely ignored in the academic community. But there were many more attempts to falsify Spearman's theories.

**Mathematical challenges.** Three mathematical-statistical challenges held the potential to falsify two factor theory, had not their mathematical complexity obscured their impact. The first occurred in 1916 when Godfrey Thomson demonstrated that he could generate correlation matrices of imaginary mental tests by rolling dice (a random number generator before the age of computers) that would yield near-zero tetrad differences, Spearman's criterion for a general factor. Thomson (1916, 1919a; Brown & Thompson, 1921) argued that two models--one based on Spearman's test intercorrelations and the other based on his dice-throw mental tests--could give rise to the types of correlation matrices that Spearman specified were required to extract a general factor, and that therefore Spearman's model was *sufficient but not necessary* for the derivation of latent factors. Thomson termed this approach his "sampling theory of ability" or what has more recently been termed a "bonds model" (Bartholomew, Deary, & Lawn, 2009), and his debates with Spearman extended for some three decades through Spearman's death.

The second challenge came from holder of the Galton Chair at the University of London, mathematician Karl Pearson, who attacked Spearman in an unsigned 1927 review of *The Abilities of Man*, critical of Spearman's misrepresentations, errors, and unfounded mathematic assumptions. Pearson (1927) wrote, "what Prof. Spearman considers proofs of his theory are not proofs, and . . . much mathematical work remains to be done before we shall even be in sight of a proof" (p. 183). Pearson and Moul (1927) followed with a detailed (and signed) analysis including identification of errors in the mathematics of the two-factor theory, ultimately rendering "a verdict of non-proven" for the theory (p. 291). The long acrimonious history between Pearson and Spearman reportedly began with Spearman's (1904b) criticisms of Pearson's product moment correlation coefficient and its failure to account for attenuation and irrelevant factors (see Spearman, 1930b, p. 323 for his account of the "embroilment with Karl Pearson"). Lovie and Lovie (2010) provide further information on this well-known, unresolved dispute.

The third serious challenge came in 1928 from Harvard public health statistician Edwin B. Wilson, who identified a problem of factor indeterminacy that appeared to fatally compromise Spearman's derivation of  $g$ . In a review of Spearman's (1927) definitive tome on intelligence, Wilson (1928) wrote, "He does not define it, he computes it" (p. 245). Wilson concluded that Spearman's model could not uniquely determine  $g$ , and between 1928 and 1939, more than a dozen articles by Wilson and Spearman addressed the uniqueness problem (Steiger, 1996). Ultimately, the charge of factor indeterminacy was stonewalled (Steiger, 1996) or just ignored by working psychologists (Thomson, 1947), although Lovie and Lovie (1995) argue that behind-the-scenes correspondence between the two principals and other interested parties resulted in a socially negotiated solution to this debate.

**The positive manifold problem.** A fourth, more philosophical problem, known as the *positive manifold*, always had the potential to damage the concept of a general factor. The positive manifold describes the finding that measures of intelligent behavior always tend to be positively correlated. E. L. Thorndike (1911), a staunch opponent to *g*, described the positive manifold in 1911:

All trustworthy studies so far made of the relations between the amounts of desirable single traits in the same individual agree in finding direct or “positive” relations between such traits. Having a large measure of one good quality increases the probability that one will have more than the average of any other good quality . . . Intellectual ability and moral worth hang together. (Thorndike, 1911, pp. 26-27)

The problem then is the vagueness by which *g* is defined beyond its mathematical derivation; if it is simply a statistical artifact of the positive manifold phenomenon, then it may have little to do with cognitive ability or intelligence *per se* (e.g., Horn & Blankson, 2005). Jensen (1998) argued, however, that the positive manifold is a meaningful, often replicated finding that supports the extraction of the *g* factor:

The striking empirical fact . . . is that, as far as has yet been determined, mental abilities *are* all positively intercorrelated beyond chance to some degree. I have yet to find a bona fide exception. For a century psychologists have made countless attempts to discover even a small number of mental test items that are truly uncorrelated. All have failed . . . The finding of ubiquitous positive correlations between mental abilities is not a psychometric artifact, but an empirical fact, a natural phenomenon. (Jensen, 1998, p. 56; emphasis in original)

**A polarizing figure.** Charles Spearman was a polarizing figure and his bold positions and willingness to engage in academic debates forced him to spend his professional life, as he wrote in his 1930 autobiography, in “one long fight” (Spearman, 1930b, p. 330). In kinder words, perhaps, Louis L. Thurstone (1946) described Spearman’s scholarly exchanges as “a series of lively controversies” (p. 103). These debates may have invigorated the general factor theory by forcing Spearman to reanalyze data, revisit statistical formulae, and rework his arguments. Some of Spearman’s high profile debates (with selected publications in parentheses) occurred with Alfred Binet (e.g., Binet, 1904, 1909/1975; Hart & Spearman, 1912; Spearman, 1915, 1923, 1937), William Brown (e.g., Brown, 1910, 1932; Brown & Thomson, 1921; Spearman, 1910), Godfrey H. Thomson (e.g., Brown & Thomson, 1921; Spearman, 1916; Thomson, 1916, 1919b, 1920, 1939-1940), Edward L. Thorndike (e.g., Hart & Spearman, 1912; Spearman, 1922, 1924; Thorndike, 1909, 1924; Thorndike, Lay, & Dean, 1909), and Thurstone (e.g., Spearman, 1939b; Thurstone, 1939, 1952). Some print debates continued over the span of two decades or more (e.g., Spearman-Pearson, Spearman-Thomson, and Spearman-Thorndike), and Spearman was willing to lay down a challenge but never willing to let his opponent have the last word. Lewis M. Terman, author of the Stanford-Binet, wrote in 1932 about reading Spearman (1904a, 1904b) for the first time: “I shall never forget . . . the impression that those articles made on me—the dogmatic tone of the author, the finality with which he disposed of everyone else, and his one-hundred-per-cent faith in the verdict of his mathematical formulae” (p. 319). In a somewhat self-serving historical account, *Psychology Down the Ages*, Spearman (1937) continued his dispute with Alfred Binet over 25 years after Binet’s death.

**“Meditation upon Facts”:** Alfred Binet and *L'échelle métrique de l' intelligence*

With researchers in the United States and Europe competing to develop a workable way to measure intelligence, the triumph fell to an innovative outsider, a self-taught researcher who worked independently of the academic and government institutions of his time. Even today, Alfred Binet's (1857-1911) work, most of it never translated from the French, remains surprisingly contemporary and ahead of its time (e.g., Nicolas & Levine, 2012; Nicolas & Sanitioso, 2012; Siegler, 1994; Wasserman, 2012). In three editions, 1905, 1908, and 1911, Binet and his *aliéniste* (psychiatrist) colleague Théodore Simon (1872-1961) created a standardized, norm-referenced, clinically- and developmentally-sensitive intelligence test that effectively identified individuals with intellectual disability and dementias (and in the final edition, intellectual giftedness), as well as predicting academic performance. Ultimately spanning the ages 3 through adulthood, the Binet-Simon Intelligence Scale (or "*L'échelle métrique de l'intelligence*" as the authors designated the 1908 revision) was a landmark, paradigm-shifting achievement. One of Binet's most forceful critics, Charles Spearman, acknowledged in 1915:

To appreciate fully the greatness of the work of Mm. Simon and Binet, it is necessary to recall the state of psychology when their epoch-making investigation first saw light. At that time, the "mental tests," initiated by the genius of Galton and taken up with characteristic energy in the United States, had so lamentably disappointed expectations as to have sunk into a general by-word of scorn.

By this one great investigation the whole scene was transformed. The recently despised tests were now introduced into every country with enthusiasm. And everywhere their practical application was brilliantly successful. (Spearman, 1915, p. 312)

The definitive account of the origins of the Binet-Simon Intelligence Scale has recently been recounted by John Carson (2007), and it is clear that Binet's initiatives, including the intelligence

tests and his proposals for educational and military applications, were wholly rejected by the prevailing institutions in his native France. In the United States, however, Binet's work changed the educational landscape and laid the groundwork for a new age in educational assessment, clinical testing, and personnel selection procedures. The authoritative biography of Binet remains Theta H. Wolf's (1973) account, but there is much about Binet's tests, their theoretical underpinnings, and his ideas about assessment that remains untranslated or lost.

Binet began a focused program of intelligence research with Victor Henri in the early 1890s, but the impetus for the educational reform that led to development of the intelligence test began in 1904 when the French Minister of Public Instruction established a commission, headed by Léon Bourgeois, to determine how France's 1882 mandatory public education laws could be applied to abnormal children, including students who were blind, deaf-mute, and backward (Carson, 2007; Wolf, 1969). The public pressure to take action on behalf of intellectually disabled children came from organizations like the educational advocacy group, *La Société Libre pour l'Étude Psychologique de l'Enfant* [Free Society for the Psychological Study of the Child], of which Binet was the president. Seeing an opportunity to objectify and systematize the examination of the intellectually disabled, Binet arranged to become a member of the 21-person Bourgeois Commission, which met numerous times in 1904 and 1905, before issuing its report in 1906. In other words, Binet's Free Society agitated for creation of the commission he planned to join and influence.

The Bourgeois Commission report, largely based on a Binet-drafted subcommittee report, recommended that students who do not benefit from education, teaching, or discipline, should undergo a medical and educational examination under the auspices of a school inspector, doctor, and special educational school director. Exceptional children were to be educated through special

classes 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 did not propose 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). In all likelihood, Binet himself hoped to draw up the scientific guide, and Binet and Simon's 1907 book *Les Enfants anormaux* (with its preface by Léon Bourgeois) was probably intended for this purpose. An English translation was published in 1914. In the end, however, Binet's efforts were completely rebuffed by the French establishment. When the French legislature enacted the law of April 15, 1909 on the education of abnormal students, it highlighted the medical examination but made no mention of any role for psychologists or special methods (i.e., intelligence tests) for assessing students (Carson, 2007). The recommendations of the Bourgeois Commission were largely ignored.

In fact, had it not been for a copy of the 1908 Binet-Simon Scale given to American Henry H. Goddard by Belgian physician and special educator Ovide Decroly, and Goddard's astute recognition and appreciation of the test's remarkable potential with intellectually disabled individuals, the historic contribution of Alfred Binet might have been lost altogether. At present, Binet's works have nearly all been brought back into print with active French scholarship on his contributions, and 2011 saw a belated celebration marking "Le centenaire de la mort d'Alfred Binet."

### **Theoretical Underpinnings**

Binet's theory of intelligence is difficult to pin down, because he demonstrated a consistent and deliberate caution about formulating theories, fearful they might bias his



observations. Binet is even reported to have angrily terminated a professional relationship in which his collaborator, Nicholas Vaschide, made measurement errors consistently in the hypothesized direction (e.g., Wolf, 1961). As a prolific writer and productive thinker, however, he could not help but theorize, and he did offer some theoretical formulations and started but did not finish a grand theoretical synthesis before his death.

**Reticence about theory.** A humiliating early career experience lies behind Binet's wariness about theory (Wolf, 1964). While working with Jean-Martin Charcot in Charcot's neurology clinic at the Salpêtrière Hospital in Paris in the mid- and late-1880s, Binet and colleague Charles Féré experimentally demonstrated the transfer of movements and perceptions from one side of the body to the other in hypnotized hysterical patients, as well as creating polar shifts in emotions (i.e., hate to love, joy to despair) in the same patients, all purportedly through the use of hidden magnets. When challenged in the literature by distinguished Belgian psychophysicist, J. R. L. Delbœuf, and later by Hippolyte Bernheim of the rival Nancy school, Binet responded with forceful defenses of the research, its conclusions, and Charcot's formulations about hypnosis. Delbœuf responded by demonstrating that the transfer and polarization phenomena disappeared when experimental conditions were carefully controlled, proving that suggestibility was the problem (insofar as Binet and Féré's hypnotized subjects were aware of what the experimenters were saying, doing, and expecting; some of the same patients were similarly hypnotized by Charcot). It was a convincing and humiliating falsification of Binet's hypnosis research, and he capitulated, considerably chastened. Binet ended his association with Charcot and Féré, taking an unpaid position at the Laboratory of Physiological Psychology at the Sorbonne, and essentially launching an entirely new line of research. He

would never again accept an authority's theoretical formulation without skepticism, and he was exceedingly cautious about proposing new theories himself:

We have sometimes been accused of being opposed with blind infatuation, to all theory and to the *a priori* method. It is an unjust reproach. We admit the use of theory before the experimental researches, in order to prepare them and afterwards to interpret them; what we strongly reject, are theoretical discussions which are either intended to take the place of an exploration of facts or which are established upon obscure, equivocal and legendary facts . . . In our opinions, the ideal of the scientific method must be a combination of theory and of experimentation. Such a combination is well defined in the following formula: prolonged meditation upon facts gathered at first hand" (Binet & Simon, 1908/1916c, p. 182; see also p. 253).

As a point of comparison, Sigmund Freud, who studied with Charcot at the Salpêtrière for a few months during the same time period, also defended Charcot's views and never fully accepted the Delbœuf-Bernheim falsification that deeply affected Binet (see Macmillan, 1991, for a full account).

**Binet reviews other theories.** Having painfully lost his scientific naïveté, Binet co-founded the journal *L'Année Psychologique* as a way to publish his research and review the work of other psychologists. It is informative not only to read Binet's original articles but also to see the work that he chose to review for insight into his thinking and interests at any given time.

In the paper that may have been the first to use the term "individual psychology" to refer to the study of individual differences (as contrasted with universal mental processes), Binet and Victor Henri (1895) directly challenge the importance given to measurement of simpler mental processes by Galton and Cattell:

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; emphasis in original)

That complex, higher order tasks do a better job of differentiating younger (and therefore lower ability) individuals from older (and therefore higher ability) individuals was critical in ultimately determining the content of the Binet-Simon Scale, but Binet also relied heavily on tests that differentiated special populations of interest, working with samples diagnosed as mentally disordered (neurotic and psychotic), intellectually disabled, language disordered, intellectually gifted, and even monosavants. The Binet-Simon Scale effectively identified intellectually disabled examinees because Binet selected tasks and test items for that very purpose. In practical terms, higher order complex processes provided better separation between individuals and between groups of interest than did simpler sensory and motor tests:

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 this article, Binet and Henri (1895) provide their first detailed descriptions of higher order tasks of attention, memory, reasoning, mental imagery, imagination, comprehension, suggestibility, aesthetic sentiment, moral sentiments, muscular force and strength of will, and visual-motor skill.

Binet's position on Charles Spearman's general factor of intelligence must be described as ambivalent but generally indicating acceptance of *g*. In a review in *L'Année Psychologique*, Binet (1904) initially disparaged Spearman's (1904b) finding that the corrected correlation between sensory discrimination and general intelligence is near-perfect:

[Spearman] regards this conclusion as profoundly important. It may possibly be. We ourselves are profoundly astonished at the conclusion because of the very defective character both of the sensory experiments of the author and of the manner in which he determined or had others determine general intelligence. (translated in Peterson, 1926, pp. 271-272; original in Binet, 1904, p. 624)

Binet remained critical of Spearman's concept of general intelligence, especially its mathematical-statistical derivation, through the end of his life (see e.g., Binet, 1909a). Although he softened his criticism over time and developed a test much better at estimating overall mental level than any specific aptitude, he never fully endorsed (or rejected) Spearman's *g*:

Two absolutely contradictory views are prevalent today and both claim to be upheld by proof. According to one view, which has been vigorously supported by Thorndike, the mind is nothing but an absolutely heteroclite collection of faculties which are as if juxtaposed but which remain rigorously independent from one another. The opposite

view, supported by Spearman [sic] with impressive mathematical data, states that intelligence is one, that in each of us there exists a faculty deserving the name of general intelligence and that this faculty accounts for the relationship which is revealed by the measurement of our performance in all our activities, even the most disparate ones . . . On the whole, these two views are extreme, and there are more moderate ones which such controversies do not invalidate. (Binet, 1909/1975, pp. 169-170)

Binet here is contrasting Spearman's (1904a) general ability factor with E. L. Thorndike's (1901) hypothesis that the number of neural bonds or connections is the foundation for intellect, because neural bonds correspond to learned stimulus-response associations.

**Conceptualizations of intelligence.** Binet's evolving conceptualizations of intelligence were comprehensively reviewed by Varon (1936) and Wolf (1969) somewhat before contemporary intelligence theories began emphasizing the hierarchical structure of cognitive abilities (e.g., Carroll, 1993). It is possible to make the case that Binet anticipated a hierarchical structure for intelligence and that he saw intelligence as integrating a fundamental superordinate factor (*judgment*), four essential processes (*direction, comprehension, invention, and criticism*), and a host of narrower mental faculties (e.g., memory, imagery, attention, judgment, and suggestibility), while emphasizing the dynamic interplay between individual aptitudes, emotions, physical health, and environment. As early as 1905, Binet had concluded that intelligence is hierarchically structured and organized:

Intellectual qualities are not superposable, and therefore cannot be measured as linear surfaces are measured, but are on the contrary, a classification, a hierarchy across diverse intelligences. (Binet & Simon, 1905/1916c, p. 40)

By 1909, decades before hierarchical factor structure was developed, Binet specified a higher order role for basic cognitive processes and a lower order role for the mental faculties:

In short then, we may conclude that the theory of the intellectual faculties and the theory of the scheme of thought belong to two different planes.

To make this distinction more clear, we borrow from biology the following comparison; the primordial biological element is the cell; in grouping themselves, cells form the tissues; tissues in their turn form the organs. In the same way one might say that the intellectual functions of memory, attention, judgment, etc., correspond to the cells; combining themselves, they form something analogous to a tissue. What corresponds to the organ is our scheme of thought, because, like the organ, this scheme has a function.

(Binet & Simon, 1909/1916e, pp. 152-153)

In spite of his ambivalence about the general factor of intelligence, Binet came to consider intelligence as *unitary*, alluding in *Modern Ideas about Children* (Binet, 1909/1975) to a superordinate general factor:

Dr. Simons [sic] and I have recently proposed a synthetic theory of the functioning of the mind, which it will surely be useful to summarize here. This theory shows clearly that the mind is a unit in spite of the multiplicity of its faculties and that it has an essential function to which all others are subordinated. (Binet, 1909/1975, pp. 90-91)

This essential function reflected *judgment* (Binet & Simon, 1905/1916c) and manifested itself by the “best possible adaptation of the individual to his environment” (Binet & Simon, 1911/1916b, pp. 300-301):

It seems to us that in intelligence there is a fundamental faculty, the alteration or lack of which, is of the utmost importance for practical life. This faculty is judgment, otherwise

called good sense, practical sense, initiative, the faculty of adapting one's self to circumstances. To judge well, to comprehend well, to reason well, these are the essential activities of intelligence. (Binet & Simon, 1905/1916c, pp. 42-43)

In a remarkable but forgotten theoretical formulation first appearing in *L'étude expérimentale de l'intelligence* in 1903, Binet specified four core mental processes or "primordial functions" (Binet, 1909/1975, p. 91), that he termed a "scheme of thought" in *The Intelligence of the Feeble-Minded* (Binet & Simon, 1909/1916e, p. 130), elaborating further in *Modern Ideas about Children* (Binet, 1909/1975). In brief, Binet posited an active cognitive organization of thought processes in relation to external events. Thought is composed of four distinct elements: a *direction*, a *comprehension*, an *invention*, and a *criticism*. "This is the essence of intelligence," he wrote (Binet, 1909/1975, p. 91; see also Binet, 1909, p. 118). He described each of these processes as follows:

- **Direction:** an "organizing force" (Binet, 1903, p. 108) that addresses the question of "what it is about" (Binet & Simon, 1909/1916e, p. 136) for one or more intellectual acts; strongly related to attention, focus, and persistence in the face of distraction, establishing the path to be taken and intent of a task or action; "All we have to do is draw a line and our schema is complete. The work described cannot be done haphazardly, without knowing what the matter is all about, without adapting a certain position from which we do not deviate. We must therefore take and maintain a definite direction" (Binet, 1909/1975, p. 91).
- **Comprehension:** perception and understanding of experiences and directing ideas (Binet & Simon, 1909/1916e); refers to the perception of things and events, across sensory modalities, and understanding of their meaning and significance, including the ability to take an in-depth, global view (distinguishing between the incidental and the essential, and the

capacity to see beyond the superficial); requires abstract understanding that transcends sensory perception; the capacity for observation and interpretation (Binet & Simon, 1909/1975, p. 92).

- ***Invention***: generation of actions, answers, choices, ideas, and responses to problems based on knowledge, reason, and imagination, including the sorting of the appropriate response options from available ideas, responding with an appropriate level of precision or complexity, and performing the optimal mental operations (e.g., addition, continuation, or synthesis) to produce a response (Binet, 1909/1975).
- ***Criticism***: a self-monitoring, evaluative “apparatus of control . . . which has for its object its own operations” (Binet & Simon, 1909/1916e, p. 148); Binet also uses words such as “the correction,” “auto-criticism,” “critical sense,” “judgment,” and “auto-censure” to describe this metacognitive process, which operates in two directions: “one is turned toward the exterior world, which we perceive and evaluate; the other, the fact of reflection, is turned back upon ourselves, and it is ourselves that we evaluate” (Binet & Simon, 1909/1916e, p. 148); the mental process of criticism continually evaluates progress towards any given objective: “the work cannot be completed successfully if the ideas which are generated in the process are not evaluated as they occur and rejected if they are found to be unsuitable to the goal that is being pursued” (Binet, 1909/1975, p. 91).

Binet’s inclusion of what is essentially an attention/intention factor (*direction*), a higher order perceptual processing factor (*comprehension*), a response generation factor (*invention*), and a metacognitive evaluative factor (*criticism*) is astonishing for his time, anticipating some of the innovative cognitive processing ideas appearing many decades later.



Finally, Binet's specification of the intellectual faculties that constitute the lower plane, narrower abilities in his hierarchy (Binet & Simon, 1909/1916e, pp. 152-153) came from the original plans he and Henri made for the intelligence test (Binet & Henri, 1895). The faculties interact dynamically with each other and the higher order cognitive processes of *direction*, *comprehension*, *invention*, and *criticism* (in this earlier version Binet uses *direction*, *correction*, and *adjustment*) in the service of the superordinate factor, *judgment* (here called *adaptation*):

The essential point of the new theory is considering thought as an action, the action consisting in adapting itself; it is around this conception that everything gravitates; furthermore, the principle of adaptation is not contained in any one of our intellectual faculties; there is in it an idea which surpasses them. On the other hand, if the principal parts of the system, direction, correction, adjustment, can be explained by a play of the attention, memory, imagination, judgment, it must be noted that any one of those faculties taken alone would be ineffectual for the work that one would wish to assign it. (Binet & Simon, 1909/1916e, pp. 151-152)

At the time of his death, Binet was developing tests of isolated aptitudes (e.g., memory, mathematics, attention) that seem identical with some of the faculties in his model (Binet, 1911).

### **Tests and Measures**

Alfred Binet completed working measures of intelligence (with Théodore Simon) and academic skills (with Victor Vaney). He had previously researched or piloted methods to measure will and personality, although he considered these to require further research. Binet's measures related to health and medical status included questionnaires and anthropometric measures of physique, motor skills, and sensory acuity (auditory and optometric tests). And as

we have mentioned, he was developing a series of aptitude tests when he passed away (Binet, 1911). Here we discuss his most important contribution, the Binet-Simon Intelligence Scale.

**1905 scale.** The initial 1905 edition of the Binet-Simon Scale consisted of 30 items sequenced from easy to hard and was the first working intelligence test. It was developmentally sequenced, with norms drawn from about 50 children between the ages 3 and 11 years of age, but the tasks selected also discriminated between normal and intellectually disabled individuals making it helpful in classification. Above all it was *practical*:

We have aimed to make all our tests simple, rapid, convenient, precise, heterogeneous, holding the subject in continued contact with the experimenter, and bearing principally upon the faculty of judgment. Rapidity is necessary for this sort of examination. It is impossible to prolong it beyond twenty minutes without fatiguing the subject. (Binet & Simon, 1905/1916c, p. 41)

The first item (“Le Regard”) was the simplest and required the subject to look at and visually track a lighted match; the hardest and most complex item required the subject to answer an abstract question based on vocabulary and an understanding of word relations (e.g., “What difference is there between ‘weariness’ and ‘sadness’?”). Intellectually disabled individuals considered *idiots* generally could not advance beyond the sixth of the 30 tasks; individuals considered *imbeciles* rarely went beyond the fifteenth task (Binet & Simon, 1905/1916c).

With the introduction of the 1905 edition, Binet and Simon emphasized standardized administration while eliciting optimal performance—“make him [the examinee] do his best without giving aid” (p. 45). They also describe the importance of accurately recording results while “making psychological observations . . . [of] such little facts as permit one to give the gross results their true value” (p. 45). Through three editions, Binet and Simon made skilled

observations of test behaviors, ranging from evaluating the quality of examinee verbalizations to distinguishing between classes of errors made during problem-solving to tracking nonverbal problem-solving strategies (such as the spatial path taken in a paper-and-pencil reproduction of a design). Administration of the Binet-Simon Scale was intended as a structured clinical interaction that went far beyond the actual test scores; results were ideally to be integrated with direct observations, information from questionnaires and background history, and educational and health status reports, with the objective of arriving at a holistic understanding of the examinee.

**1908 scale.** The 1908 Binet-Simon Intelligence Scale became an *age-graded scale of intelligence* [*L'échelle métrique de l' intelligence*] featuring 56 tests arranged by difficulty so that tests were placed at levels corresponding to about a 75% pass rate for children of any given age. The test was administered in a spiral omnibus format, so that all items of different types were administered at a given level before items on the next level were given. Normative expectations were generated by testing about 200 children between the ages of 3 and 15 years. The 1908 scale permitted a student's *mental level* [*niveau mental*] to be estimated, based on the highest age at which a child passed four or five tests (the basal year) with an additional year added for each of the five tests passed beyond the basal.

**1911 scale.** In the 1911 edition, Binet's scales were extended from age 3 years through adulthood, with 11 levels and five items administered at each level (Binet & Simon, 1911/1916b). The extension of the Binet-Simon Scale into adulthood enabled its use with the intellectually disabled *and* the gifted, because it could now measure mental levels several years older than student's chronological age (Binet, 1911, p. 10):

The practical applications of this study [of intelligence] are evident in recruitment for classes of the abnormal, in the formation of classes for the supernormal, in the determination of the degree of responsibility of certain feeble-minded [*débiles*], etc., without even taking account of the great interest that a parent or a schoolmaster could find in knowing if a child is intelligent or not, if his scholastic performance [*succès*] is related to his idleness or intellectual incapacity, and towards what kind of career it is fitting to direct him. (Binet, 1911, p. 10)

For school children, Binet identified an intellectual disability for a mental level delay of “two years when the child is under [age] nine, and three years when he is past his ninth birthday” (Binet & Simon, 1907/1914, p. 42). The Binet-Simon scale never yielded an intelligence quotient (IQ), and Binet and Simon considered the use of a single summary score to be a betrayal of the scale’s intent (see Wolf, 1973, p. 203). Ahead of the American army mental tests, Binet and Simon (1910) also argued, albeit unsuccessfully, for the value of intelligence testing in French military conscription.

### **Ultimate Outcomes**

“His measuring instrument virtually perished with him in France,” wrote Carson (2007, p. 149), concluding that the Binet-Simon scale was almost wholly ignored by French educators, scholars, and the government after Binet’s death. Even his laboratory at the Sorbonne came under the supervision of Henri Piéron who was deeply at odds with Binet, discontinuing all research on his test and omitting any mention of Binet in test compendiums published in 1904 and 1911 (Toulouse & Piéron, 1911; Toulouse, Vaschide, & Piéron, 1904; see also discussion in Lautrey & de Ribaupierre, 2004).

The credit for bringing the Binet-Simon Scale to the United States belongs to Henry H. Goddard, Director of Research at the Training School for the Retarded at Vineland, New Jersey, who arranged for a translation by Elizabeth S. Kite and national distribution from the Vineland School. Goddard became a leading advocate for the test, distributing 22,000 copies and 88,000 response sheets by 1915 (Goddard, 1916). Goddard's biographer, Leila Zenderland, notes that Goddard's importance to intelligence testing was soon surpassed by Lewis M. Terman at Stanford University, author of the Stanford adaptation of the Binet-Simon Scale, but she concludes that Goddard's early contributions were critical:

In less than ten years, Goddard had won legitimacy for intelligence testing in ways only dreamed of by Binet himself. By 1910, he had convinced American physicians to try intelligence testing. By 1911, he had used the same tests in public schools. By 1913, he had tried out his tests at Ellis Island. By 1914, he had become the first psychologist to present evidence from Binet tests in a court of law. By 1918, he had even helped to introduce intelligence testing into the United States Army. (Zenderland, 1998, p. 2)

The success of the Binet-Simon Intelligence Scale was profound and paradigm shifting by almost any standards. An editor of the *Journal of Educational Psychology* asked only somewhat rhetorically, "What is the reason for the remarkable influence which the Binet-Simon scale for the measurement of intelligence has exerted on the educational thinking of the civilized world during the past five years?" (Bell, 1917, p. 45). By 1917, over 700 studies, reviews, or books covering the scale had been published (Kohs, 1914a,b,c; 1917a,b,c,d) and the scale had been adapted or translated for use in Canada, England, Australia, New Zealand, South Africa, Germany, Switzerland, Italy, Russia, China, Japan, and Turkey (Goddard, 1916).

In her *A Bibliography of Mental Tests and Rating Scales*, Gertrude H. Hildreth (1933) begins with a section called “Binet Tests and Revisions” (p. 25), in which she lists an astonishing 74 Binet-Simon Scale tests and adaptations, including translations, that had been developed through 1932. The list included downward and upward age extensions, several point scale adaptations, several abbreviated versions, several special editions for disabled populations (Deaf, Mute, Blind), and more than two dozen translations. But the English-language adaptations, translations, and revisions were authored by an unusually distinguished group, here partially listed in alphabetical order: Cyril Burt, Henry H. Goddard, John P. Herring, Edmund B. Huey, Fred Kuhlmann, Norbert J. Melville, Stanley D. Porteus, Lewis M. Terman, J. E. Wallace Wallin, Guy M. Whipple, and Robert M. Yerkes. Of these, the Stanford revision and extension, as authored by Terman (1916), quickly established itself as the best of the many available alternatives. It was the quality of his Binet-Simon adaptation and the extraordinary ways in which he applied test results that distinguished the *Stanford-Binet Intelligence Scale* and its author, Lewis M. Terman (1877-1956), from numerous competitors in mental testing. The theoretical underpinnings of the Stanford-Binet, however, were almost entirely derived from Binet’s original ideas.

In summarizing Alfred Binet’s contributions to mental ability testing, it may be said that Binet built on Francis Galton’s work by developing an objective, norm-referenced, and systematic assessment approach that shifted the focus from lower order anthropometric characteristics to higher order mental processes. While Galton recognized the value of measuring individual differences *per se*, Binet identified the types of individual differences that had the greatest capacity to change over the life span and to discriminate between (and within) individuals in both normal and special populations. While Galton conceptualized the

measurement of individual differences in mental capacity by “sinking shafts . . . at a few critical points” (Galton, 1890c, p. 380), it fell to Alfred Binet to recognize that a diverse sampling of mental processes (what Spearman, 1930b, p. 325, termed a “hotchpotch”; see also Spearman, 1930c, p. 344) yielded the most powerful prediction of behavior. And although Binet would not formally endorse the Spearman’s concept of general intelligence, his test ended up being its greatest champion.

The history of Alfred Binet’s brilliant work on the measurement of intelligence constitutes a significant challenge to history of science accounts endeavoring to explain how science advances in terms of theory and practice. After Binet’s death, his test effectively disappeared in France but was transformative for the field of psychology in the United States. Binet’s work was never falsified from a scientific perspective, but it was treated in France as if it had been thoroughly falsified through its wholesale rejection by the French academic, educational, medical, military, and government institutions. Accordingly, the fate of Binet’s body of work in France provides an informative illustration of how nonscientific social factors can contribute to the demise of an approach.

### **“Pragmatic Mental Testing”: David Wechsler and the Wechsler Intelligence Scales**

At the time this chapter is being written, the Wechsler intelligence scales and memory scales still rank as the most frequently used tests of their respective types by school and clinical psychologists, having held this top-ranked position for a half-century or more. The success of the Wechsler scales probably stems from a number of factors, the most substantial of which is their derivation from leading applied diagnostic measures by a gifted clinician, David Wechsler (1896-1981), who synthesized and streamlined the diverse tests into practical batteries that enhanced clinical and educational decision-making.

Wechsler's career and experiences coincide with twentieth century growth of the field of psychology, and his personal and professional story includes participation in many milestone events in psychological science including its expansion from an experimental science to an applied discipline; its rise in public prominence after the army mental tests of World War I; the gradual professionalization of its practitioners; and its commercialization in clinical, educational, and industrial/organizational practice. As one of the first practicing psychologists, Wechsler's orientation was more applied than academic and more practical than theoretical. In the words of Lee J. Cronbach (1958), "His scale represents the highest flowering of the pragmatic mental testing initiated early in this century, rather than a break into any new understanding of intellectual processes" (p. 1133).

Elsewhere, we have provided accounts of David Wechsler's personal and professional history (Wasserman, 2011; Wasserman & Kaufman, in press). He was an improbable success story for any number of reasons: his arrival as a child immigrant in New York at the age of 6 in the middle of a huge influx of impoverished Eastern European Jews, his family almost entirely working in the garment industry in New York City's Lower East Side, both parents perishing from cancer by the time he was 10 years old, his education at a time when university-based anti-Semitism in academics was high, and his self-identified status as a conscientious objector (then a risky decision) in World War I. His older brother, Israel S. Wechsler, would raise and mentor him after the death of their parents. Israel Wechsler was a distinguished neurologist who served as president of the American Neurological Association and would become a professional role model, as a successful academician and practicing clinician, for David. David Wechsler would serve as psychologist (and later chief psychologist) at New York's Bellevue Psychiatric Hospital from 1932 to 1967, as a clinical professor at the New York University College of Medicine from



1942 to 1970, and as author of *The Range of Human Capacities* (two editions, in 1935 and 1952) and *The Measurement and Appraisal of Adult Intelligence* (through five editions, beginning in 1939).

Through planning or good fortune, David Wechsler's graduate education and training were with some of the most influential psychologists in the early twentieth century. He earned his master's degree in 1917 and his doctoral degree in 1925 from Columbia University, in a program that Robert S. Woodworth (1942) described as teaching "experimental abnormal psychology" (p. 11), an antecedent to clinical psychology. He counted Woodworth, J. McKeen Cattell, and Edward L. Thorndike among his primary influences. During his World War I military service as an army mental examiner, he was trained in both the group tests Alpha and Beta and individually-administered tests; after completing training, he was stationed at Camp Logan, Texas, where he primarily administered individual intelligence tests to recruits who had failed the group tests. Fluent in German, Wechsler volunteered as an outpost listener near the end of World War I and was shipped to Europe but did not serve due to the Armistice. He then participated in an educational program for soldiers that permitted him to study for four months in 1919 at the University College, London with Charles Spearman and Karl Pearson, learning about Spearman's general factor and Pearson's correlation coefficient. Following his military service, Woodworth recommended him for an American Field Service (AFS) fellowship in France, where Wechsler conducted research from 1919 to 1921 at the University of Paris under the supervision of Henri Piéron and Louis Lapique (Wechsler, 1925, p. 8; Rock, 1956, p. 675). During this time just a few years after Binet's death, he reportedly met Pierre Janet and Théodore Simon (Matarazzo, 1972).

Upon his return to the United States, Wechsler continued to seek clinical experience in Boston (learning from Frederic L. Wells, William Healy, and Augusta F. Bronner) and New York (at the New York Bureau of Children's Guidance). Following a stint of several years in private practice, two years as the acting secretary of The Psychological Corporation, and three months learning psychoanalysis with Anna Freud in Vienna, Wechsler was hired in 1932 as a psychologist by the Psychiatric Division of Bellevue Hospital, New York, one of the oldest public hospitals in the United States. By 1941, Wechsler had become chief psychologist and a clinical faculty member at the N.Y.U. College of Medicine, supervising more than 15 clinical psychologists, five interns, and two research psychologists on grants. Wechsler retired from Bellevue 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.

### **Theoretical Underpinnings**

“First and foremost,” Dr. Wechsler often reminded one of us (A.K.), “IQ tests are clinical tests—not psychometric tests, but clinical tests.” It was not that Wechsler was unaware of the importance of theory, factors, and psychometric quality, but instead that he placed a premium on the clinical utility of his tests. With the ongoing stream of patients needing assessment at Bellevue, where he was chief psychologist, tests were specifically used to inform and enhance clinical decision-making. Moreover, Wechsler used his intelligence tests to make inferences about the functioning of the *whole* person, not just his or her cognitive ability. Upon learning that a substantial amount of intelligence test performance variance remained statistically unexplained by general, group, and specific factors, Wechsler became convinced that personality and other non-intellective factors accounted for the remaining performance variance (e.g., Wechsler, 1939,

1950), and his attempts to measure these other factors are one of the great untold stories of his life.

Wechsler's definition of intelligence evolved over time, becoming increasingly relativistic by the end of his life. Wechsler considered most authorities on intelligence to be essentially correct in their definitions, but he noted that intelligence is multifaceted and that many definitions are offered with respect to a narrow research orientation or specific setting. For example, when intelligence is discussed relative to school settings the matter of *learning* tends to be especially relevant, because learning is a primary activity in school.

Like many psychologists before him, Wechsler accepted adaptation as a fundamental aspect of intelligence. Herbert Spencer (1855) is generally credited with initially emphasizing adaptation, defining intelligence as a "continuous adjustment" of "inner to outer relations" (Spencer, 1855, p. 486). In a 1926 article written for the *New York Times Current History* magazine, Wechsler wrote,

In general, by intelligence the psychologist means the ability to adapt one's self to a new situation and handle it successfully. In children this ability can best be estimated by how quickly or how well they learn; in adults, by how well they profit from past experience and how successfully they can apply their knowledge to new problems." (p. 360)

He reiterated this definition in his earliest scholarly statement on intelligence in 1927 in his brother's neurology book: "... 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). Just two years later, Wechsler (1929) offered a variation on this definition: "In brief, it [intelligence] is the ability first to perceive a new situation and then to handle it successfully" (p. 291).

It is Wechsler's 1939 definition, however, that remains his most-cited definition 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).

According to an interview with Wechsler (Wechsler, Doppelt, & Lennon 1975), this definition was intended to reflect Spearman's general factor ("global capacity"), Thurstone's group factors ("elements or abilities"), Terman's capacity for "abstract thinking" ("to think rationally"), and Binet's emphasis on adaptation ("to deal effectively with one's environment"). Binet (1910) also took the position that intelligence is a dynamic synthesis, more than the different "pieces of the machine" that comprise it, possibly influencing Wechsler's statement that intelligence is more than the "mere sum" of its constituent abilities.

In an APA address published in 1975, Wechsler's conceptualization changed again, now to seeing intelligence as a multifaceted concept that can only be considered *relative to context*: "Intelligence cannot be equated with cognitive or intellectual ability . . . To be rated intelligent, behavior must not only be rational and purposeful; it must not only have meaning but it must also have value, it must be esteemed" (p. 136). Put otherwise, intelligence is determined in part by a person's environment and the qualities valued there. Intelligence can only be evaluated with respect to how appropriate, effective, and worthwhile its corresponding behaviors are:

What we measure with tests is not what tests measure—not information, not spatial perception, not reasoning ability. These are only means to an end. What intelligence tests measure, what we hope they measure, is something much more important: the capacity of an individual to understand the world about him and his resourcefulness to cope with its challenges. (Wechsler, 1975, p. 139)

Wechsler's conviction that intelligence is best understood in sociocultural context dates back to 1935, when he wrote that the concept of genius requires superiority of ability, rarity of incidence, and pursuit of an activity "that is socially esteemed and considered worthwhile by those who are qualified to judge" (Wechsler, 1935b, p. 109). By this last phrase, he meant that the exceptional abilities that characterize genius may be of value in one society but not another and therefore should not be evaluated independent of the environment. He reiterated the importance of sociocultural context in a paper on mental deficiency, in which he recalled classification errors made by the army tests:

The first case of this type that came to my attention was a native, white Oklahoman of 28, who had come up for individual psychological examinations because he had failed to pass the Army Alpha and Army Beta intelligence tests. On both Stanford-Binet and the Yerkes Point Scale he obtained a mental age of less than 8 years. Nevertheless, before entering the Army he had gotten along very well, was supporting a family, had been working as a skilled oil-driller for several years and, at time of draft, was earning from \$60 to \$75 per week. Incidentally, he was making the grade as a soldier, and would not have come to the attention of the authorities had he not failed on the psychological tests. (Wechsler, 1935a, p. 234)

Wechsler (1939) drew the conclusion that blind interpretation of test scores, without considering environmental adjustment and background history, is not best practice:

The kind of life one lives is itself a pretty good test of a person's intelligence. When a life history (assuming it to be accurate) is in disagreement with the "psychometric", it is well to pause before attempting a classification on the basis of tests alone. Generally it will be found that the former is a more reliable criterion of the individual's intelligence. (p. 48)

He cautioned against group testing which yields scores interpreted according to "*apersonal* psychometrics" (Wechsler, 1939, p. 48; emphasis in original).

The Wechsler intelligence scales may be considered theoretically ecumenical, embracing multiple perspectives but none too tightly. Wechsler emphasized Spearman's general factor *g* (e.g., Kaufman, 2009) but included subtests in his intelligence scales with poor *g* saturation (e.g., Digit Span). He explained in 1939 that a scale built entirely of tests rich in *g* would be a mistake: "Such a scale would not be a very good measure of general intelligence because it would eliminate a number of abilities essential for effective behavior" (p. 8). Wechsler wrote early and reverently of Spearman's general intelligence factor, but he resisted adopting Spearman's dogmatism:

We . . . can only indicate our own position by saying that Professor Spearman's generalized proof of the two-factor theory of human abilities constitutes one of the great discoveries of psychology. (Wechsler, 1939, p. 6)

The present writer is far from being in full agreement either with Professor Spearman's concept of general intelligence or even his views regarding the best mode of measuring it, but as regards the demonstration of the existence of "g" as a common factor, there seems

to be no possibility of doubt. Psychometrics, without it, loses its basic prop. (Wechsler, 1939, pp. 7-8)

Wechsler also acknowledged a variety of group factors, the most enduring of which were those derived from factor analyses of his own intelligence scales: verbal ability, nonverbal / performance ability, and memory (Wechsler, 1958). Other factors considered and discussed by Wechsler included Cattell's (1943) fluid and crystallized abilities (discussed in Wechsler, 1971); Thorndike's (1920) abstract, social, and practical intelligences (discussed in Wechsler, 1939, 1950); and Thurstone's (1938) primary mental abilities (discussed in Wechsler, 1950). Wechsler did not, however, dedicate much attention to differentiating these factors, commenting,

While intellectual abilities can be shown to contain several independent factors, intelligence cannot be so broken up. Hence, no amount of refinement of tests or addition of factors will account for the total variance of an intelligence test battery, because the variance in intelligence test performance is due not only to the direct contributions of the factors themselves but also to their collective behavior or integration" (Wechsler, 1958, p. 23).

Influenced by A. W. Stern (1956), he felt that the interaction and connectivity of primary cognitive abilities was of far more value in predicting behavior than the isolated factors.

### **Tests and Measures**

Wechsler's intelligence scales included the *Wechsler-Bellevue* (1939), the *Wechsler Intelligence Scale for Children* (WISC; 1949), the *Wechsler Adult Intelligence Scale* (WAIS; 1955), and the *Wechsler Preschool and Primary Scale of Intelligence* (WPPSI; 1967). The *Wechsler Memory Scale* (WMS), also an industry leader, was first published in 1945. Wechsler participated in several revisions, and after his death, the test publisher continued to issue

revisions with external expert advisors and internal project directors guiding test development. Wechsler never proposed or wrote about achievement tests or nonverbal tests like those that currently carry his name.

The majority of measurement procedures included in Wechsler's intelligence scales he had learned in the army (see e.g., Yerkes, 1921). The procedures and in some cases the test stimuli he included in the original Wechsler Memory Scales originated from the Army Performance Tests, the Binet-Simon and Stanford-Binet Intelligence Scales, and Wells' (1927) *Mental Tests in Clinical Practice*. Wechsler's criteria for selecting procedures for his intelligence tests included mental functions measured, psychometric quality, evidence for validity, and clinical value. 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). Boake (2002) has chronicled the origins of the individual Wechsler subtests from existing tests, a fact which Wechsler (1939) openly acknowledged even while he streamlined them into 5-10 minute point scale subtests and made scoring rules relatively uniform. In some ways, the Wechsler scales may be said to work well because their original antecedent measures were already known to work well.

Wechsler (1939) may be credited with popularizing the deviation IQ as part of the Wechsler-Bellevue, which provided scores expressed as deviation from the normative mean (with the normative mean set at 100 and standard deviation set at 15). With the assumption of a normal probability distribution, the deviation IQ permitted test scores to correspond to uniform percentile ranks. Wechsler was certainly not the first psychologist to express scores in terms of standard deviations from the mean; Robert S. Woodworth (1912) provided formulae to transform raw scores to standard scores and standard deviations from the normative mean, thus facilitating



the comparison of individual performance across different tests to the group average and individual deviation to group variability. Wechsler was also not the first to describe or implement a deviation IQ (see e.g., Burt, 1914; Otis, 1917; and Pintner and Marshall, 1922), but he popularized its use and offered a test with sufficient appeal to cause practitioners to make the shift from the ratio IQ. Ironically, just two years before Wechsler (1939) published the Wechsler-Bellevue, the authors of the Stanford-Binet LM declined to adopt a deviation IQ on the grounds that it was too statistical for the professional community (Terman & Merrill, 1937, pp. 27-28).

Wechsler's subtests were divided into verbal and performance/nonverbal content for *practical* reasons, just as the army group-administered mental tests had included the *Alpha* and the *Beta*, and the convenience of having both types of contents balanced and normed together in a single test was one of the reasons for the Wechsler-Bellevue's popularity (e.g., Rabin, 1945; Rabin & Guertin, 1951). The verbal-performance dichotomy 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 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, Doppelt, & Lennon, 1975, p. 55):

To the extent that tests are particular modes of communication, they may be regarded as different *languages*. These languages may be easier or harder for different subjects, but it cannot be assumed that one language is necessarily more valid than another. Intelligence

can manifest itself in many forms, and an intelligence scale, to be effective as well as fair, must utilize as many different languages (tests) as possible. (Wechsler, 1974, p. 5; emphasis in original)

Wechsler was consistent in not “singling out any ability, however esteemed (e.g., abstract reasoning), as crucial or overwhelmingly important” (Wechsler, 1974, p. 5). He cautioned against using only a single testing modality, noting that for nonverbal and performance tests, “they correlate poorly with verbal aptitudes and are poor prognosticators of over-all learning ability as well as school achievement. Above all, they have turned out to be neither culture-free nor culture fair” (Wechsler, 1966, p. 63).

Wechsler introduced many psychometrically-based procedures to interpretation of his tests, including assignment of descriptive labels based on the Full Scale IQ (based on the probable errors from the normative mean), specification of the commonality and statistical significance of specific VIQ-PIQ differences, and identification of intraindividual subtest patterns. In 1932, Wechsler had already described an approach to profile analysis of Army Alpha subtests, an antecedent to interpreting relative strengths and weaknesses on individual Wechsler scale subtests. Following F. L. Wells’ introduction of the Memory Quotient (MQ) concept, Wechsler (1945) recommended that memory should be compared to measured intelligence (IQ-MQ) “because it makes possible comparison of the subject’s memory impairment with his loss in other intellectual functions” (p. 90). The presence of a substantial IQ-MQ discrepancy grew to be considered diagnostic of clinical amnesic disorders (e.g., Milner, 1958).

David Wechsler is reported to have personally administered and interpreted his own tests in a flexible and individualized manner, at odds with test standardization guidelines today. Arthur Weider (personal communication, August 10, 1995), who was mentored by Wechsler at

Bellevue, disclosed, “He never gave the Wechsler the same way twice.” In practice David Wechsler was known to administer a single verbal subtest alone, without other subtests, for an assessment of intelligence and personality (Adam F. Wechsler, personal communication, December 3, 1993). Alan S. Kaufman (1992, 2009) recalls Wechsler’s assertion that he could even diagnose adult personality based on the response to a single item!

It is evident from a review of Wechsler’s (1939) *The Measurement of Adult Intelligence* that he readily interpreted test behaviors, response quality and errors, and problem-solving strategies. For example, Wechsler (1958; see also Kaufman, 1992, 2009) appreciated the interpretive value of responses to emotionally-loaded test stimuli, both words (e.g., *beer and wine, knife, gamble*) and pictures (e.g., the Taxi item on Picture Arrangement). Many of these evocative stimuli have been removed since his death.

### **Ultimate Outcome**

“The bankruptcy of the alleged practical approach,” wrote Raymond B. Cattell (1998b, p. 29), “is shown in the response of the WAIS, WISC, and other Binet derivatives to the question, ‘What is intelligence?’ their innocent reply is, ‘Intelligence is what intelligence tests measure.’” Cattell’s observation belies the intrinsic appeal of Wechsler’s clinically-useful measures, because a practical measure may be helpful with clinical and educational decisions while a theory-driven measure may not. Still, a pragmatic test without an associated theory cannot lead to explanations, extrapolations, and predictions beyond its own parameters.

With the Wechsler scales, however, something else seems to have happened—the test structure seems to have bootstrapped itself into a sort of theoretical framework. As a practical manifestation of Spearman’s *g*, the Full Scale IQ has become enthroned as a monarch of mental functioning. The Wechsler scales have been used by several generations of practitioners who

learned to think in terms of verbal-nonverbal (or verbal-visuospatial or verbal-performance) distinctions in intelligence, rather than fluid-crystallized distinctions. We further suspect that what most practicing psychologists know about information processing is not far removed from the working memory and processing speed composites in the Wechsler scales. It is not clear whether the incremental changes across new Wechsler editions will eventually unite practice with contemporary intelligence theory, but the Wechsler scales have clearly had a role in shaping scientific thought about intelligence.

David Wechsler (1943, 1950, 1981) was acutely aware of some limitations of his tests—specifically that they failed to systematically capture the *non-intellective factors*, such as drive, persistence, and interest, that substantially influence test (and life) performance. Wechsler (1943) defined non-intellective factors as “all affective and conative abilities which in any way enter into global behavior” (p. 103), and elsewhere he described them as “factors of personality” (Wechsler, 1950, p. 78). Wechsler became convinced of the importance of non-intellective factors after recurrent findings that factor analyses of his intelligence tests never extracted more than 60 to 70 percent of the total extractable variance, leading him to try (unsuccessfully) to develop tests for the remaining 30 to 40 percent. Despite several efforts, he was never able to successfully devise effective tests of the non-intellective factors.

Still the Wechsler scales have evolved, albeit slowly. Cohen’s (1952a, 1952b, 1957a, 1957b, 1959) exploratory factor analyses of the Wechsler scales with normal and clinical samples identified the third factor that he named Freedom from Distractibility or Memory, and Cohen’s five factor solutions anticipated contemporary five-factor CHC solutions (e.g., Weiss, Keith, Zhu, & Chen, 2013a,b). Edith Kaplan’s work (Kaplan, 1988, 1990; Wechsler, Kaplan, Fein, Kramer, Morris, Delis, & Maerlender, 2004) has served as a laboratory within which

Wechsler scores may be fractionated and special Wechsler process-based measures may be tested for applied, clinical value. As Kaufman (2013) recently expressed, changes in the Wechsler scales need to be driven by both practical-clinical and theoretical considerations:

No discussion of Wechsler's scales is complete without consideration of Wechsler the man and clinician; . . . no discussion of test profiles can be done in a psychometric vacuum, without keeping in the foreground Wechsler's credo that precision and technical sophistication cannot substitute for what real-world accomplishments tell us about a person's intelligence. (Kaufman, 2013, p. 225)

**“Not One Factor but Two”: Cattell, Horn, and Carroll,  
and the Extended Theory of Fluid and Crystallized Intelligence**

If one remembers that Charles Spearman was the discoverer and lifelong advocate for the primacy of the general intelligence factor  $g$ , it may be surprising to learn that his theoretical distinction between *eductive* and *reproductive* mental processes anticipated the dichotomy that would come much later between fluid and crystallized abilities. Spearman (1930a,c) contrasted the *eductive* mental processes, associated with reasoning abilities, with *reproductive* processes, especially repetition and rote retention. Spearman's work deeply influenced the development of Raymond B. Cattell's (1940) *Culture-Free Intelligence Test*, which contained three variants of matrix relations tasks depicting both pictorial objects and geometric forms. Cattell was Spearman's last doctoral student, completing his Ph.D. in 1929 and proposing his own theory of fluid and crystallized abilities in a 1941 American Psychological Association convention presentation and wide-ranging 1943 review of adult intelligence instruments. In 1943, Cattell hypothesized:

- (1) Adult mental capacity is of two kinds, the chief characteristics of which may be connoted by the use of the terms “fluid” and “crystallized.”
  - (2) Fluid ability has the character of a purely general ability to discriminate and perceive relations between any fundamentals, new or old. It increases until adolescence and then slowly declines. It is associated with the action of the whole cortex. It is responsible for the intercorrelations, or general factor, found among children’s tests and among the speeded or adaptation-requiring tests of adults.
  - (3) Crystallized ability consists of discriminatory habits long established in a particular field, originally through the operation of fluid ability, but not longer requiring insightful perception for their successful operation.
  - (4) Intelligence tests test at all ages the combined resultants of fluid and crystallized ability, but in childhood the first is predominant whereas in adult life, owing to the recession of fluid ability, the peaks of performance are determined more by the crystallized abilities.
- (R. B. Cattell, 1943, p. 178)

When introducing his Culture-Free Test, Cattell (1940) excoriated tests like the Binet-Simon that “measure a good deal of obviously acquired knowledge and skill” (p. 162). He was openly critical of test contents that “are frequently tests of scholastic attainment and life experience rather than ‘*G*’” (p. 163). Like Spearman, Cattell was interested in measures of education:

In general the kinds of test showing the best “*G*” saturation are those involving relation and correlate education [*sic*; he means *education*] in a high degree and reproduction in the lowest degree. (R. B. Cattell, 1940, p. 168)

Interestingly, this 1940 paper has been cited by Cattell as the place where he first put forth the

theory of *fluid* and *crystallized* abilities (Cattell & Butcher, 1968, p. 18), although neither of those words appear in the article; instead *eduction* and *reproduction* appear. As Cattell stated in 1943, he saw eductive/fluid intelligence tasks as being most strongly related to general intelligence. Looking back, Cattell (1987) regretted the timing of his introduction of fluid and crystallized intelligences, not only because of his high regard for Spearman who had spent his life defending a monolithic concept of *g*, but also because *g* theory had been important in the development of a scientific psychology, having “taken intelligence testing out of the realm of guesswork and quackery twenty years before” (p. 89).

Raymond B. Cattell (1905-1998) ranks among the most influential and prolific psychologists of the twentieth century, having introduced the theory of fluid (Gf) and crystallized (Gc) intelligence but being best-known for his personality work. Cattell authored over 500 articles and 43 books during his career. Cattell did not return to fluid-crystallized experimental studies until 1963, when school officials requested that intelligence tests be added to a study of personality he had initiated in order to predict school achievement (Carroll, 1984). In the resulting article, Cattell asserted that the general ability factor was “not one factor but *two*” (Cattell, 1963, emphasis in original). Cattell (1963) saw two highly cooperative, second-order general factors, Gf and Gc, that were difficult to disentangle, explaining why this joint factor had been mistaken others as one factor. In the best traditions of science, Cattell (1963) also generated nine testable propositions to support the theory.

Cattell completed several of the seminal publications in the theory of fluid and crystallized intelligence with John L. Horn (1928-2006). Horn was a Fulbright Fellow before working as Cattell’s research assistant at the University of Illinois from 1958 to 1961. He completed a remarkable 1965 dissertation on fluid and crystallized intelligence with Cattell as

his chair—remarkable not just because of his facility with the nuances of hierarchical factor analysis but also for an unusual erudition about the history of intelligence measurement and structure. Horn had, it seemed, read nearly everything ever written on the structure of intelligence, developing an encyclopedic range of knowledge on the subject.

In short order, Cattell and Horn expanded the number of broad ability factors from two to five (adding visualization, retrieval capacity, and cognitive speed; Horn, 1965; Horn & Cattell, 1966). In the next three decades, Horn arrived at nine higher order ability factors (Horn, 1985; Horn & Blankson, 2005; Horn & Stankov, 1982), while Cattell's list had grown to six ability factors (adding distant memory and retrieval) plus three provincial factors (visual, auditory, and kinesthetic; Cattell, 1971, 1987, 1998b) organized hierarchically. Beginning with Horn and Cattell (1966, 1967), numerous investigations demonstrated a characteristic developmental course--that Gc capabilities are maintained or improved over much of adulthood, while Gf reasoning abilities peak in late adolescence or young adulthood and then steadily decline over the rest of the life span.

### ***Theoretical Underpinnings***

In this section, we highlight some commonalities between Cattell and Horn as founders of the extended theory of fluid and crystallized intelligence. Toward the end, we will describe some differences. Cattell and Horn both resisted the imposition of single higher order general factor.

**Two intelligences.** *Fluid* ability (Gf) was described by Cattell (1943, 1963, 1971, 1987, 1998b; Cattell & Horn, 1978) and Horn (1965, 1976; Horn & Blankson, 2005; Horn & Cattell, 1966, 1967) as a facility in reasoning, particularly where adaptation to new situations is required and previously-acquired knowledge and learning assemblies are of limited utility. Ability is said



to be *fluid* when it takes different forms or utilizes different cognitive skill sets according to the unique demands of the problem requiring solution. Fluid ability is typically measured reasoning tasks (inductive, deductive, conjunctive, and distinctive), and it involves identifying relationships, comprehending implications, and/or drawing inferences from content that is either novel or equally familiar to all. When there is no advantage in task performance due to previous knowledge, fluid ability plays an increasing role as task complexity increases (Cattell, 1998b). According to Cattell's (1963) *investment theory*, fluid ability is the main determinant of performance in infancy (i.e., before culture-specific knowledge has been accumulated), driving learning processes and setting an upper limit on the possible acquisition of knowledge and crystallized skills over maturation. Horn cites a paucity of evidence to support investment theory but concludes it may operate in the early years of development, while adding that as development proceeds Gc may precede and do more to determine Gf than the reverse (e.g., Horn & Blankson, 2005).

*Crystallized* ability (Gc) refers to the depth and breadth of accessible stores of culture-specific knowledge and skills, as well as the capacity to acquire further knowledge via familiar learning strategies (Cattell, 1943, 1963, 1971, 1987, 1998b; Cattell & Horn, 1978; Horn, 1965, 1976; Horn & Blankson, 2005; Horn & Cattell, 1966, 1967). It is typically measured by recitation of language-based knowledge (e.g., factual information, word knowledge, quantitative skills, and language comprehension tasks), because these are the domains of knowledge that are culturally valued and educationally relevant in the Western world (Cattell, 1943, 1963, 1971, 1987; Horn & Cattell, 1966). Knowledge may be said to be *crystallized* when it takes a definite form, analogous to the hard calcium carbonate skeletons produced by the marine corals. Cattell used the analogy of the coral reef to convey that the configurations by which crystallized

knowledge is represented (i.e., the dead coral formations) reflect the properties, limitations, and intercorrelations of the originating fluid abilities (i.e., the living corals that first secreted the exoskeleton) (Cattell, 1943, p. 178-179; see also Cattell & Butcher, 1968, p. 21). In terms of predictions of criteria, Cattell (1998b) specified that crystallized ability looks backward because it represents the effects of past applications of fluid ability and the amount, intensity, and opportunity for education. Cattell (1998b) also noted that crystallized ability tends to have significantly less variability than fluid ability, perhaps due to pressure for educational conformity: “The busy teacher puts upward pressure on the dull and leaves the really bright more to their own devices” (p. 32).

**Triadic theory.** Cattell’s (1971, 1987, 1998b) final statements on the nature and structure of intelligence were found in triadic (*three-in-one*) theory, in which he described three taxonomic classes of ability involved in any cognitively-oriented observed behavior: general capacities or *g*’s, provincial organizations or *p*’s, and agencies (primary abilities) or *a*’s. General capacities were conceptualized as broad, higher order factors that function as limiting properties of the brain; Cattell saw them as operating across all cognitive performances. He specified four general ability factors: fluid ability (*g<sub>f</sub>*), crystallized ability (*g<sub>c</sub>*), retrieval/fluency efficiency (*g<sub>m</sub>*), and speed (*g<sub>s</sub>*). Cattell also identified a group of local provincial factors, some broader and some narrower, that are involved with basic pattern perception and organization of sensory and motor information: auditory ability (*p<sub>a</sub>*), visual ability (*p<sub>v</sub>*), tactile ability, kinesthetic ability, olfactory and gustatory judgment, as well as cerebellar and general motor abilities. Finally, the narrow primary mental abilities were termed *agencies* and were conceptualized as aids and proficiencies of sorts that are learned. These three classes of abilities may interact with motivational variables (shaped through previous reinforcement schedules) in the production of behaviors.

**Horn's eight broad factors.** Near the end of his life, Horn considered factor analytic evidence to support some eight (or nine) broad, second-order factors: acculturation-knowledge (Gc), fluid reasoning (Gf), short-term apprehension and retrieval (SAR), fluency of retrieval from long-term storage (TSR), processing speed (Gs), visual processing (Gv), auditory processing (Ga), and quantitative knowledge (Gq); he also enumerated some 80 first-order factors, or primary mental abilities, although he suspected that many more exist (e.g., Horn & Blankson, 2005). Horn rejected a third-order *g* factor, noting that the structural and developmental evidence for a single general factor remains weak, and the use of the positive manifold to justify *g* could equally apply to all non-cognitive traits that are positively intercorrelated.

**Fluid reasoning synonymous with *g*.** In the 1980s, several investigations by Johan O. Undheim and Jan-Eric Gustafsson (Gustafsson, 1984, 1988; Undheim, 1981a,b; and Undheim & Gustafsson, 1987 ) advanced a new synthesis of Spearman's *g* with Cattell and Horn's extended *Gf-Gc* model within the context of a hierarchical structure of cognitive abilities. Using multiple methodologies including hierarchical exploratory factor analyses and confirmatory factor analyses, Undheim and Gustafsson argued for a hierarchical cognitive structure and retention of a higher order *g* factor that was functionally equivalent to a second order fluid reasoning factor, *Gf*.

**Carroll's three stratum theory.** Just a few years later, these findings would be reinforced by polymath John B. Carroll (1916-2003) who had been a one-time student of Thurstone as well as serving as director of the University of North Carolina's L. L. Thurstone Psychometric Laboratory from 1974 to 1982. Carroll spent his retirement reanalyzing over 460 archival data sets in producing the 1993 classic, *Human Cognitive Abilities: A Survey of Factor-*

*Analytic Studies*, that remains definitive two decades after its publication. In *Human Cognitive Abilities*, Carroll built upon the work of Cattell and Horn by proposing a hierarchical, multiple-stratum model of human cognitive abilities with the general intelligence factor, psychometric *g*, at the apex (or highest stratum); eight broad group factors of intelligence at the second stratum; and at least 69 narrow factors at the first (or lowest) stratum. Carroll (1983, 1993, 1994) accumulated over a thousand archival datasets related to human cognitive test performance, ultimately using 461 of the datasets for his analyses. He conducted hierarchical exploratory factor analyses, subjecting datasets to the Schmid-Leiman orthogonalized hierarchical factor procedure, with factor interpretations being based on the resulting hierarchical factor matrix. Carroll's hierarchical analyses led him to report eight broad factors at the stratum below general intelligence; listed by their descending strength of association with *g*, they are 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). Each of these eight broad factors could be further divided into narrow first stratum cognitive abilities. Carroll's three stratum theory has been integrated with the work of Horn and Cattell on 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).

### ***Tests and Measures***

There are several plausible candidates as the first tests to measure the theory of fluid and crystallized intelligence, and some of them predate articulation of the theory by Cattell (1941, 1943).

**Shipley Institute of Living Scale.** Early tests of crystallized ability frequently relied upon *vocabulary* as the measure of a person's knowledge base. Following the lead of Harriet Babcock (1930) who proposed that vocabulary scores are resistant to deterioration from psychiatric and neurological disorders, Walter C. Shipley (1940) developed the *Shipley Institute of Living Scale* (SILS) as a quick, objective, self-administering measure consisting of a Vocabulary Test (*crystallized* in our view) and an Abstraction Test (*fluid* in our view). The magnitude of the discrepancy between Vocabulary and Abstraction was to be interpreted as an index of deterioration, based on the clinical and experimental finding that "in mental deterioration *vocabulary level tends to be affected but slightly*, while the *ability to see abstract relationships declines rapidly*" (Shipley, 1940, p. 371, emphasis in original). The SILS was revised in 2009, adding an optional nonverbal fluid measure (Shipley, Gruber, Martin, & Klein, 2009). In a 2010 review, Thorndike-Christ reiterated its theoretical foundation: "The test was and is grounded in the assumption that crystallized and fluid abilities, taken together, provide a good estimate of overall cognitive skill and that these abilities are differentially resistant to impairment and cognitive deterioration with crystallized knowledge being more stable than fluid reasoning skills."

**Raven Progressive Matrices and Mill Hill Vocabulary.** The second fluid and crystallized measures we will mention, although they did not use those terms, were a pair of measures developed by John C. Raven in the late 1930s. As an operationalization of Spearman's distinction between eductive and reproductive cognitive processes, Raven developed the Raven's Progressive Matrices (Penrose & Raven, 1936; Raven, 1936, 1938) and its complementary vocabulary test, the Mill Hill Vocabulary Scale (Raven & Walshaw, 1944). The Raven Progressive Matrices consisted of a series of designs presented in a rectangular pattern with a

part missing; examinees were asked to select the correct part from multiple choices to complete the pattern. The designs making up a pattern could be analyzed horizontally or vertically. In general, Penrose and Raven (1936) considered items requiring analogic reasoning to select the correct design as the best measures of education, but they developed many conceptual principles by which items could be solved including analysis of design similarities, opposites, addition or subtraction of elements, ordinal sequence, and others. The most difficult items could be generated by using combinations or reduplications of ordinal and additive relations. After the Standard Progressive Matrices (1938) were created, new forms were added including Colored Progressive Matrices and Advanced Progressive Matrices. Raven's Progressive Matrices have been in continuous use since their development, with normative studies in dozens of countries. Several recent forms, most importantly the Standard Progressive Matrices Plus (SPM Plus; Raven, Raven, & Court, 1998), have been developed, and the use of matrix reasoning measures have become common in intelligence assessment. In a 1949 review, David Wechsler praised Raven's Progressive Matrices but expressed concern about the problems inherent in the use of only a single assessment methodology to appraise intelligence:

The main limitation of the Progressive Matrices is that they attempt to measure general intelligence (or even educative ability) through a single modality of performance. To this there are serious theoretical objections; but leaving these aside there still remains the practical consideration that there is no test, no matter how excellent, but that certain individuals may fail to do justice to themselves because they simply are not good in doing the test or type of task used. The Progressive Matrices are no exception. (Wechsler, 1949, p. 259)

If the Progressive Matrices were eductive (and *fluid*), then Raven's Mill Hill Vocabulary Scale (Raven & Walshaw, 1944) provided the reproductive (and *crystallized*) counterpart. The Mill Hill Vocabulary Scale consists of 88 words, equally divided between two sets (A and B). Examinees are asked to give a verbal (or written) statement about the meaning of each word in Set A, and choose the best synonym from six multiple choices in Set B. An entirely multiple choice version and a short form version of this test are also available. The test takes about 15 to 40 minutes to administer. Raven (1948) considered the Mill Hill Vocabulary Scale as a test of the general fund of knowledge a person has acquired as a result of intellectual activity in the past.

The use of Raven's Progressive Matrices in Great Britain grew quickly, with Slater (1948) reporting that it was second only to the Binet-Simon Scale among intelligence tests. Slater also reported Philip Vernon's factor analytic findings that showed the Progressive Matrices to yield a *g* loading of about 0.8 and negligible loadings on other factors. In 1948, Raven reported distinctive life span normative performance curves for the Progressive Matrices and Mill Hill Vocabulary Scale in a large sample between the ages of 4 and 65 years. From matrix relations results, he reported that "the capacity to form comparisons and reason by analogy increases rapidly during childhood, appears to have reached its maximum somewhere about the age of 14, stays relatively constant for about 10 years, and then begins to decline, slowly but with remarkable uniformity" (pp. 15-16). From vocabulary results, he found that "the average person's ability to recall information increases steadily up to the age of about 25-27, remains almost constant for the next 25 years, and then declines a little" (p. 17). These growth curves anticipated those later reported for fluid and crystallized abilities.

We would be remiss if we neglected Raymond B. Cattell's own multidimensional measure of education, the *Culture-Free Intelligence Test* (Cattell, 1940; Cattell, Feingold, &

Sarason, 1941), which was developed using some of the same nonverbal test procedures that Spearman had pioneered. Cattell selected seven subtests: Mazes, Series, Classification, Progressive Matrices I, Progressive Matrices II, Progressive Matrices III, and Mirror images. All but Mazes were considered sound measures of *g*, and the use of multiple subtests was thought to avoid placing disproportionate weight on any single procedure. This test never reached the test usage rate of the Raven's measures and suffered from an unfortunate name, with most psychologists believing that no test can completely remove all cultural influences.

**Omnibus measures.** The first omnibus intelligence test to explicitly tap Cattell and Horn's theory of fluid and crystallized intelligence was the *Stanford-Binet Intelligence Scale, Fourth Edition* (SB IV; R. L. Thorndike, Sattler, & Hagen, 1986). The SB IV was structured hierarchically, with Spearman's *g* at the apex, four broad group factors at a level lower (crystallized ability, fluid-analytic ability, quantitative reasoning, and short-term memory), and individual subtests at the lowest level.

Along with the 1986 Stanford-Binet Intelligence Scale, the *Woodcock-Johnson Psycho-Educational Battery--Revised* (WJ-R; Woodcock & Johnson, 1989) cognitive battery was the first test developed principally around extended Gf-Gc theory and the approach that would come to be known as the Cattell-Horn-Carroll model. Woodcock began work on the first edition of this battery of tests in 1963 in a series of controlled experiments for measuring learning ability. During construction of the revised edition, he met with John L. Horn and John B. Carroll in the fall of 1986 (Newton & McGrew, 2010), and perhaps not surprisingly, the WJ-R normative dataset is the most complete measure of CHC theory to appear in Carroll (1993). Another meeting with Horn and Carroll occurred in 1999 in preparation for the WJ III and the Stanford-Binet 5th edition (Newton & McGrew, 2010). McGrew (2009) suggests that Carroll's (1993)



survey was a “pivotal landmark or tipping point that provided the first working map of the human cognitive ability terrain” (p. 9), and following this logic, the WJ III Tests of Cognitive Abilities may represent the first way that intelligence practitioners can use the map for practical travel. The historical significance of the WJ III Tests of Cognitive Abilities, however, cannot be fully understood at this early juncture.

### *Ultimate Outcome*

Several major intelligence theorists have cautioned about building theories principally from factor analyses. Acutely aware of the limitations of factor analysis, Carroll (1980) cautioned that “Factor analysis, as such, is not a theory of intelligence; it is a method of analysis” (p. 14). In 1983, he specified a number of additional forms of theory validation that transcend factor analysis including establishing the nature of a factor, its psychometric characteristics, its place in a hierarchy of abilities, its developmental characteristics, its genetic and environmental determinants, the presence of any demographically-based group mean differences, its susceptibility to intervention or more transient influences such as drugs or fatigue, its relationship to noncognitive variables, its ecological relevance and validity, and its implications for psychological theory as a whole.

Even Raymond B. Cattell expressed some caution about factorially-derived theories. At a 1994 Human Cognitive Abilities conference also attended by Carroll, Horn, and Woodcock, Raymond B. Cattell disparaged contemporary attempts to build theory based on factor analyses:

The only thing that I found disappointing is the building of necessarily rickety theory on the basis of factor analysis that is published prematurely without being checked. There is no doubt in my mind that we have a plague of poor factor analyses—that is to say, factor analyses that are not rotated definitely to simple structure, to a significant simple

structure. I can understand this because you get very weary in long factor analysis. It reminds me of the early attempts to find the North Pole. There were many attempts before people got there, but they all thought they got there. (Cattell, 1998a, 305-306)

The founder of multiple factor analysis, L. L. Thurstone (1937) also saw mathematical treatment of constructs in psychology as just one step, and “in no sense a substitute for the exploratory and descriptive types of experimentation” (p. 7). One method of derivation cannot define a theory.

Still, at the time of this writing, the Cattell-Horn-Carroll theory of cognitive ability must be considered the most esteemed of intelligence theories in the mental testing literature. The historical status of this theory in its current incarnation, however, is quite young and few serious challenges to it have been mounted. Perhaps the most trenchant critique was offered by Carroll (1998), himself. Shortly before his death, Carroll (2003) even expanded his model to include ten second stratum factors, signaling that even this definitive model may be subject to change and expansion. Horn and Blankson (2005) put it wisely:

The extended theory of fluid and crystallized (Gf and Gc) cognitive ability is wrong, of course, even though it may be the best account we currently have of the organization and development of abilities thought to be indicative of human intelligence. All scientific theory is wrong. It is the job of science to improve theory. (p. 41)

CHC theory, as it currently stands, has considerable promise with a number of substantial limitations: a) although it encompasses and aggregates many relatively independent cognitive abilities, it lacks the integrative “glue” of good theory; b) the model remains mainly descriptive, and as a result it is difficult to extract testable predictions capable of falsifying the model; c) CHC theory offers few neural underpinnings and remains disconnected from brain function; d) because factors have continually been added for the last 40-50 years, the CHC model appears

interminable; and e) the continued accumulation of factors in the model undermines its foundation, because the inclusion of new marker tests in factor analyses may fundamentally alter the existing factor structures (see Woodcock, 1990, for an illustration of how factors break apart when new markers tests are added). To date, the CHC model has not been integrated well with the enormous bodies of knowledge that have accumulated in clinical psychology, developmental psychology, the cognitive sciences, neuropsychology, and other fields of applied psychology. Most importantly, the CHC model is limited in its ability to inform practitioners about the relevance of its constituent factors and their dynamic interactions to clinical, educational, or occupational decision-making. Carroll (1993) 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).

**“The Computer Itself”: Cognitive Neuroscience, Luria, and Information Processing**

It was because the activities of the computer itself seemed in some ways akin to cognitive processes. Computers accept information, manipulate symbols, store items in “memory” and retrieve them again, classify inputs, recognize patterns and so on . . . Indeed the assumptions that underlie most contemporary work on information processing are surprisingly like those of the nineteenth century introspective psychology, though without introspection itself. (Neisser, 1976, pp. 5-7)

So wrote Ulric Neisser to describe the *computer* as a metaphor for (and simulator of) the human mind and brain, capable of elucidating elements of information processing. The revolution in psychology associated with this metaphor was cognitive psychology, beginning in the late 1940s and 1950s and borne out of a rejection of the behaviorist tenet that internal mental events cannot be part of a scientific psychology (e.g., Baars, 1986; Gardner, 1985; Leahey, 1992; Miller, 2003).

Through the contributions of such luminaries as Broadbent, Chomsky, Miller, Newell and Simon, and Tanner and Swets, among others, it became possible to conceptualize mental operations as part of an information processing system that manipulates symbolic structures (e.g., Newell & Simon, 1972). Consequently, the cognitive revolution made it acceptable for mental tests to aspire to measure internal cognitive processes, leaving behind the disparagement that “intelligence is what the tests test” (Boring, 1923, p. 35).

An examination of cognitive processing typically seeks to determine *how* problem-solving takes place; which mental operations, mechanisms, and strategies are used; and how operations are sequenced or simultaneously engaged en route to solving a problem. Earl Hunt (1980) specified that the value of the computer analogy stems from universal requirements of any physical information processing system, whether it is a computer or the human mind. According to Hunt, every problem-solving machine must have: (a) a *structural* capacity to store, retrieve, and transform information; (b) a *process or program* to function in a specified (and highly flexible) order ; and (c) access to previously acquired *knowledge* that can be readily compared, and coordinated with the present situation. Although the distinctions are not always fixed, cognitive processing theories tend differ from multifactor theories of intelligence in their origins (usually clinical-theoretical derivation versus empirical-factorial derivation), research foundations (cognitive/neuroscience versus more strictly psychometric), and emphases (operations performed versus enduring performance traits).

Cognitive processes are, however, difficult to quantify and analyze statistically, and in his Presidential address to the American Psychological Association, Quinn McNemar (1964) called for mental testers to move beyond their reliance on response products alone (i.e., abilities, capacities, aptitudes, and skills) in measuring intelligence:

Studies of individual differences never come to grips with the process, or operation, by which a given organism achieves an intellectual response. Indeed, it is difficult to see how the available individual difference data can be used even as a starting point for generating a theory as to the process nature of general intelligence or of any other specified ability. (McNemar, 1964, p. 881)

Some breakthroughs would come from neuroscience. Pioneering neuroscientists Alexander R. Luria (1902-1977) and Roger W. Sperry (1913-1994) provided fertile ground for major advances in cognitive psychology and neuropsychology. As Luria—a clinician—conveyed consistently about his patient assessments, which could be conducted over days and in many variations:

The person investigating intellectual activity must pay particular attention not so much to the result or solution of the particular problem as to the analysis of the course of the intellectual process, and the identification of the difficulties which the patient experiences when solving the problem” (Luria, 1980, p. 565)

Among others, the work of neuroscientists like Luria and Sperry would inspire Alan S. Kaufman (the co-author of this chapter), who would influence the theory and practice of intelligence testing in the last quarter of the twentieth century. Kaufman completed his doctorate in psychology at Columbia University in 1970, with Robert L. Thorndike as his major professor. From 1970 to 1974, he was employed as an assistant test development director at The Psychological Corporation, working closely with David Wechsler on the *Wechsler Intelligence Scale for Children—Revised* (WISC-R). When he subsequently wrote *Intelligent Testing with the WISC-R* (1979), it was in part keeping a promise he had made to Wechsler (Kaufman, 1992). Through his revisions of *Intelligence Testing*, his *Essentials* assessment series, his assessment textbooks, and his cognitive processing tests co-authored with his wife, Kaufman has educated

several generations of psychologists. As evidenced through the remarkable number of psychological tests created by his former graduate students at the University of Georgia, where he taught from 1974 to 1979, Kaufman served as a catalyst to a distinguished group of psychological test authors. In the words of his students Randy W. Kamphaus and Cecil R. Reynolds (2009), Kaufman's contribution having the greatest long-term impact was his "joining of the two disciplines of measurement science and clinical assessment practice" (p. 148).

In 1978, Alan and his psychologist-wife-coauthor Nadeen L. Kaufman began developing what would become the *Kaufman Assessment Battery for Children* (K-ABC; Kaufman & Kaufman, 1983; see also KABC-II published in 2004). The K-ABC was developed to measure intelligence based on a brain-based model of problem-solving and information processing (sequential and simultaneous processing; e.g., Luria, 1966, 1970a), to separate acquired factual knowledge (i.e., crystallized intelligence, measured through an Achievement scale) from the ability to solve unfamiliar problems (i.e., fluid intelligence, measured through the Mental Processing scale), to yield scores that translate to educational intervention and remediation, and to be sensitive to diverse needs of preschool, minority group, and exceptional children. The K-ABC was notable not only because of its practical innovations, but more so because it offered a new model for appraising cognitive processes based on the neuropsychological literature, finally realizing the vision of intelligence testing rooted in brain science (e.g., E. L. Thorndike, 1901). A revision (KABC-II; Kaufman & Kaufman, 2004) was published some two decades later, and an independent measure derived from the same Lurian cognitive processing theory has also been published (Naglieri & Das, 1997).

### ***Theoretical Underpinnings***

The K-ABC was developed by Kaufman and Kaufman (1983) to measure the cognitive processing theory of Alexander R. Luria, the Russian neuropsychologist, who has inspired development of several intelligence tests and neuropsychological test batteries. Luria ranks as the most cited Russian scholar in the Western psychology literature (Solso & Hoffman, 1991).

In brief, Luria (1966, 1970a, 1973, 1980) conceptualized the brain as being organized in distinctive yet integrated systems, each hierarchical and multilayered and serving different mental processes:

Human mental processes are complex functional systems and . . . they are not ‘localized’ in narrow, circumscribed areas of the brain, but take place through the participation of groups of concertedly working brain structures, each of which makes its own particular contribution to the organization of this functional system. (Luria, 1973, p. 43)

Luria’s (1966) notion of “dynamic localization” (p. 37) and “working constellations” (p. 38) underscored the concept of a combined, interdependency of brain-based processes to produce complex behaviors; the cerebral cortex was neither undifferentiated (as argued by the antilocalizationists like Marie Jean Pierre Flourens in the nineteenth century and Karl Lashley in the twentieth) nor marked by localized functions like cities on a map (as argued by the phrenologists).

Luria (1966, 1970a, 1973, 1980) described three principal functional units of the brain that participate in all mental activities: (1) a system seated in the upper and lower parts of the brain stem, particularly the reticular formation, that controls arousal, alertness, wakefulness, cortical tone, and responsiveness to environmental stimuli; (2) a system located in the posterior parts of the cerebral cortex, particularly primary sensory and association areas for vision, audition, and tactile sensation, involved in analysis, coding, and storage of information according

to sensory modality and level of modal integration; and (3) a system seated in frontal lobes, particularly the prefrontal cortex, that is involved in attention, intentions, programs, regulation, and verification of mental activity. Each of these basic units was conceptualized as being hierarchical in structure, consisting of primary, secondary, and tertiary cortical zones. The *primary* (projection) areas receive impulses from or send impulses to the periphery, the *secondary* (projection-association) areas conduct processing and programming, and the *tertiary* (overlapping) zones integrate complex events across modalities.

Higher forms of activity and behavior involve all parts and levels of the brain, each of which “makes its own special contribution to the work of the functional systems as a whole” (Luria & Majovski, 1977, p. 962). These three systems could be further subdivided into areas of specialization mediating dozens of mental processes (e.g., Luria, 1980). Among the processes described by Luria (1966, 1973, 1980) were *simultaneous* (all-at-once) processing seated in the parietal-occipital region and *sequential* (serially-ordered, or successive) processing, seated in the frontal-temporal region of the brain. Luria (1970b, pp. 221-222) provides an elegant example of these different forms of processing in his analysis of visual perception of Leonardo da Vinci’s *The Last Supper*: To draw a general impression of the entire picture, it is necessary to expand one’s field of vision and simultaneously integrate the many figures at once, but to consider one element at a time (e.g., the face of Christ) one must constrict the visual field and suppress attention to the periphery, moving from one area of focus to another to arrive at an understanding of the painting through sequential analysis of its personages.

After a 1975 visit with Luria, J. P. Das and his colleagues articulated the simultaneous-successive synthesis of cognitive abilities (Das, Kirby, & Jarman, 1975). In 1994, Das, Naglieri, and Kirby (1994) expanded the Lurian-derived model to encompass four processes given the



acronym PASS (planning, attention, simultaneous, and successive processes). Based on Luria's extensive writings about memory (Luria, 1973, 1976, 1980), Kaufman and Kaufman (2004) added learning processes to the model so the KABC-II now measures planning, learning, simultaneous, and sequential processes. Korkman (1999) listed 29 components of cognitive processes specified by Luria (1973) in his introduction to neuropsychology, *The Working Brain*, so there may be a rationale for thoughtful expansion of these batteries in the future.

### ***Tests and Measures***

Luria's work provides the theoretical underpinnings for two major contemporary intelligence tests: the *Kaufman Assessment Battery for Children* (K-ABC and KABC-II; Kaufman & Kaufman, 1983, 2004) and the *Cognitive Assessment System* (CAS; Naglieri & Das, 1997). Two comprehensive neuropsychological test batteries have also been inspired by Luria: the *Luria-Nebraska Neuropsychological Battery* (Golden, Purisch, & Hammeke, 1979) and the *NEPSY: A Developmental Neuropsychological Assessment* (Korkman, Kirk, & Kemp, 1998). Luria's theory also had a profound effect upon Edith Kaplan (1988, 1990), who became the leading advocate for the Boston Process Approach to neuropsychological assessment, an approach which formally measures problem-solving strategies, error types, and performance under different conditions for a variety of tasks—all with the goal of quantifying test performance in terms of component functions and processes. One of the first formal process-based measures derived from the Boston Process Approach was the *WAIS-R as a Neuropsychological Instrument* (WAIS-R NI; Kaplan, Fein, Morris, Kramer, & Delis, 1991).

To illustrate one well-known Luria-inspired cognitive processing measure, we will briefly describe the KABC-II (Kaufman & Kaufman, 2004). The KABC-II measures cognitive abilities and processing in children and adolescents from age 3 years through 18 years. It may be

interpreted according to a dual theoretical foundation, which should be selected before testing, either through Luria's cognitive processing approach or the Cattell-Horn-Carroll psychometric model. Kaufman and Kaufman's (2004) design for the KABC-II included *Planning* ability measures with low speed requirements (also tapping *Fluid Reasoning*, or Gf, in the CHC model); *Learning* ability measures (also tapping *Long-Term Storage or Retrieval*, or Glr), *Simultaneous Processing* measures (also tapping *Visual Processing*, or Gv), and *Sequential Processing* measures (also tapping *Short-Term Memory*, or Gsm). If the CHC model is selected *a priori*, then additional measures of *Crystallized Ability*, or Gc, are also administered. Accordingly, the KABC-II Lurian framework features planning, learning, simultaneous processing, and sequential processing, yielding a composite Mental Processing Index (MPI). The CHC framework tests Gf, Glr, Gsm, Gv, and Gc, yielding a composite Fluid-Crystallized Index (FCI). The authors explain that the CHC model is generally preferred unless there is a rationale to deemphasize the verbal expression of knowledge (e.g., the examinee has a background or diagnosis that depresses verbal functioning and knowledge acquisition):

The CHC model should generally be the model of choice, except in cases where the examiner believes that including measures of acquired knowledge/crystallized ability would compromise the validity of the Fluid-Crystallized Index (FCI). In those cases, the Luria-based global score (MPI) is preferred. The CHC model is given priority over the Luria model because the authors believe that Knowledge/Gc is, in principle, an important aspect of cognitive functioning. (Kaufman & Kaufman, 2004, pp. 4-5)

Consistent with Luria's emphasis on observation of problem-solving processes during task performance, every KABC-II subtest has an optional list of observable Qualitative Indicators (QIs) that may either disrupt or enhance task performance.

### *Ultimate Outcomes*

Cognitive psychology and neuroscience began to recognizably influence more traditional intelligence tests by the 1980s. The application of cognitive processing concepts to the assessment of intelligence was highlighted by Robert L. Thorndike, lead author of the *Stanford-Binet Fourth Edition*, in his 1984 monograph, *Intelligence as Information Processing: The Mind and the Computer*. Beyond more complex decision-making algorithms and strategy heuristics, Thorndike singled out four fundamental processes shared by both humans and computers: information processing speed, short-term or working memory processes, long-term memory storage processes, and executive control processes (which he saw as most similar to the computer's central processor). Several of these processes are now included in traditional intelligence tests, like the Wechsler intelligence scales.

Luria's influence continues to be pervasively felt through neuropsychology. Daniel Tranel (2009) recently wrote that Luria "stressed the critical importance of understanding the distinct ingredients that made up complex mental functions" and concluded, "Luria's contributions have survived, and perhaps even thrived, with the test of time, and contemporary students and scholars would do well to heed his teachings" (p. 35).

Although it is beyond the scope of our immediate chapter, one comprehensive neuropsychological test battery modeled after Luria's neuropsychological investigation—the *Luria-Nebraska Neuropsychological Battery*—appears to have been falsified on multiple grounds including clinical content validity, inconsistent statistical reporting, psychometric problems, and concerns over scientific integrity (Adams, 1980a, 1980b, 1984; Crosson & Warren, 1982; Delis & Kaplan, 1982, 1983; Russell, 1980; Spiers, 1981). Throughout these many challenges, there were virtually no aspersions on Luria's original writings, only concerns

about the way in which his writings were adapted into the LNNB by Golden, Purisch, and Hammeke (1979).

The K-ABC introduced a new model for appraising intelligence based on the neuropsychological processing literature, and the 2004 KABC-II expanded its breadth to measure four forms of processing but can also be interpreted using either the Luria processing approach or a Cattell-Horn-Carroll approach in a dual theoretical formulation. The K-ABC is probably too young as an intelligence test for its impact and influences to be fully appraised (and we are admittedly not neutral about its qualities).

### **Conclusion**

“The history of science” wrote J. McKeen Cattell in 1893, “is the history of measurement” (p. 316). He envisioned mental testing as enabling the discipline of psychology to join the other natural sciences and harbored visions for testing in American society. He saw tests given to schoolchildren as helpful in “determining their progress and the classes to which they should be assigned” (Cattell, 1923, p. 168), with tests used in the business world in “selecting individuals for the work for which they are best fit” (p. 169). Cattell’s visions have arguably been realized to a significant degree.

If we accept the proposition that mental measurement is a science, and we do, then we are obligated to demonstrate that it employs the methods of science. In terms of theory building, mental measurement has offered sensory-motor processing theories (e.g., Galton and Cattell’s anthropometry), few brain-based theories (e.g., Kaufman and Kaufman’s simultaneous-sequential processing), several mathematically-derived theories (Spearman’s two-factor theory, various factor analytic models including Carroll’s three stratum model), and several practical approaches that perhaps do not fully qualify as theories (Binet and Simon, Terman, and

Wechsler). Cattell and Horn combined elements from several of these approaches in arriving at their extended fluid and crystallized intelligence theory.

Evaluation of the qualities of these theories by Kuhn's (1977) five characteristics (accuracy, consistency, scope, simplicity, and fruitfulness) suggests that mental testers generally do not form particularly fruitful theories, if the capacity to predict that which has not yet been observed is considered. Ideally, the scope of theories should not be limited by the properties of a single measurement tool, and most of these theories have yielded but one corresponding test (an exception is the Lurian processing theory, which has yielded tests from Kaufman and Kaufman [1983, 2004] and Naglieri and Das [1997]). Several theories fail to articulate how their core elements meaningfully interact, making them more compilations rather coherent theories to our thinking.

In their seminal paper on construct validity, Cronbach and Meehl (1955) describe the multiple sources of evidence that should be accumulated and subjectively evaluated when determining the validity of a test and theory. They also describe the implications of negative evidence, when results are inconsistent with what a theory predicts. From their scholarly perspective, the processes of validation and falsification are rational, balanced, and sustained. Our review of the history of mental testing suggests that there may be much more subjectivity in the falsification process than previously thought, and we are reminded of the lung capacity and sustained effort required to inflate a large balloon and the brevity of the prick that punctures it. We begin with the methodologically flawed falsifications, Sharp (1899) and Wissler (1901), puncturing anthropometry. Even though both theories were fatally flawed to begin with (and merited falsification), there is now agreement that the falsifiers used flawed, one-dimensional methodologies (e.g., Buckhalt, 1991; Deary, 1994; Jensen, 2006). Even today, we see single

confirmatory factor analyses misrepresented as a sole and sufficient basis for falsification (e.g., Keith & Kranzler, 1999) when multiple methods are always optimal for theory validation or falsification. We may have some lessons to learn from the history of (false) falsifications.

The sociopolitical subjectivity of falsification is also troubling. Binet's work is emblematic, because after his death his work was largely abandoned in France and treated as if it had been falsified, the victim of antipathy from a rigid French establishment and a laboratory successor, Henri Piéron, with his own theory to promulgate. The fate of Binet's work in the United States was a striking contrast, as the Binet-Simon scale popularized by Henry H. Goddard was embraced by psychologists and spawned the revolution in mental testing. That the same work was rejected in one society and embraced in another indicates the subjectivity (and unreliability) of sociopolitical decision-making labeled as scientific.

The tenacity and reputation of the theoretician also seems to predict the longevity of a theory. Charles Spearman's general factor of intelligence survived due to his remarkable willingness to engage and debate critics, never letting a challenge go without a response, although there were several challenges that could potentially have falsified the theory. On the other hand, J. McKeen Cattell appears to have given up on his anthropometric tests after the Wissler (1901) and Sharp (1899) studies. Cattell was temperamentally tenacious to a fault, but he may have felt overmatched by Wissler's study, possibly the first correlational study completed in the United States. The lesson to theoreticians is to mount a vigorous (and rational) defense of one's theory.

Finally, our historical review suggests that practical measures without testable theory (e.g., the Binet-Simon, Stanford-Binet, and Wechsler scales) seem to have a longer life than measures that are more easily tested (e.g., Galton-Cattell's anthropometry). Even though we

strongly believe that theory advances science, test developers who invest hundreds of thousands of dollars into test development might be advised that their tests should remain theoretically-light, so as not to be easily disproven. In a spirited defense of one falsified intelligence test we have not discussed (the *Structure of Intellect Learning Abilities* tests [Meeker, Meeker, & Roid, 1985]), Roid (1985) rued the tendency of critical test reviewers to “degenerate into a negative approach which does not recognize the practical difficulties of developing and standardizing worthwhile clinical instruments” (p. 121), asking, “Isn’t it fair to give time for developing tests (for which a stock of printed materials may already be warehoused), to shift to newly developed techniques?” (p. 122). Speaking for ourselves, we believe that it is in the best traditions of science to build tests with well-articulated, testable theoretical underpinnings.

### **Future Directions**

Our examination of the history of mental testing and theory identifies some unresolved strands that one of us has called “loose threads” (Wasserman, 2012, p. 36), persisting over the last 100 or more years. It is in part the failure to definitively resolve these ongoing questions that explains why long-time observers apply words like “little progress” (Buros, 1977, p. 10) and “stagnation” (Carroll, 1978, p. 93) to the history of mental measurement. In these closing comments, we list a few of these threads needing investigation and resolution.

First, there needs to be more research into the nature of Spearman’s general ability factor; more than one hundred years after its introduction, it must be explained and understood in terms beyond those reached by Charles Spearman (1923, 1927) eight to nine decades ago. There are some indicators of progress in this area, with some important research in this direction having been completed by Linda Gottfredson (1997b, 2002a, 2002b), for example, but to apply the old chestnut, more research is needed.

Second, as David Wechsler knew, clinical and educational considerations need to be fully integrated with existing models of cognitive ability. Within the Cattell-Horn-Carroll model, the number of broad ability factors has grown in various accounts to ten (Carroll, 2003) or 15 or 16 (McGrew, 2009; Newton & McGrew, 2010). More important than this growing number are questions about which factors best identify disabling clinical conditions like specific learning disabilities, which deficient cognitive processes they tap, and which factors identify the need for nontraditional intervention approaches. Complex structural models of cognitive ability become most meaningful when individual factors are relevant to impaired performance and intervention, when needed. Theory needs to meet practice.

Third, any science of mental testing needs ultimately to be based on a neuroscientifically-grounded theory of brain functioning. This aspiration, first expressed in a misbegotten manner by the phrenologists and again at the beginning of the twentieth century by E. L. Thorndike in his writings on connectionism, has been partially realized through operationalizations of Luria's theory of functional neural systems in the brain but appears absent from Cattell-Horn-Carroll theory. Mental testing ultimately needs to be reconciled with what is known (or to be learned) about the functioning of the brain.

Finally, we see the need for more training and education in the philosophy of science, the philosophy of psychology, and the art of theory-building. To advance psychological science, measures of intelligence and cognition need to be undergirded by coherent, testable theories. As we have recounted, the success of intelligence scales may be due more to their practical applied value than the soundness of their theories, but progress as a scientific discipline requires that we improve theory construction and theory testing.



### References

- Adams, K. M. (1980a). An end of innocence for behavioral neurology? Adams replies. *Journal of Consulting and Clinical Psychology, 48*, 522-524.
- Adams, K. M. (1980b). In search of Luria's battery: A false start. *Journal of Consulting and Clinical Psychology, 45*, 511-516.
- Adams, K. M. (1984). Luria left in the lurch: Unfulfilled promises are not valid tests. *Journal of Clinical Neuropsychology, 6*(4), 455-458.
- Baars, B. J. (1986). *The cognitive revolution in psychology*. New York: Guilford.
- Babcock, H. (1930). An experiment in the measurement of mental deterioration. *Archives of Psychology, 18, No. 117*, 1-105.
- Bartholomew, D. J., Deary, I. J., & Lawn, M. (2009). A new lease of life for Thomson's bonds model of intelligence. *Psychological Review, 116*(3), 567-579.
- Bell, J. C. (1917). Review of Yerkes, Bridges, and Hardwick "A Point Scale for Measuring Mental Ability." *Journal of Educational Psychology, 8*(1), 45-47.
- Binet, A. (1904). Spearman: The proof and measurement of association between two things; General intelligence objectively determined and measured [Review]. *L'Année Psychologique, 11*, 623-624.
- Binet, A. (1909a). Burt: Experimental tests of general intelligence [Review]. *L'Année Psychologique, 16*, 459-460.
- Binet, A. (1909b). *Les idées modernes sur les enfants*. Paris : Ernest Flammarion.
- Binet, A. (1911). Avant-propos : Le bilan de la psychologie en 1910. *L'Année Psychologique, 17*, 5-11.

- Binet, A. (1975). *Modern ideas about children*. (S. Heisler, Trans.). Menlo Park, CA: Suzanne Heisler, Publisher. (Original work published 1909)
- Binet, A., & Henri, V. (1895). La psychologie individuelle. *L'Année Psychologique*, 2, 411-465.
- Binet, A., & Simon, T. (1907). *Les enfants anormaux : Guide pour l'admission des enfants anormaux dans les classes de perfectionnement*. Paris : Librairie Armand Colin.
- Binet, A., & Simon, T. (1910). Sur la nécessité d'une méthode applicable au diagnostic des arriérés militaires. *Annales Médico-Psychologiques*, 11, 9<sup>e</sup> série, 123-127.
- Binet, A., & Simon, T. (1914). *Mentally defective children* (W. B. Drummond, Trans.). New York: Longmans, Green & Co. (Original work published in 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. *L'Année Psychologique*, 11, 245-336. In H. H. Goddard (Ed.), *The development of intelligence in children (The Binet-Simon Scale)* (E. S. Kite, Trans.; pp. 91-181). Baltimore, MD: Williams & Wilkins. (Original work published 1905; Translation published 1916)
- Binet, A., & Simon, T. (1916b). New investigation upon the measure of the intellectual level among school children. *L'Année Psychologique*, 1911, 145-201. In H. H. Goddard (Ed.), *The development of intelligence in children (The Binet-Simon Scale)* (E. S. Kite, Trans.; pp. 274-329). Baltimore, MD: Williams & Wilkins. (Original work published 1911; Translation published 1916)
- Binet, A., & Simon, T. (1916c). New methods for the diagnosis of the intellectual level of subnormals. *L'Année Psychologique*, 11, 191-244. In H. H. Goddard (Ed.), *The development of intelligence in children (The Binet-Simon Scale)* (E. S. Kite, Trans.; pp.

- 37-90). Baltimore, MD: Williams & Wilkins. (Original work published 1905; Translation published 1916)
- Binet, A., & Simon, T. (1916d). The development of intelligence in the child. In H. H. Goddard (Ed.) & E. S. Kite (Trans.), *The development of intelligence in children* (pp. 182-273) Baltimore, MD: Williams & Wilkins. Originally published in *L'Année Psychologique*, 1908, 1-90.
- Binet, A., & Simon, T. (1916e). The intelligence of the feeble-minded (*L'Année Psychologique*, 1909, pp. 1-147). In H. H. Goddard (Ed.), *The intelligence of the feeble-minded* (E. S. Kite, Trans., pp. 8-155). Baltimore, MD: Williams & Wilkins. (Original work published 1909; Translation published 1916)
- Boake, C. (2002). From the Binet-Simon to the Wechsler-Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology*, 24(3), 383-405.
- Bolton, T. L. (1892). The growth of memory in school children. *The American Journal of Psychology*, 4, 362-380.
- Boring, E. G. (1923, June 6). Intelligence as the tests test it. *The New Republic*, 35(444), 35-37.
- Brown, A. W., et al. (1935). Report of committee of clinical section of American Psychological Association: I. The definition of clinical psychology and standards of training for clinical psychologists. II. Guide to psychological clinics in the United States. *The Psychological Clinic*, 23, 1-140.
- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 3(3), 296-322.

- Brown, W. (1932). The mathematical and experimental evidence for the existence of a central intellectual factor (*g*). *British Journal of Psychology*, 23(2), 171-179.
- Brown, W., & Thomson, G. H. (1921). *The essentials of mental measurement*. Cambridge, UK: Cambridge University Press.
- Buckhalt, J. A. (1991). Reaction time and intelligence: Contemporary research supports Cattell's abandoned hypothesis. *School Psychology International*, 12, 355-360.
- Buros, O. K. (1977). Fifty years in testing: Some reminiscences, criticisms, and suggestions. *Educational Researcher*, 6(7), 9-15.
- Burt, C. (1914). The measurement of intelligence by the Binet tests: Part II. *Eugenics Review*, 6(2), 140-152.
- Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice*, 31(2), 141-154.
- Carr, H. W., Wolf, A., & Spearman, C. (1925). Symposium: The nature of intelligence. *Proceedings of the Aristotelian Society Supplementary Volumes*, 5, 1-27.
- 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, D.C.: National Academy of Education.
- Carroll, J. B. (1980). Remarks on Sternberg's "Factor theories of intelligence are all right almost." *Educational Researcher*, 9(8), 14-18.
- 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. (1984). Raymond B. Cattell's contributions to the theory of cognitive abilities.

*Multivariate Behavioral Research*, 19, 300-306.

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, New Jersey: Ablex Publishing Corporation.

Carroll, J. B. (1998). Human cognitive abilities: A critique. In J. J. McArdle & R. W. Woodcock

(Eds.), *Human cognitive abilities in theory and practice* (pp. 5-23). Mahwah, NJ:

Lawrence Erlbaum Associates.

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, MA: Pergamon.

Carson, J. (2007). *The measure of merit: Talents, intelligence, and inequality in the French and*

*American republics, 1750-1940*. Princeton, NJ: Princeton University Press.

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

Cattell, J. M. (1893). Mental measurement. *The Philosophical Review*, 2(3), 316-332.

Cattell, J. M. (1923). The Psychological Corporation. *Annals of the American Academy of*

*Political and Social Science*, 110, 165-171.

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

Cattell, J. M., & Farrand, L. (1896). Physical and mental measurements of the students of

Columbia University. *Psychological Review*, 3, 618-648.

- Cattell, R. B. (1940). A culture-free intelligence test. I. *Journal of Educational Psychology*, 31(3), 161-179.
- Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing [Abstract]. *Psychological Bulletin*, 38, 592.
- Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, 40(3), 153-193.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1), 1-22.
- Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.
- Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. New York: North-Holland.
- Cattell, R. B. (1998a) Discussions of Human Cognitive Abilities conference papers. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 297-321). Mahwah, NJ: Lawrence Erlbaum.
- Cattell, R. B. (1998b). 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: Lawrence Erlbaum Associates.
- Cattell, R. B. (Ed.). (1987). *Intelligence: Its structure, growth and action*. New York: Elsevier Science.
- Cattell, R. B., & Butcher, H. J. (1968). *The prediction of achievement and creativity*. Oxford, England: Bobbs-Merrill.
- Cattell, R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement*, 15, 139-164.

- Cattell, R. B., Feingold, S. N., & Sarason, S. B. (1941). A culture-free intelligence test: II. Evaluation of cultural influence on test performance. *Journal of Educational Psychology*, 32(2), 81-100.
- Cohen, J. (1952a). A factor-analytically based rationale for the Wechsler-Bellevue. *Journal of Consulting Psychology*, 16(4), 272-277.
- Cohen, J. (1952b). Factors underlying Wechsler-Bellevue performance of three neuropsychiatric groups. *Journal of Abnormal and Social Psychology*, 47(2, Suppl.), 359-365.
- Cohen, J. (1957a). A factor-analytically based rationale for the Wechsler Adult Intelligence Scale. *Journal of Consulting Psychology*, 21(6), 451-457.
- Cohen, J. (1957b). The factorial structure of the WAIS between early adulthood and old age. *Journal of Consulting Psychology*, 21(4), 283-290.
- Cohen, J. (1959). The factorial structure of the WISC at ages 7-6, 10-6, and 13-6. *Journal of Consulting Psychology*, 23(4), 285-299.
- Cronbach, L. J. (1958). [Review] *The Measurement and Appraisal of Adult Intelligence* (4<sup>th</sup> ed.) by David Wechsler. *Science, New Series*, 128(3332), 1133.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Crosson, B., & Warren, R. L. (1982). Use of the Luria-Nebraska Neuropsychological Battery in aphasia: A conceptual critique. *Journal of Consulting and Clinical Psychology*, 50(1), 22-31.
- Cudeck, R., & MacCallum, R. C. (Eds.) (2007). *Factor analysis at 100: Historical developments and future directions*. Mahwah, NJ: Lawrence Erlbaum.

- Darwin, C. (1899). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: John Murray.
- Das, J. P., Kirby, J. R., & Jarman, R. F. (1975). Simultaneous and successive synthesis: An alternative model for cognitive abilities. *Psychological Bulletin*, 82, 87-103.
- Das, J. P., Naglieri, J. A., & Kirby, J. R. (1994). *Assessment of cognitive processes*. New York: Allyn & Bacon.
- Deary, I. J. (1994). Sensory discrimination and intelligence: Postmortem or resurrection? *American Journal of Psychology*, 107, 95-115.
- Deary, I. J. (2012). 125 years of intelligence in The American Journal of Psychology. *American Journal of Psychology*, 125(2), 145-154.
- 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.
- Delis, D. C., & Kaplan, E. (1982). The assessment of aphasia with the Luria-Nebraska Neuropsychological Battery: A case critique. *Journal of Consulting and Clinical Psychology*, 50(1), 32-39.
- Delis, D. C., & Kaplan, E. (1983). Hazards of a standardized neuropsychological test with low content validity: Comment on the Luria-Nebraska Neuropsychological Battery. *Journal of Consulting and Clinical Psychology*, 51(3), 396-398.
- Ebbinghaus, H. (1911). *Psychology: An elementary text-book* (M. Meyer, Trans.). Boston, MA: D. C. Heath & Co. Original German edition published in 1908.
- Fancher, R. E. (1989). Galton on examinations: An unpublished step in the invention of correlation. *Isis*, 80, 446-455.



- Feyerabend, P. (1975). *Against method: Outline of an anarchistic theory of knowledge*. London: Verso.
- Galton, F. (1865). Hereditary talent and character. *Macmillan's Magazine*, 12, 157-166, 318-327.
- Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences*. London: Macmillan.
- Galton, F. (1874). Proposal to apply for anthropological statistics from schools. *Journal of the Anthropological Institute*, 3, 308-311.
- Galton, F. (1879a). Generic images. *Nineteenth Century*, 6, 157-169.
- Galton, F. (1879b). Psychometric experiments. *Brain: A Journal of Neurology*, 2, 149-162.
- Galton, F. (1879c). Psychometric facts. *Nineteenth Century*, 5, 425-433.
- Galton, F. (1880). Mental imagery. *Fortnightly Review*, 28, 312-324.
- Galton, F. (1882). The anthropometric laboratory. *Fortnightly Review*, 31, 332-338.
- Galton, F. (1883a). *Inquiries into human faculty and its development*. London: Macmillan.
- Galton, F. (1883b). On apparatus for testing the delicacy of the muscular and other senses in different persons. *Journal of the Anthropological Institute*, 12, 469-477.
- Galton, F. (1883c). *Outfit for an anthropometric laboratory*. London: Author. Retrieved from <http://www.galton.org/>
- Galton, F. (1884a). *Anthropometric Laboratory* [pamphlet]. London: William Clowes and Sons.
- Galton, F. (1884b). Measurement of character. *Fortnightly Review*, 36, 179-185.
- Galton, F. (1884c). Notes. "Rede" Lecture: The measurement of human faculty. *Nature*, 30, 129-131.
- Galton, F. (1885a). Anthropometric percentiles. *Nature*, 31, 223-225.

- Galton, F. (1885b). On the Anthropometric Laboratory at the late International Health Exhibition. *Journal of the Anthropological Institute of Great Britain and Ireland*, 14, 205-219.
- Galton, F. (1885c). Some results of the Anthropometric Laboratory. *Journal of the Anthropological Institute*, 14, 275-287.
- Galton, F. (1886). Supplementary notes on "prehension" in idiots. *Mind*, 12, 79-82.
- Galton, F. (1887). *A descriptive list of anthropometric apparatus . . . designed under the direction of Mr. Francis Galton*. Cambridge, England: Cambridge Scientific Instrument Company.
- Galton, F. (1889a). An instrument for measuring reaction time. *Report of the British Association for the Advancement of Science*, 59, 784-785.
- Galton, F. (1889b). *Natural inheritance*. London: Macmillan and Co.
- Galton, F. (1890a). Cambridge anthropometry. *Nature*, 41(1063), 454.
- Galton, F. (1890b). Exhibition of instruments (1) for testing perception of differences of tint, and (2) for determining reaction-time. *Journal of the Anthropological Institute*, 19(1), 27-29.
- Galton, F. (1890c). Remarks [after J. M. Cattell's "Mental tests and measurements"]. *Mind*, 15, 380-381.
- 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. (1908). *Memories of my life* (2<sup>nd</sup> ed.). London: Methuen & Co.
- Galton, F. (1910). Eugenic qualities of primary importance. *Eugenics Review*, 1, 74-76.

- Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.
- Gardner, H., Feldman, D., & Krechevsky, M. (1998a). *Project Spectrum—Building on children's strengths: The experience of Project Spectrum*. New York: Teachers College Press.
- Gardner, H., Feldman, D., & Krechevsky, M. (1998b). *Project Spectrum: Early learning activities*. New York: Teachers College Press.
- Gardner, H., Feldman, D., & Krechevsky, M. (1998c). *Project Spectrum: Preschool assessment handbook*. New York: Teachers College Press.
- Gates, A. I. (1971). An autobiography. In R. J. Havighurst (Ed.), *Leaders in American education. Seventieth yearbook of the National Society for the Study of Education, Part II* (pp. 189-221). Chicago, IL: NSSE.
- Geschwind, N. (1970). The organization of language and the brain. *Science, New Series*, 170(3961), 940-944.
- 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). Introduction. In H. H. Goddard (Ed.), *The development of intelligence in children (The Binet-Simon Scale)* (pp. 5-8). Baltimore, MD: Williams & Wilkins.
- Golden, C. J., Purisch, A. D., & Hammeke, T. A. (1979). *The Luria-Nebraska Neuropsychological Battery. A manual for clinical and experimental uses*. Lincoln, NE: University of Nebraska Press.
- Goodenough, F. L. (1949). *Mental testing: Its history, principles and applications*. New York: Rinehart and Company.

- Gottfredson, L. J. (1997a). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, 24, 13-23.
- Gottfredson, L. S. (1997b). Why g matters: The complexity of everyday life. *Intelligence*, 24(1), 79-132.
- Gottfredson, L. S. (2002a). g: Highly general and highly practical. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 331-380). Mahwah, NJ: Erlbaum.
- Gottfredson, L. S. (2002b). Where and why g matters: Not a mystery. *Human Performance*, 15(1/2), 25-46.
- Gray, J. (2007). *Straw dogs: Thoughts on humans and other animals*. New York: Farrar, Straus and Giroux.
- Gustafsson, J. -E. (1984). A unifying model of the structure of intellectual abilities. *Intelligence*, 8, 179-203.
- Gustafsson, J. -E. (1988). Hierarchical models for individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4; pp. 35-71). Hillsdale, NJ: Erlbaum.
- Hamid, S. A. (1925). Some factors of effectiveness in mental ("Intelligence") tests. *British Journal of Psychology*, 16(2), 100-115.
- Hardie, J. L. (1928). Spearman's Measure of 'Intelligence': A statistical analysis. *British Journal of Psychology*, 19(2), 188-197.
- Hart, B., & Spearman, C. (1912). General ability, its existence and nature. *British Journal of Psychology*, 5, 51-84.
- Hearnshaw, L. S. (1964). *A short history of British psychology, 1840-1940*. London: Methuen &

Co.

- Hildreth, G. H. (1933). *A bibliography of mental tests and rating scales*. New York: The Psychological Corporation.
- Holzinger, K. J. (1936). Recent research on unitary mental traits. *Journal of Personality*, 4(4), 335-343.
- Holzinger, K. J. (1946). Spearman as I knew him. *Psychometrika*, 10(4), 231-235.
- Holzinger, K. J., & Swineford, F. (1939). A study in factor analysis: the stability of a bi-factor solution. *Supplementary Educational Monographs*, No. 48. Chicago, IL: University of Chicago Department of Education.
- Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 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. (1985). Remodeling old models of intelligence. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurements and applications* (pp. 267-300). New York: John Wiley.
- Horn, J. L., & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2<sup>nd</sup> ed., pp. 41-68). New York: Guilford.
- 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., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 26, 107-129.
- Horn, J. L., & Stankov, L. (1982). Auditory and visual factors of intelligence. *Intelligence*, 6, 165-185.
- Hunt, E. (1980). Intelligence as an information-processing concept. *British Journal of Psychology*, 71(4), 449-474.
- Jastrow, J. (1893). The section of psychology. In M. P. Hardy (Ed.), *Official Catalogue -- World's Columbian Exposition* (Part. Vii; pp. 50-60). Chicago: W. B. Conkey.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Jensen, A. R. (2006). *Clocking the mind: Mental chronometer individual differences*. Amsterdam, Netherlands: Elsevier.
- Kamphaus, R. W., & Reynolds, C. R. (2009). Kaufman's work in the penumbra between measurement science and clinical assessment. In J. C. Kaufman (Ed.), *Intelligence testing: Integrating psychological theory and clinical practice* (pp. 148-156). New York: Cambridge University Press.
- Kaplan, E. (1988). A process approach to neuropsychological assessment. In T. Boll & B. K. Bryant (Eds.), *Clinical neuropsychology and brain function: Research, measurement, and practice* (pp. 129-167). Washington, DC: American Psychological Association.
- Kaplan, E. (1990). The process approach to neuropsychological assessment of psychiatric patients. *Journal of Neuropsychiatry and Clinical Neurosciences*, 2(1), 72-87.
- Kaplan, E., Fein, D., Morris, R., Kramer, J. H., & Delis, D. C. (1991). *WAIS-R NI: WAIS-R as a Neuropsychological Instrument*. San Antonio, TX: The Psychological Corporation.
- Kaufman, A. S. (1979). *Intelligent testing with the WISC-R*. New York: John Wiley & Sons.

Kaufman, A. S. (1992). Dr. Wechsler remembered. *The School Psychologist*, 46(2), 4-5, 17.

Reprinted in Lichtenberger, E. O., & Kaufman, A. S. (2009). *Essentials of WAIS-IV assessment* (pp. xiii-xx). New York: John Wiley.

Kaufman, A. S. (2009). *IQ testing 101*. New York : Springer.

Kaufman, A. S. (2013). Intelligent testing with Wechsler's fourth editions: Perspectives on the Weiss et al. studies and the eight commentaries. *Journal of Psychoeducational Assessment*, 31(2), 224-234.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children*. Circle Pines, MN: American Guidance.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children, Second Edition (KABC-II): Manual*. Bloomington, MN: Pearson.

Kaufman, A. S., & Lichtenberger, E. O. (2006). *Assessing adolescent and adult intelligence* (3<sup>rd</sup> ed.). Hoboken, NJ: John Wiley.

Keith, T. Z., & Kranzler, J. H. (1999). The absence of structural fidelity precludes construct validity: Rejoinder to Naglieri on what the Cognitive Assessment System does and does not measure. *School Psychology Review*, 28(2), 303-321.

Kohs, S. C. (1914a). The Binet-Simon measuring scale for intelligence: An annotated bibliography. Part I. *Journal of Educational Psychology*, 5(4), 215-224.

Kohs, S. C. (1914b). The Binet-Simon measuring scale for intelligence: An annotated bibliography. Part II. *Journal of Educational Psychology*, 5(5), 279-290.

Kohs, S. C. (1914c). The Binet-Simon measuring scale for intelligence: An annotated bibliography. Part III. *Journal of Educational Psychology*, 5(6), 335-346.

- Kohs, S. C. (1917a). An annotated bibliography of recent literature on the Binet-Simon Scale (1913-1917). Part I. *Journal of Educational Psychology*, 8(7), 425-438.
- Kohs, S. C. (1917b). An annotated bibliography of recent literature on the Binet - Simon Scale (1913 - 1917). Part II. *Journal of Educational Psychology*, 8(8), 488-502.
- Kohs, S. C. (1917c). An annotated bibliography of recent literature on the Binet-Simon Scale (1913-1917).Part III. *Journal of Educational Psychology*, 8(9), 559-565.
- Kohs, S. C. (1917d). An annotated bibliography of recent literature on the Binet-Simon Scale (1913-1917) Part IV. *Journal of Educational Psychology*, 8(10), 609-618.
- Korkman, M. (1999). Applying Luria's diagnostic principles in the neuropsychological assessment of children. *Neuropsychology Review*, 9(2), 89-105.
- Korkman, M., Kirk, U., & Kemp, S. L. (1998). *NEPSY: A Developmental Neuropsychological Assessment*. San Antonio, TX: The Psychological Corporation.
- Kuhn, T. S. (1977). Objectivity, value judgment, and theory choice. In T. S. Kuhn, *The essential tension: Selected studies in scientific tradition and change* (pp. 320-339). Chicago, IL: University of Chicago Press.
- Kuhn, T. S. (2000). *The road since Structure* (J. Conant & J. Haugeland, Eds.). Chicago, IL: University of Chicago Press.
- Kuhn, T. S. (2012). *The structure of scientific revolutions* (4<sup>th</sup> ed.). Chicago, IL: University of Chicago Press. (Original work published 1962)
- Lautrey, J., & de Ribaupierre, A. (2004). Psychology of human intelligence in France and French-speaking Switzerland. In R. J. Sternberg (Ed.), *International handbook of intelligence* (104-134). New York: Cambridge University Press.



- Leahey, T. H. (1992). The mythical revolutions of American psychology. *American Psychologist*, 47(2), 308-318.
- Louttit, C. M., & Browne, C. G. (1947). The use of psychometric instruments in psychological clinics. *Journal of Consulting Psychology*, 11(1), 49-54.
- Lovie, P., & Lovie, A. D. (1995). The cold equations: Spearman and Wilson on factor indeterminacy. *British Journal of Mathematical and Statistical Psychology*, 48(2), 237-253.
- Lovie, P., & Lovie, A. D. (1996). Charles Edward Spearman, F.R.S. (1863-1945). *Notes and Records of the Royal Society of London*, 50(1), 75-88.
- Lovie, S., & Lovie, P. (2010). Commentary: Charles Spearman and correlation: A commentary on 'The proof and measurement of association between two things.' *International Journal of Epidemiology*, 39, 1151-1153.
- Lubin, B., Larsen, R.M., & Matarazzo, J.D. (1984). Patterns of psychological test usage in the United States: 1935-1982. *American Psychologist*, 39, 451-454.
- Lubin, B., Wallis, R., & Paine, C. (1971). Patterns of psychological test usage in the United States: 1935-1969. *Professional Psychology*, 2, 70-74.
- Luria, A. R. (1966). *Human brain and psychological processes*. New York: Harper & Row.
- Luria, A. R. (1970a). The functional organization of the brain. *Scientific American*, 222(3), 66-78.
- Luria, A. R. (1970b). *Traumatic aphasia: Its syndromes, psychology and treatment* (D. Bowden, Transl). The Hague, The Netherlands: Mouton & Co.
- Luria, A. R. (1973). *The working brain: An introduction to neuropsychology* (B. Haigh, Trans.). New York: Basic Books.

- Luria, A. R. (1976). *The neuropsychology of memory* (B. Haigh, Trans.). Washington, D.C.: V. H. Winston & Sons.
- Luria, A. R. (1980). *Higher cortical functions in man* (2<sup>nd</sup> ed.; B. Haigh, Trans.), New York: Basic Books.
- Luria, A. R., & Majovski, L. V. (1977). Basic approaches used in American and Soviet clinical neuropsychology. *American Psychologist*, 32, 959-968.
- Macmillan, M. (1991). *Freud evaluated: The completed arc*. Amsterdam, The Netherlands: Elsevier Science.
- Matarazzo, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5<sup>th</sup> ed.). New York: Oxford University Press.
- 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. (1964). Lost: Our intelligence? Why? *American Psychologist*, 19(12), 871-882.
- Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, 1(2), 108-141, 173-180.
- Meeker, M. N., Meeker R. J., & Roid. G. H. (1985). *Structure of Intellect Learning Abilities Test (SOI-LA) manual*. Los Angeles, CA: Western Psychological Services.
- Meredith, G. P. (1948-1949). A revision of Spearman's noëgenetic principles. *Proceedings of the Aristotelian Society, New Series*, 49, 91-114.
- Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, 7(3), 141-144.
- Milner, B. (1958). Psychological defects produced by temporal lobe excision. *Research Publications -- Association for Research in Nervous and Mental Disease*, 36, 244-257.

- Minton, H. L., & Schneider, F. W. (1980). *Differential psychology*. Prospect Heights, IL: Waveland Press.
- Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System*. Itasca, IL : Riverside Publishing.
- National Academy of Sciences. (1999). *Science and creationism: A view from the National Academy of Sciences (2<sup>nd</sup> ed.)*. Washington, DC: National Academy Press.
- Neisser, U. (1976). *Cognition and reality*. San Francisco, CA: W. H. Freeman.
- Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*, 77-101.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- 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., & Levine, Z. (2012). Beyond intelligence testing: Remembering Alfred Binet after a century. *European Psychologist*, *17*, 320-325
- Nicolas, S., & Sanitioso, R. B. (2012). Alfred Binet and experimental psychology at the Sorbonne laboratory. *History of Psychology*, *15*, 328-363.
- Norton, B. (1979). Charles Spearman and the general factor in intelligence: Genesis and interpretation in the light of sociopersonal considerations. *Journal of the History of the Behavioral Sciences*, *15* (2), 142-154.
- Otis, A. S. (1917). A criticism of the Yerkes-Bridges Point Scale, with alternative suggestions. *Journal of Educational Psychology*, *8*, 129-150.

- Pearson, K. (1927). The abilities of man, their nature and measurement, by Prof. C. Spearman, *Nature*, 120(3014), 181-183.
- Pearson, K., & Moul, M. (1927). The mathematics of intelligence. I. The sampling errors in the theory of a generalised factor. *Biometrika*, 19(3/4), 246-291.
- Penrose, L. S., & Raven, J. C. (1936). A new series of perceptual tests: Preliminary communication. *British Journal of Medical Psychology*, 16, 97-105.
- Peterson, J. (1926). *Early conceptions and tests of intelligence*. Yonkers-on-Hudson, New York: World Book Company.
- Pintner, R., & Marshall, H. (1921). A combined mental-educational survey. *Journal of Educational Psychology*, 12(1), 32-43.
- Popper, K. (2005). *The logic of scientific discovery*. New York: Routledge. (Original work published 1935)
- Rabin, A. I. (1945). The use of the Wechsler-Bellevue scales with normal and abnormal persons. *Psychological Bulletin*, 42(7), 410-422.
- Rabin, A. I., & Guertin, W. H. (1951). Research with the Wechsler-Bellevue test: 1945-1950. *Psychological Bulletin*, 48, 211-248.
- Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual. Section 3: Standard Progressive Matrices (1998 edition). Introducing the Parallel and Plus versions*. Oxford: Oxford Psychologists Press.
- Raven, J. C. (1936). *Mental tests used in genetic studies: The performances of related individuals in tests mainly educative and mainly reproductive*. Unpublished master's thesis. University of London.

- Raven, J. C. (1938). *Progressive Matrices: A perceptual test of intelligence, 1938, Individual form*. London: H. K. Lewis.
- Raven, J. C. (1948). The comparative assessment of intellectual ability. *British Journal of Psychology, 39(1)*, 12-19.
- Raven, J. C., & Walshaw, J. B. (1944). Vocabulary tests. *British Journal of Medical Psychology, 20*, 185-194.
- 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. (1985). Limitations of the test reviewing process: A response to Clarizio and Mehrens. *Gifted Child Quarterly, 29*, 121-123.
- Russell, E. W. (1980, August). *Theoretical bases of Luria-Nebraska and Halstead-Reitan batteries*. Paper presented at the meeting of the American Psychological Association, Montreal, Canada.
- Sharp, S. E. (1899). Individual psychology: A study in psychological method. *American Journal of Psychology, 10*, 329-391.
- Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology: Interdisciplinary and Applied, 9(2)*, 371-377.
- Shipley, W. C., Gruber, C. P., Martin, T. A., & Klein, A. M. (2009). *Shipley-2 manual*. Los Angeles, CA: Western Psychological Services.
- Siegler, R. S. (1992). The other Alfred Binet. *Developmental Psychology, 28(2)*, 179-190.

- Slater, P. (1948). Comment on "The comparative assessment of intellectual ability." *British Journal of Psychology*, 39(1), 20-21.
- Sokal, A., & Bricmont, J. (1998). *Fashionable nonsense: Postmodern intellectuals' abuse of science*. New York: Picador.
- Sokal, M. M. (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. (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: Lawrence Erlbaum Associates.
- Sokal, M. M. (2010). Scientific biography, cognitive deficits, and laboratory practice: James McKeen Cattell and early American experimental psychology, 1880-1904. *Isis*, 101, 531-554.
- Solso, R. L., & Hoffman, C. A. (1991). Influence of Soviet scholars. *American Psychologist*, 46, 251-253.
- Spearman, C. (1904a). "General intelligence," objectively determined and measured. *American Journal of Psychology*, 15 (2), 201-292.
- Spearman, C. (1904b). The proof and measurement of association between two things. *American Journal of Psychology*, 15(1), 72-101.
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3(3), 271-295.
- Spearman, C. (1915). The measurement of intelligence, *Eugenics Review*, 6(4), 312-313.
- Spearman, C. (1916). Some comments on Mr. Thomson's paper. *British Journal of Psychology*, 8, 282-284.

- Spearman, C. (1922). A friendly challenge to Professor Thorndike. *Psychological Review*, 29(5), 406-407.
- Spearman, C. (1923). *The nature of intelligence and the principles of cognition*. London: Macmillan.
- Spearman, C. (1924). A challenge still open. *Journal of Educational Psychology*, 15(6), 393.
- Spearman, C. (1925). *A measure of "intelligence" for use in schools*. London: Methuen.
- Spearman, C. (1927). *The abilities of man: Their nature and measurement*. London: Macmillan
- Spearman, C. (1930a). *Creative mind*. London: Cambridge University Press.
- Spearman, C. (1930b). C. Spearman [Autobiography]. In C. Murchison (Ed.), *A history of psychology in autobiography* (Vol. 1) (pp. 299-331). Worcester, MA: Clark University Press.
- Spearman, C. (1930c). "G" and after—A school to end schools. In C. Murchison (Ed.), *Psychologies of 1930* (pp. 339-366), Worcester, MA: Clark University Press.
- Spearman, C. (1933). *The Spearman Visual Perception Test*. Cited on page 167 in Cattell, R. B. (1940). A culture-free intelligence test: I. *Journal of Educational Psychology*, 31(3), 161-179).
- Spearman, C. (1937). *Psychology down the ages* (Vol. 1). London: Macmillan and Co.
- Spearman, C. (1939a). "Intelligence" tests. *Eugenics Review*, 30(4), 249-254.
- Spearman, C. (1939b). Thurstone's work re-worked. *Journal of Educational Psychology*, 30(1), 1-16.
- Spearman, C. (1946). Theory of general factor. *British Journal of Psychology*, 36, 117-131.
- Spearman, C., & Wynn Jones, LL. (1950). *Human ability: A continuation of "The Abilities of Man."* London: Macmillan & Co.

- Spencer, H. (1855). *The principles of psychology*. London: Longman, Brown, Green, and Longmans.
- Spiers, P. (1981). Have they come to praise Luria or to bury him? The Luria-Nebraska Battery controversy. *Journal of Consulting and Clinical Psychology*, 49, 331-341.
- Steiger, J. H. (1996). Coming full circle in the history of factor indeterminacy. *Multivariate Behavioral Research*, 31(4), 617-630.
- Stern, A. W. (1956). The nature of *g* and the concept of intelligence. *Acta psychologica*, 12, 282-289.
- Sundberg, N. D. (1961). The practice of psychological testing in clinical services in the United States. *American Psychologist*, 16, 79-83.
- 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, MA: Houghton Mifflin.
- Terman, L. M. (1932). Lewis M. Terman [Autobiography]. In C. Murchison (Ed.), *A history of psychology in autobiography* (Vol. 2, pp. 297-331). Worcester, MA: Clark University Press.
- Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence: A guide to the administration of the new revised Stanford-Binet tests of intelligence*. Boston, MA: Houghton Mifflin.
- Thomson, G. H. (1916). A hierarchy without a general factor. *British Journal of Psychology*, 8, 271-281.
- Thomson, G. H. (1919a). On the cause of hierarchical order among the correlation coefficients of a number of variates taken in pairs. *Proceedings of the Royal Society of London*, 95 A, 400-408.



- Thomson, G. H. (1919b). The proof or disproof of the existence of general ability. *British Journal of Psychology*, 9, 321-336.
- Thomson, G. H. (1920). The general factor fallacy in psychology. *British Journal of Psychology*, 10, 319-326.
- Thomson, G. H. (1939-1940). The factorial analysis of ability: Agreement and disagreement in factor analysis: A summing up. *British Journal of Psychology*, 30, 105-108.
- Thomson, G. H. (1947). Charles Spearman. 1863-1945. *Obituary Notices of Fellows of the Royal Society*, 5(15), 373-385.
- Thorndike, E. L. (1901). The evolution of the human intellect. *Popular Science Monthly*, 60(5), 58-65.
- Thorndike, E. L. (1909). A note on the specialization of mental functions with varying content. *Journal of Philosophy, Psychology, and Scientific Methods*, 6(9), 239-240.
- Thorndike, E. L. (1911). *Individuality*. Cambridge, MA: Houghton Mifflin.
- Thorndike, E. L. (1920). Intelligence and its use. *Harper's Magazine*, 140, 227-235.
- Thorndike, E. L. (1924). "A Challenge Still Open": Reply. *Journal of Educational Psychology*, 15(6), 394-395.
- 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. (1984). *Intelligence as information processing: The mind and the computer* [A CEDR monograph]. Bloomington, IN: Phi Delta Kappa.
- Thorndike, R. L. (1994). *g*. [Editorial]. *Intelligence*, 19, 145-155.
- Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet Intelligence Scale: Fourth Edition*. Itasca, IL: Riverside Publishing.

- Thorndike-Christ, T. (2010). Review of the Shipley-2. In R. A. Spies, J. F. Carlson, & K. F. Geisinger (Eds.), *The eighteenth mental measurements yearbook*. Available from <http://buros.unl.edu/buros/>
- Thurstone, L. L. (1937). Psychology as a quantitative rational science. *Science*, 85, 227-232.
- Thurstone, L. L. (1938). *Primary mental abilities* (Psychometric Monographs, No. 1). Chicago, IL: University of Chicago Press.
- Thurstone, L. L. (1939). *Primary mental abilities*. Chicago, IL: University of Chicago Press.
- Thurstone, L. L. (1946). Theories of intelligence. *The Scientific Monthly*, 62(2), 101-112.
- Thurstone, L. L. (1952). L. L. Thurstone. In E. G. Boring, H. Werner, H. S. Langfeld, & R. M. Yerkes (Eds.), *A history of psychology in autobiography* (Vol. 4, pp. 295-321). Worcester, MA: Clark University Press.
- Toulouse, E., & Piéron, H. (1911). *Technique de psychologie expérimentale* (2<sup>nd</sup> ed.). Paris : Octave Doin.
- Toulouse, E., Vaschide, N., & Piéron, H. (1904). *Technique de psychologie expérimentale (examen des sujets)*. Paris : Octave Doin.
- Tranel, D. (2008). Theories of clinical neuropsychology and brain-behavior relationships: Luria and beyond. In J. E. Morgan & J. H. Richter (Eds.), *Textbook of clinical neuropsychology* (pp. 25-37). New York: Taylor & Francis.
- Undheim, J. O. (1981a). On intelligence: II. A neo-Spearman model to replace Cattell's theory of fluid and crystallized intelligence. *Scandinavian Journal of Psychology*, 22, 181-187.
- Undheim, J. O. (1981b). On intelligence: IV. Toward a restoration of general intelligence. *Scandinavian Journal of Psychology*, 22, 251-265.
- Undheim, J. O., & Gustafsson, J. -E. (1987). The hierarchical organization of cognitive abilities:

- Restoring general intelligence through the use of linear structural relations (LISREL). *Multivariate Behavioral Research*, 22(2), 149-171.
- Varon, E. (1936). Alfred Binet's concept of intelligence. *Psychological Review*, 43, 32-58.
- Venn, J. (1889). Cambridge anthropometry. *Journal of the Anthropological Institute of Great Britain and Ireland*, 18, 140-154.
- Venn, J. (1890). Cambridge anthropometry. *Nature*, 41(1063), 450-454.
- Walters, E. H., & Thomas, F. C. (1929). Some notes on the standardisation of Professor Spearman's "Measure of 'Intelligence' for Use in Schools." *Forum of Education (Training College Association, Great Britain)*, 7, 35-42.
- Washburn, M. F. (1929). Energy, engines, and the engineer: A critique of C. Spearman. *American Journal of Psychology*, 41(2), 322-326.
- Wasserman, J. D. (2012). A history of intelligence assessment: The unfinished tapestry. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues (3<sup>rd</sup> ed.)* (pp. 3-55), New York: Guilford Press.
- Wasserman, J. D., & Kaufman, A. S. (in press). David Wechsler (1896-1981). In R. Cautin & S. Lilienfeld (Eds.), *Encyclopedia of clinical psychology*. Wiley-Blackwell.
- Wechsler, D. (1926). Psychology as a practical science in modern life. *Current History*, 25, 358-364.
- Wechsler, D. (1927). Psychometric tests. In I. S. Wechsler, *Textbook of clinical neurology* (pp. 104-116). Philadelphia, PA: W. B. Saunders.
- Wechsler, D. (1929). General intelligence, mental level and the psychoneuroses. In I. S. Wechsler, *The neuroses* (pp. 289-319). Philadelphia, PA: W. B. Saunders Co.
- Wechsler, D. (1932). Analytic use of the Army Alpha examination. *Journal of Applied*

- Psychology*, 16, 254-256.
- Wechsler, D. (1935a). The concept of mental deficiency in theory and practice. *Psychiatric Quarterly*, 9, 232-236.
- Wechsler, D. (1935b). *The range of human capacities*. Baltimore, MD: Williams & Wilkins.
- Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore, MD: Williams & Wilkins.
- Wechsler, D. (1943). Non-intellective factors in general intelligence. *Journal of Abnormal and Social Psychology*, 38(1), 101-103.
- Wechsler, D. (1944). The psychologist in the psychiatric hospital. *Journal of Consulting Psychology*, 8(5), 281-285.
- Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology: Interdisciplinary and Applied*, 19(1), 87-95.
- Wechsler, D. (1949a). Review of Progressive Matrices: A perceptual test of intelligence, 1938. In O. K. Buros (Ed.), *The third mental measurements yearbook* (pp. 257-258). New Brunswick NJ: Rutgers University Press.
- Wechsler, D. (1949b). *Wechsler Intelligence Scale for Children (WISC)*. New York: The Psychological Corporation.
- Wechsler, D. (1950). Cognitive, conative, and non-intellective intelligence. *American Psychologist*, 5(3), 78-83.
- Wechsler, D. (1955). *Wechsler Adult Intelligence Scale—Restandardized (WAIS)*. New York: The Psychological Corporation.
- Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore, MD: Williams & Wilkins Co.

- Wechsler, D. (1966, June 26). The I.Q. is an intelligent test. *New York Times*, pp. SM12, 13, 63-66..
- Wechsler, D. (1967). *Wechsler Preschool and Primary Scale of Intelligence (WPPSI)*. New York: The Psychological Corporation.
- Wechsler, D. (1971). Intelligence: Definition, theory, and the IQ. In R. Cancro (Ed.), *Intelligence: Genetic and environmental influences* (pp. 50-55). New York: Grune & Stratton.
- Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children—Revised*. New York: The Psychological Corporation.
- Wechsler, D. (1975). Intelligence defined and undefined: A relativistic appraisal. *American Psychologist*, 30(2), 135-139.
- 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, D., Kaplan, E., Fein, D., Kramer, J., Morris, R., Delis, D., & Maerlender, A. (2004). *Wechsler Intelligence Scale for Children Fourth Edition—Integrated (WISC-IV Integrated) administration and scoring manual*. San Antonio, TX: Harcourt Assessment.
- Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013a). WAIS-IV and clinical validation of the four- and five-factor interpretative approaches. *Journal of Psychoeducational Assessment*, 31(2), 94-113.
- Wells, F. L. (1927). *Mental tests in clinical practice*. Yonkers, NY: World Book.

- Williams, R. H., Zimmerman, D. W., Ross, D. C., & Zumbo, B. D. (2006). *Twelve British statisticians*. Raleigh, NC: Boson Books.
- Wilson, E. B. (1928). The abilities of man, their nature and measurement [Review]. *Science, New Series*, 67(1731), 244-248.
- Wissler, C. (1901). The correlation of mental and physical tests. *Psychological Monographs*, 3(6), 1-62.
- Wolf, T. H. (1961). An individual who made a difference. *American Psychologist*, 16(5), 245-248.
- Wolf, T. H. (1964). Alfred Binet: A time of crisis. *American Psychologist*, 19(9), 762-771.
- Wolf, T. H. (1969). The emergence of Binet's conception and measurement of intelligence: A case history of the creative process. *Journal of the History of the Behavioral Sciences*, 5, 113-134.
- Wolf, T. H. (1973). *Alfred Binet*. Chicago, IL: University of Chicago Press.
- Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment*, 8, 231-258.
- Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery—Revised*. Itasca, IL: Riverside Publishing.
- Woodworth, R. S. (1912). Combining the results of several tests: A study in statistical method. *Psychological Review*, 19, 97-123.
- Yerkes, R. M. (Ed.). (1921). *Psychological examining in the United States Army. Memoirs of the National Academy of Sciences* (Vol. 15). Washington, DC: Government Printing Office.
- Zenderland, L. (1998). *Measuring minds: Henry Herbert Goddard and the origins of American intelligence testing*. New York: Cambridge University Press.

Zubin, J. (1933). Factor analysis of the Spearman Visual Perception Test [Abstract]. *The Psychological Bulletin*, 30(8), 568.