IQ Testing 101

Alan S. Kaufman, PhD

SPRINGER PUBLISHING COMPANY

3

History, Part 2: At Long Last—Theory Meets Practice

Theory was alive and well in psychological laboratories throughout the world from the early years of the 20th century (Spearman, 1904), and so were IQ tests (Binet & Simon, 1905). But apart from a few attempts to apply theory directly to the interpretation of IQ tests, or to the development of group-administered IQ tests (Meeker et al., 1975; Thurstone & Thurstone, 1949), the decade of the 1970s ended with theory failing to make a dent in the construction of a clinically based IQ test.

In the 1980s, as Guilford's (1967) SOI theory was going out of favor, two important things occurred. First, neurological theories of mental processing, notably Sperry's (1968) ideas about cerebral specialization and Luria's (1966, 1973) notions of successive and simultaneous processing, formed the basis of a clinical test of intelligence, the Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983); and (2) Horn's (1985) expansion of *Gf-Gc* theory to include additional broad abilities, such as short-term memory (*Gsm*), visualization (*Gv*), and processing speed (*Gs*)—and its subsequent merger with Carroll's (1993) model to form Cattell-Horn-Carroll (CHC) theory—zoomed in popularity and soon formed the solid foundation for several IQ tests, most notably the Woodcock-Johnson—Revised (WJ-R; Woodcock & Johnson, 1989) and its sequel, the WJ III (Woodcock, McGrew, & Mather, 2001).

These two innovative pathways would radically change the face of IQ test development and IQ test interpretation, starting in the early 1980s and continuing to the present day. But before following these two roads that led to the present-day breed of theory-based IQ tests, two topics must be mentioned.

First, a final look at Guilford's impact is warranted, as articulated by Linda Silverman (personal communication, July 8, 2008): "In spite of the abounding criticism, the Structure-of-Intellect model has had an enormous influence on modern conceptions of intelligence. Even the most severe critics (e.g., Carroll, 1968; Horn & Knapp, 1973; Humphreys, 1962) have indicated that the model has provided a stimulus to creative test development and has provoked considerable re-evaluation of the nature of human abilities."

Second, the two main theoretical approaches featured in this chapter—neuropsychological processing and CHC—are two among many theories of intelligence. I am well aware that some of the most popular and ingenious theories, notably Sternberg's (1988b, 1999) triarchic theory of successful intelligence and Gardner's (1993) multiple-intelligence theory, are far more comprehensive and encompass many more abilities than the handful of theories that have formed the foundation of modern IQ tests. As Sternberg (1988b) has said for years, IQ tests measure only one of the three prongs of his theory—*analytic* abilities, but not *practical* intelligence or *creativity.* (He is correct.) From the perspective of Gardner's eight multiple intelligences, IQ tests assess only three: *linguistic, logical-mathematical,* and *spatial* (Chen and Gardner, 2005, give credit only for the first two, but IQ tests have measured spatial intelligence for 70 years). No question, though, IQ tests do not measure Gardner's other five intelligences, many of which are noncognitive: *musical, bodily-kinesthetic, naturalistic, interpersonal,* and *intrapersonal* (i.e., self-insight).

Two decades ago, Sternberg (1988a) said, "Intelligence tests of the present are anachronisms" (p. 8), and Gardner (1988) said, "The whole concept [of IQ tests] has to be challenged; in fact, it has to be replaced" (p. 4). More recently, Sternberg has softened his stance: "[N]ew intelligence tests developed during the past twenty years (including the fifth edition of the Stanford-Binet) have been built from theories of intelligence. . . . Indeed, it would be hard for a new or revised test *not* based on theory to be competitive" (Sternberg et al., 2008, p. 12). Gardner's current perspective is still hard-line: "We recognize that some current intelligence tests do measure more than two cognitive abilities . . . and that Carroll's work . . . measures up to eight different intellectual components . . . These intelligence tests are based on 'horizontal' theories of intelligence" (Chen & Gardner, 2005, p. 81). Never mind what Gardner means by *horizontal.* Suffice it to say that his theory is *vertical,* and, therefore, much better—though not every IQ expert agrees. Lloyd Humphries said, "Gardner has debased the meaning of intelligence by grouping everything but the kitchen sink under that rubric" (Cordes, 1986, p. 8). For a thorough treatment of Sternberg's, Gardner's, and other influential theories of intelligence, broadly defined and not limited to the lens of IQ testing, consult Flanagan and Harrison's (2005) comprehensive edited text or Plucker's (in press) *Intelligence 101*.

My own view? The existing IQ tests, simply by following their own theoretical approach to what intelligence is, are immediately wrong or invalid from boxloads of other, sometimes opposite, viewpoints. Develop an IQ test from one theory, and you instantly alienate and incur the wrath of a multitude of ivory-tower researchers who preach the righteousness of their own Word without understanding the constraints of the clinical assessment of IQ. I don't dispute the obvious fact that IQ tests measure only an aspect of intelligence. Formal testing may not even be the best way to measure some other aspects. Gardner wants to do away with contemporary IQ tests because they measure only a few of his eight intelligences. But what rational person would try to claim than an IQ test can, or should, try to measure *all* that is intelligence? That IQ tests can't measure the totality of intelligence is axiomatic. David Wechsler knew that quite well. His theory of intelligence—the overall capacity to understand and cope with one's environment (Wechsler, 1958, 1975)—was far more comprehensive than his measurement of it. How much time can a psychologist reasonably spend giving an IQ test to an adult or child referred for clinical evaluation? The answer, Wechsler knew, is usually an hour-and-a-half or two, tops. That is long enough to understand a person's strong abilities and weak abilities, general level of mental functioning, and learning style—but not long enough to explore every crevice in the person's brain.

About a half century ago, psychologists named Pinard and Laurendeau (1964) developed an intelligence test based on Jean Piaget's well-respected developmental theory of intelligence (e.g., Inhelder & Piaget, 1958), and they even standardized it on a large sample of Canadian children. Trouble is, the authors were a bit compulsive and developed a test with 27 long tasks that took about 14 hours to administer. The term *experimental mortality* is used figuratively in psychology to indicate those subjects who drop out of a research study because of lack of interest, illness, moving away, and so forth. I have a feeling that experimental mortality might have taken on a more literal meaning during the norming of that all-encompassing Piaget test! (The test itself died a quiet death.) I also suspect that near-death experiences (NDEs according to Connie Willis in her brilliant sci-fi novel *Passage*), which are now associated with tunnels of light and blissful peace, may take on a new meaning if Gardner enters the realm of clinical assessment of IQ with a test that reliably measures all eight of his intelligences.

NEUROPSYCHOLOGICAL THEORIES AND IQ TESTS

The clinical field of neuropsychology (Luria, 1966) and the laboratory research field of psychobiology (Sperry, 1968) have contributed much to the field of IQ testing, especially Luria's neuropsychological theory. But it was the fact that Sperry's brilliant work was ignored by IQ test developers in the 1970s that impelled me to register a strong complaint: "Individual intelligence testing has been remarkably resistant to change, despite advances in related fields such as psychology and neurology. . . . The item content and mental processes assessed by conventional intelligence tests have not changed very much since the turn of the century when Alfred Binet and his coworkers engaged in their pioneering test development research" (Kaufman, 1979a, p. 96). Happily, that complaint is no longer true.

Sperry's Split-Brain Research

In the late 1970s I was enamored with Roger Sperry's (1968) research on patients who had "split-brain" surgery and with the cerebral specialization theory that evolved from this research (Kaufman, 1979a). This radical surgery, which was sometimes given to patients with severe epilepsy, involved cutting the corpus callosum—a thick band of nerve fibers that runs across the top of the skull and connects the two hemispheres of the brain. Patients who had this surgery saw a lessening of their symptoms of epilepsy (sometimes violence), but the doctors and psychologists who evaluated them saw something else entirely: They saw people who seemed to have two separate brains. With the two hemispheres surgically separated, and the two halves of the brain no longer in regular communication, it was possible to test one hemisphere at a time using tachistoscopes and a little imagination. The results were astonishing. The same person would respond to the exact same test item differently depending on which half of the brain was being assessed. For example, one item showed a picture of a birthday cake on a plate and the person had to point to the picture that was most closely associated with the cake. When the left hemisphere was asked to respond, the person pointed to a knife and fork, which was a conceptual, functional, stimulus-response association. In contrast, the right hemisphere pointed to a cowboy hat, because the wide-brimmed hat looked like the cake on a plate.

Once the professionals recovered from the shock of watching a person give two different answers to the same question, the surgeons began to realize the full impact of the split-brain procedure and modified it to keep a few of the interhemispheric fibers intact. But before this insight was reached, there were already a number of children and adults who effectively had two brains. And they were studied, and studied again, by medical and psychological researchers (Bogen, 1969). Research revealed that the two hemispheres had different styles of solving problems (Levy & Trevarthen, 1976; Levy-Agresti & Sperry, 1968). The left brain was analytic, time oriented, and dependent on language, and it tended to think in a logical-sequential manner. In contrast, the right brain integrated many stimuli at once in a gestalt-holistic fashion, tended to be nonverbal and spatial, and processed information in a simultaneous format (hence, the right hemisphere selected the cowboy hat as the right answer because it *looked like* the cake on the plate).

I believed that cerebral specialization theory was a perfect foundation for an IQ test: "The time has come for individual intelligence tests, *the* construct of intelligence to many people, to be substantially modified in accordance with the implications of the vital and dynamic research relating to brain functioning. . . . There can be little justification for being blind to the impact of split-brain research" (Kaufman, 1979a, p. 96).

Sequential Versus Simultaneous Processing

At first glance, Wechsler's armchair division of his subtests into Verbal and Performance scales seemed to mirror almost exactly what the hemispheres were specialized to do—the left was the verbal half of the brain and the right was the nonverbal. But that simple verbal-nonverbal distinction was known for years before Sperry's innovative research. Ralph Reitan (1955), who popularized the field of neuropsychology, had already conducted numerous studies of patients with known brain damage to a single hemisphere (e.g., adults who had a tumor in the left hemisphere or a stroke in the right hemisphere). He hypothesized that patients with left-brain damage should have a relatively low Verbal IQ whereas those with right-brain damage should have a low Performance IQ, and he conducted clinical research studies to try to prove his point (Reitan, 1955). However, that line of research never produced the promised fruits: In general, patients with right-hemisphere damage tended to earn Performance IQs that were lower than their Verbal IQs (V > P profiles), but left-damaged patients tended *not* to earn the opposite P > V profile (Kaufman & Lichtenberger, 2006, chapters 8 and 9).

Why did the research not fully support the prediction? That is where Sperry's innovative research came in. The original notion that equated the left hemisphere with verbal ability and the right hemisphere with nonverbal ability was a distinction that was based on the *content* of the test items—Did the questions involve verbal content (like Wechsler's Vocabulary subtest) or pictorial, figural content (like Wechsler's Block Design subtest)? At first the cerebral specialization researchers were also thinking content, but the more they evaluated split-brain patients, the more they focused on the *process* preferred by each hemisphere. The left half of the brain was analytical and sequential, which is useful for understanding language; but it was the *analytic-sequential processing style* that distinguished this half of the brain, not its capacity for language. By contrast, the right hemisphere favored a *simultaneous-holistic processing style*—again, this type of processing facilitated the handling of spatial, nonverbal stimuli, but the key was the process, not the content. The simplistic example of the birthday cake and the cowboy hat is helpful here. The content was held constant when each half of the brain was given the test (i.e., nonverbal, pictorial content)—but each hemisphere solved the problem differently because of its own distinctive style of processing information.

Ultimately, however, Sperry's cerebral specialization theory did not revolutionize the field of neurology or the field of IQ testing. Similar processing dichotomies had been springing up all over the broad field of psychology in the middle of the 20th century. The problem was that the field of psychology had become so specialized at that time that researchers in one area (e.g., psychobiology) didn't read journals in other areas (e.g., neuropsychology), and brain-related disciplines tended to ignore publications in the more traditional laboratory science journals (e.g., cognitive psychology). How else could one explain the psychobiologist Sperry's (1968) "discovery" that the right hemisphere was intelligent at nonverbal problem solving when the neuropsychologist Reitan (1955) had been conducting research on that very topic for years?

At the same time as Sperry was uncovering the mysteries of the right hemisphere's unique processing style, the great Russian neuropsychologist Alexander Luria was publishing his innovative clinical findings based on investigations of patients with damage to a single hemisphere—the left hemisphere. Luria (1966) was writing about two distinct mental processes: successive and simultaneous. And Luria's descriptions of these two fundamental processes were in lockstep with Sperry's distinction between left-brain (successive) and right-brain (simultaneous) processing. But Luria was not operating out of a left-right distinction (how could he, when he studied patients only with left-brain damage?). Instead, Luria described a *front-back* division of the brain. He considered successive processing to be primarily a function of the fronto-temporal regions of the brain, in contrast to the occipital-parietal localization (at the back of the skull) that accounts mainly for simultaneous syntheses. Luria had rotated the brain 90 degrees. Or maybe Sperry had rotated it 90 degrees. Did it matter? What was most important was that these two pioneers from different fields of neurology and different world views agreed that there were two basic, fundamental styles of solving problems and processing information that characterized human behavior.

And they were not alone in that belief. Outside of the brain sciences, researchers in cognitive psychology unearthed a processing dichotomy based on studies of visual search, attention, perception, detection, memory, and the like (e.g., Neisser, 1967). Only they referred to these two processes as *serial* and *parallel.* For example, Seller (1970) showed that the same type of stimuli (letters) could be processed either serially or in parallel depending on the demands of the task. When letters had to be matched based on physical identity, parallel processing was performed; in contrast, serial processing was a more efficient style that had to be used when the subjects could not rely on the physical properties of each letter to solve the problem. Many studies like Seller's established the existence of the two modes of information processing, but even more intriguing are the investigations by cognitive and other experimental psychologists who related the two processes to the left and right cerebral hemispheres. G. Cohen (1972), for example, extended Seller's work by showing the left hemisphere to be superior at matching names and the right hemisphere at matching shapes.

But the most consistent finding of the body of cognitive research was not the association of one type of processing to one hemisphere and a different type to the other half of the brain. The bottom line of all the research is that *process* hypotheses were supported far more often than *content* hypotheses. How the person goes about solving a problem was found to be more important than whether the questions involve handling verbal or pictorial or numerical stimuli.

Sequential or successive processing involves solving problems in a step-by-step fashion, placing a premium on the serial or time-related order of stimuli; in contrast, simultaneous processing demands a gestalt-like, frequently spatial, integration of stimuli to solve problems with maximum efficiency. We all tend to have our own preferred style of learning and problem solving. Suppose someone gives you a hand-drawn map so you can navigate a trip from his or her home to yours, a map that includes a snapshot of the entire trip from starting point to end point, including major roads and highways. Does the map make you feel happy and secure? If so, then you probably prefer a simultaneous-holistic processing approach to problem solving. You like a visual representation that shows the trip from start to finish—that is, the whole trip at a glance. But if the map strikes fear in your heart ("But where do I turn? Is it a right or left turn?"), then you are conceivably a sequential processor. You'd like to have a carefully spelled-out list of instructions: (1) Left at the third light (Spruce Street), (2) Right at the stop sign (Third Avenue), (3) Go about 1 mile and get onto Interstate 5 going north, and so forth. Simultaneous processors also have their moments of panic—like when they ask for directions at a toll booth and are told in rapid succession, "Take the second exit, then go about 4 blocks past the train station and turn right, then make a quick left onto a one-way street, and follow that with a sharp right at a T intersection; 2 or 3 miles ahead you'll see the shopping mall on the left, just past the post office."

And if you want the map *and* the written-out list of directions? Maybe you are insecure. Or maybe you are an integrated problem solver who relies on both sequential and simultaneous processing about equally. Or maybe you need to be given a comprehensive test to determine your best way of solving problems and processing information because who would trust a one-item test (i.e., using a map or not) in the first place? But one thing is true: People who give you directions are doing it the way *they* prefer to solve problems; they haven't got a clue about your preferred approach. Yet, suppose you know the other person's style of processing information and you are a teacher. Then you can adapt your teaching methods to the other's preferred processing style. And that is what Nadeen and I had in mind when we developed the K-ABC in the late 1970s and early 1980s.

The Kaufman Assessment Battery for Children (K-ABC)

As I reminisced during an invited address a few years ago (Kaufman, 2005a, 2005b):

I remember so clearly the day Nadeen and I came up with the ideas that would become the K-ABC. When we lived in Athens, Georgia, in the mid- to late-1970s, we frequently drove with our three kids to shopping malls in Atlanta for entertainment. One time we were going to see Luis from the television show *Sesame Street,* and while we were driving we were trying to plan our next project. We decided we were going to develop a test; on the entire 2-hour ride to Atlanta we were talking about Roger Sperry's cerebral specialization theory and the distinction between right and left hemisphere brain functioning. We were talking about cognitive styles, translations of test scores to educational remediation, nonverbal assessment, and fairness to ethnic minorities. We discussed the importance of developing interesting and novel tasks and the need to emphasize process instead of content. We were going to Atlanta in the first place, in part, because we felt guilty about not spending enough time with our kids and then we proceeded to ignore them the whole ride there! On the ride home we asked each other, "Who are we fooling? We will develop a test and nobody will publish it." So we decided to forget about that idea. Until the next day, a Monday, when we got a call from Dr. Gary Robertson, director of test development at American Guidance Service (AGS), a Minnesota test publisher. Gary asked whether Nadeen or I or both of us would like to develop a new test of intelligence. So being on the impulsive side of the impulsive-reflective cognitive style—at least at that moment—I blurted out, "Oh we developed an IQ test yesterday." It took a little bit of time to regain credibility with AGS, but that was the beginning of the K-ABC. Of course, if we had actually developed the precise test we had designed in the car, it would have taken about 24 hours to administer. Being idealists, we wanted to measure *everything* that we thought was important about children's mental ability and cognitive style. Being realists, we knew we had to be a bit more practical in what just *had* to be included in the K-ABC.[1]

Ultimately, we accomplished our most important goals: (a) to be rooted in theory; (b) to include new and interesting tasks; (c) to reduce IQ differences between White and African American children (the typical differences of about 15 points on Wechsler's scales were cut in half on the K-ABC); (d) to separate the ability to solve new problems (mental processing or "intelligence") from acquired knowledge and language skills ("achievement"), thereby providing a less language-based assessment of bilingual children's intelligence; and (e) to include "teaching items." These innovative teaching items ensured that children understood exactly what was expected of them for each task. Research had shown that young children have difficulty understanding basic concepts like "same" or "first" or "under" (Boehm, 1967). Similar concepts are commonly included in the directions spoken by the examiner when testing children on IQ tests (Kaufman, 1978). Because the tests are standardized, examiners aren't allowed to change the words of a question or an instruction. So we built in teaching items—an unscored sample item and the first two items of each subtest—enabling the examiner to feel confident that children understood the test directions. Examiners were told to teach the child, whenever necessary, by using different words, gestures, or a different language, including American Sign Language. These teaching items helped ensure that a low subtest score reflected low ability, not just bad communication between the examiner and child.

As we said in the test manual (Kaufman & Kaufman, 1983), "Intelligence, as measured by the K-ABC, is defined in terms of an individual's *style* of solving problems and processing information; this definition, which also stresses *level of skill* in each style

1. Copyright 2005 by the National Association of School Psychologists. Bethesda, MD. Adapted with the permission of the publisher. www.naspOnLine.org

of information processing, has a strong theoretical foundation in the domains of both neuropsychology and cognitive psychology" (p. 2). So we began by trying to develop an IQ test from Sperry's cerebral specialization theory but wound up building the K-ABC on a sequential-simultaneous foundation that spanned multiple theories and disciplines. We learned so much during the test development process, especially from our team of graduate students who went on to become international leaders in school psychology and assessment (Bruce Bracken, Jack Cummings, Patti Harrison, Randy Kamphaus, Jack Naglieri, and Cecil Reynolds). They helped us realize, ultimately, that the key was the distinction between the two processes, not the possible link-up between process and hemisphere. For children, especially, whose brains are "plastic," it is not really feasible to figure out whether Sperry's right-left or Luria's front-back distinction is more plausible. Moreover, the research on patients with brain damage is predominantly based on adults (Kaufman & Lichtenberger, 2006; Matarazzo, 1972), and the intact parts of a damaged brain might no longer function precisely as they did before the damage.

Word Order is an example of a K-ABC sequential processing subtest. The examiner says the names of objects (e.g., car—lamp—horn) and then the child has to point to pictures of the objects in the order in which they were named. For more difficult items, the child has to name pictures of colors before responding, an "interference task" that prevents rehearsal. Word Order, including the verbal interference task, is an adaptation of the clinical tests that Luria (1966) used to measure the higher brain functions of patients with brain damage.

Figure 3.1 depicts a K-ABC simultaneous processing subtest (Gestalt Closure). For this task, the child has to name the object or scene pictured in a partially completed inkblot drawing. This type of task was important to include in the K-ABC because "it has proved so valuable as an accepted prototype of simultaneous processing and right hemispheric functioning . . . [and] has produced approximately equal mean scores for [African Americans], Hopi Indians, and whites" (Kaufman & Kaufman, 1983, pp. 40–41).

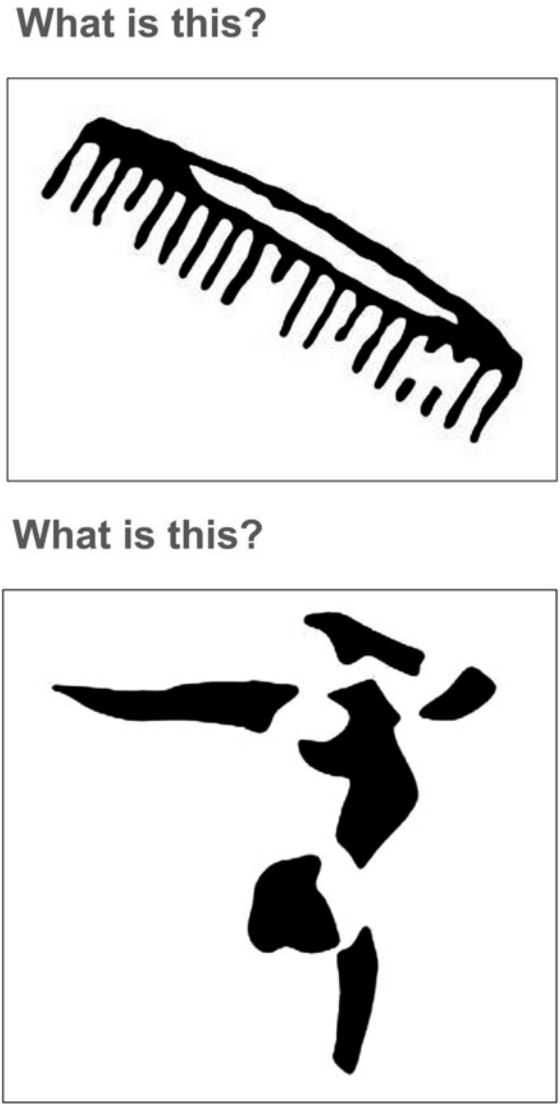

FIGURE 3.1 Sample simultaneous processing items (similar to items on the K-ABC and KABC-II Gestalt Closure subtest).

Note. From *Kaufman Assessment Battery for Children, Second Edition (KABC-II)*, by A. S. Kaufman and N. L. Kaufman, 2004, Circle Pines, MN: American Guidance Service.

When the K-ABC was published, there was much media hype (TV interviews, radio interviews, radio call-in shows, cable TV features) for a couple of years, and controversy was rampant within the professional community as well. The media was interested primarily in the K-ABC's greatly reduced ethnic differences for African American, Hispanic, and Native American children. Psychologists and special educators debated the theoretical foundation of the K-ABC, the reasons underlying the reduced ethnic differences, and our decision to keep verbal tasks and measures of acquired knowledge (both staples of the Stanford-Binet and Wechsler IQs) out of the IQ scale. The *Journal of Special Education* published a special issue devoted to the K-ABC (Miller & Reynolds, 1984) in which the test was either praised or damned, depending on the perspective of the expert invited to contribute to the special issue. Ultimately, the key point was the theory on which the test was based, and there were wildly differing opinions. Raymond Dean, a leader in the field of school psychology who would later coauthor a neuropsychological battery, was complimentary, stating that "the K-ABC represents a theoretically consistent battery of tests that offers insights into children's cognitive processing beyond presently available measures of intelligence" (Dean, 1984, p. 251). Noted theorist Robert Sternberg disagreed, stating that the K-ABC "is based on an inadequate conception of intelligence, and as a result, it is not a good measure of intelligence" (Sternberg, 1984, p. 277). (My son James would eventually earn his PhD at Yale under Dr. Sternberg, who would prove to be a wonderful mentor for James as well as a collaborator on numerous books. And James would eventually edit the *Psychology 101* series.)

Despite our initial inspiration from Sperry's model and our clear statements in the test manual that our theoretical foundation was built on a research and theoretical base that encompassed both brain-related and cognitive perspectives, our test soon became known simply as a Luria-based test that addressed only a portion of Luria's neuropsychological model and was, therefore, incomplete. That criticism was not heard much in Europe and Asia, where adapted and translated versions of the K-ABC thrived and its sequential-simultaneous model (with language and factual items excluded from the IQ measure) was respected no less than Wechsler's traditional verbal-performance distinction. The K-ABC flourished, for example, in Germany (Melchers & Preuß, 1991), France (Voyazopolous, 1994), and Japan (Matsubara et al., 1994).

But in the United States, the K-ABC model of intelligence would come to be seen as an incomplete measure of Horn's (1989) broad abilities that focused too much on short-term memory (*Gsm*) and visual processing (*Gv*) and too little on fluid reasoning (*Gf*) and comprehension-knowledge (*Gc*) (Keith, 1985). Carroll (1993) concluded that most K-ABC mental processing tasks were the well-known ability factors of what he termed VZ (visualization), LD (language processing), and MS (memory span). The great theorist dismissed the K-ABC by stating: "With respect to factorial content, there is little if anything that is new in the K-ABC test" (Carroll, 1993, p. 703). I guess it's better than being ignored!

Nonetheless, the K-ABC served some useful functions from a historical perspective. It was the first *theory-based* individually administered, clinical IQ test, and it showed that the Wechsler-Binet monopoly could be challenged, opening the door for the spate of theory-based tests of cognitive processing and cognitive abilities that would follow it during the 1980s and that continue in the 21st century. It included truly novel IQ tasks, not the same old recycled verbal and nonverbal tasks that traced their lineage to Alfred Binet and American World War I psychologists. (In this regard, Dick Woodcock, 1978, was also a pioneer in developing innovative cognitive tasks for his Woodcock-Johnson Psycho-Educational Battery.) The K-ABC shifted the focus from content to process in the eyes of many clinicians, even those who continued to administer the content-based Wechsler scales. It showed that it was possible to greatly reduce ethnic differences in IQ when care was taken to ensure fairness. It included more than 40 research studies in the test manual to demonstrate that the test was valid, whereas previous test manuals barely provided validity evidence. And we encouraged our test publisher (then called AGS, now Pearson Assessments) to hire a school psychologist as project director (they hired Randy Kamphaus, now an eminent leader in the field of cognitive and behavioral assessment), in contrast to the statistical-mathematical experts who served as project directors for previous IQ tests. This latter change ensured that a scientist-practitioner, someone who actually administered clinical tests to children and adults, was in charge of test development, not someone who excelled as a scientist but lacked real-world experience.

Today, most of these K-ABC innovations have become standard practice in current tests, including revisions of the Wechsler and Stanford-Binet. Theory-based tests of exceptional quality abound. Teaching items and novel subtests are included in nearly every IQ test. The emphasis on processing characterizes all major IQ tests, including the WISC-IV and WAIS-IV, which now yield four process-based indexes instead of two content-based IQs (i.e., Verbal and Performance). And practitioner-scientists are the rule, not the exception, as project directors and executives of test publishers. The chief executives of the two test publishers we work with directly are a clinical neuropsychologist (Dr. Aurelio Prifitera of Pearson Assessments) and a clinical psychologist (Dr. Mireille Simon, of ECPA in Paris).

But the sequential-simultaneous theory on which the K-ABC was based did not stand the test of time, at least in the United States. From the perspective of neuropsychologists, the K-ABC was criticized as measuring only Block 2 of Luria's three-block model. As discussed in the next section, the K-ABC model has been superseded by a more complete representation of Luria's neuropsychological model.

Luria's Three-Block Neuropsychological Theory

Luria's (1970) goal as a neuropsychologist was to map out the brain's systems and functions responsible for complex behavioral processes, especially the high-level processes associated with the intake and integration of information and with problem-solving abilities. Luria perceived the brain's basic functions to be represented by three main blocks, or functional systems.

- Block 1 is responsible for arousal and attention.
- Block 2 uses successive (sequential) and simultaneous (holistic) processing to analyze, code, and store information.
- Block 3, associated with the frontal lobes of the brain, is responsible for planning, decision making, and what clinical neuropsychologists refer to as "executive functions."

Figure 3.2 summarizes the functions associated with each of Luria's three blocks. The arrows between adjacent blocks reflect Luria's emphasis that the *integration* of the three blocks is necessary to permit complex thinking.

Many empirical studies support Luria's (1970, 1973) clinical documentation of the three functional units (see, for example,

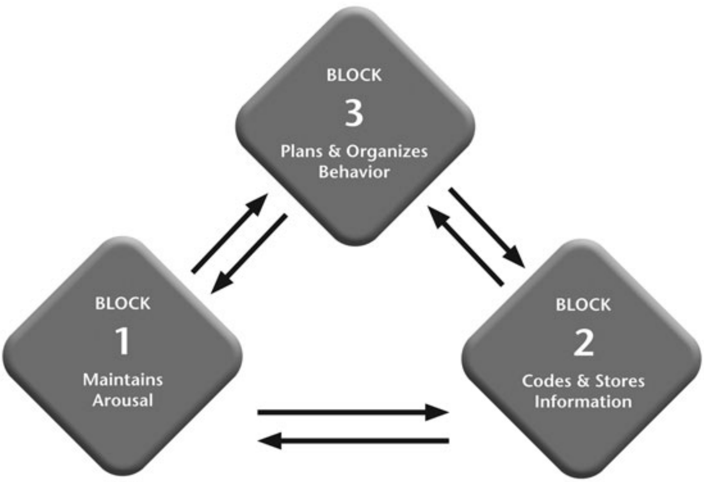

FIGURE 3.2 Luria's three blocks or functional units.
Note. From *Kaufman Assessment Battery for Children, Second Edition (KABC-II)*, by A. S. Kaufman and N. L. Kaufman, 2004, Circle Pines, MN: American Guidance Service.

Naglieri, 1999). Much neurological evidence supports the ages of 11 to 12 as crucial for the development of the prefrontal cortex, leading to the refinement of Block 3 executive functions, such as working memory, for making decisions, thinking abstractly, and solving complex problems (Golden, 1981). (Working memory is the mental scratchpad that allows us to hold onto information long enough to solve complex problems.)

Naglieri and Das's PASS Theory

Jack Naglieri and J. P. Das were instrumental in translating Luria's three blocks to the practice of cognitive assessment. Jack worked closely with Nadeen and me to develop the K-ABC, and J. P. studied with Luria in Russia. Together they developed the Luria-based PASS theory, which was an expansion of the K-ABC's sequential-simultaneous processing distinction. The **P** in PASS refers to Planning, the Block 3 function; the **A** denotes Block 1's Attention; and the two **S**'s refer to Luria's Block 2 coding processes, Successive and Simultaneous. They used PASS theory as the theoretical foundation of the Cognitive Assessment System (CAS; Naglieri & Das, 1997), a test for ages 5 to 17 years that has proved to be useful for developing educational interventions and treatment. Improving planning ability, for example, has been shown to lead to improvement in math achievement (Naglieri & Gottling, 1997; Naglieri & Johnson, 2000). In addition, groups of children with specific disorders such as ADHD or reading disabilities tend to display characteristic profiles on the CAS. Children with ADHD, for example, tend to perform considerably better on the Successive and Simultaneous scales than on Attention or Planning, whereas children with reading disorders tend to do worst on Successive Processing (Naglieri, 1999, Figure 6.3).

The two types of mental processing have already been defined. Let's look at the P and the A that make up the PASS model (Naglieri, 1999).

Planning is a mental process that requires a person to "develop a plan of action, evaluate the value of the method, monitor its effectiveness, revise or reject a previous plan as the task demands change, and control the impulse to act without careful consideration" (Naglieri, 1999, p. 13). This process is illustrated in Figure 3.3 by the CAS subtest Planned Codes, which requires the child to write a code (e.g., XO) under the appropriate letter (A, B, C, or D). The child's success is facilitated by the choice of an

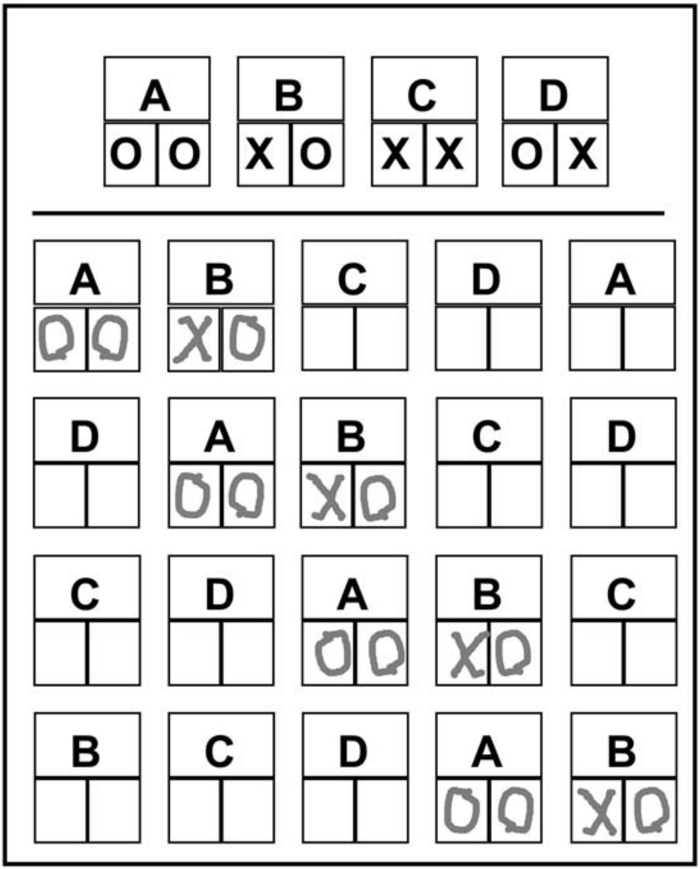

FIGURE 3.3 Example of a planning test item from the Cognitive Assessment System (CAS).
Note. From *Essentials of CAS Assessment* (p. 13, Figure 1.4) by J. A. Naglieri, 1999, New York: Wiley. Reproduced with permission.

effective strategy to permit very quick responding (e.g., doing all the A's first, then the B's, etc.). It is fascinating to realize that Binet was so far ahead of his peers such as Francis Galton and James McKeen Cattell when he theorized about intelligence, because he was talking about the Block 3 functions of the prefrontal cortex before anyone knew very much about these executive functions. Binet and Simon (1916/1973) considered intelligent thought to require *direction*, *adaptation*, and *criticism*. These aspects of intelligence form the essence of Luria's notion of planning ability: (a) direction "consists of knowing what has to be done and how to do it"; (b) adaptation "refers to the selection and monitoring of our strategy"; and (c) criticism "is our ability to criticize our own thoughts and actions . . . and to change our behavior in such a way as to improve our performance" (Sternberg et al., 2008, p. 10).

Attention is a mental process that requires a person to selectively focus on specific stimuli while inhibiting responses to competing stimuli. Figure 3.4 illustrates an Attention item from the CAS subtest Number Detection.

When we revised the K-ABC and developed the KABC-II (Kaufman & Kaufman, 2004a), for ages 3 to 18 years, we followed the lead of Naglieri and Das (1997) and expanded the neuropsychological theory underlying the KABC-II to include the Block 3 functions of planning ability, and we also added a learning scale to evaluate a child's ability to learn and retain new material

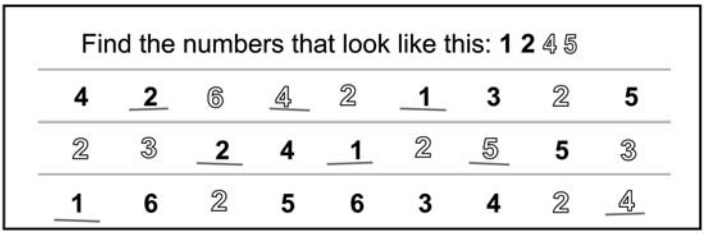

FIGURE 3.4 Example of an attention test item from the Cognitive Assessment System (CAS).
Note. From *Essentials of CAS Assessment* (p. 16, Figure 1.7) by J. A. Naglieri, 1999, New York: Wiley. Reproduced with permission.

during the assessment. Importantly, success on the learning tasks requires an integration of the three blocks. The KABC-II tasks demand Block 1's focused and selective attention, Block 2's coding and storage of auditory and visual stimuli, and Block 3's capacity to generate strategies to learn the material efficiently. Reitan (1988) said about Luria's theory that "integration of these systems constitutes the real key to understanding how the brain mediates complex behavior" (p. 333). To Luria (1970), "It is clear that every complex form of behavior depends on the joint operation of several faculties located in different zones of the brain" (p. 68).

Figure 3.5 shows sample items from a learning task that teaches the person a new language, namely, the word or concept

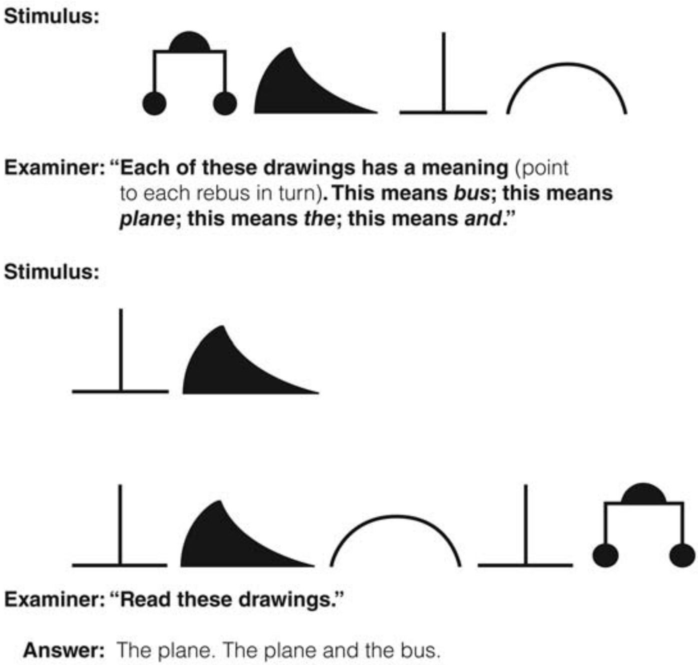

FIGURE 3.5 Example of a learning subtest (Rebus Learning).
Note. From *Kaufman Adolescent and Adult Intelligence Test (KAIT)*, by A. S. Kaufman and N. L. Kaufman, 1993, Circle Pines, MN: American Guidance Service. Copyright © 1993 by NCS Pearson, Inc. Reproduced with permission. All rights reserved.

that corresponds to specific pictures and abstract symbols (rebuses). Then the person has to "read" both simple and complex sentences composed of these symbols. The figure illustrates Rebus Learning from the KAIT, which is very similar to the KABC-II Rebus subtest.

In developing the KABC-II we relied on a dual theoretical model—*both* Luria's three functional units or blocks *and* CHC theory. We believe that the two theories complement each other well, and that both theories have a great deal to offer for the measurement of intelligence. Which leads us to the next topic.

CATTELL-HORN-CARROLL (CHC) THEORY

Neuropsychological processing theories made their mark on IQ tests, but the theory that has most influenced today's intelligence tests is an amalgam of two related theories of cognitive abilities: Horn and Cattell's (1966) theory of fluid and crystallized intelligence and Carroll's (1993) three-stratum theory, known as the CHC (Cattell-Horn-Carroll) model. Let's take the original theories in sequence; Cattell-Horn's *Gf-Gc* theory (Horn & Cattell, 1966) had the first, dramatic impact on the interpretation of IQ tests, most notably Wechsler's scales, before Carroll's (1993) exhaustive research was even known by the field of IQ testing.

Cattell and Horn's *Gf-Gc* Theory

Some years after the elaboration of Spearman's (1904, 1927) influential *g* theory, Thurstone (1938) and other leading psychologists argued strongly against *g* and advocated theories that hypothesized group factors over and above Spearman's *g* and *s* (Jensen, 1998). Even the noted learning theorist Clark Hull wrote the book *Aptitude Testing* (Hull, 1928), which foreshadowed the shift in emphasis from *g* to multiple ability approaches (Thorndike et al., 1986). In contrast to the trend either to advocate *g* or to argue against *g* (by proposing a half-dozen or so multiple abilities), Raymond Cattell simply split *g* into two pieces. As Spearman's doctoral student, Cattell (1941, 1963) built upon his mentor's approach to intelligence. His new system embraced *g* but posited two types of *g* abilities, not just one:

- *Fluid intelligence (Gf),* the ability to solve novel problems by using reasoning; Cattell believed that *Gf* was largely a function of biological and neurological factors and was vulnerable to the effects of aging.
- *Crystallized intelligence (Gc),* a knowledge-based ability that Cattell considered to be extremely dependent on education and acculturation and largely resistant to the impact of aging.

Raven's (1938) abstract Progressive Matrices Test has always been used as the paradigm of fluid reasoning and was considered by Spearman to be the very best measure of *g* (Silverman, 2009). Raven's measure of abstract reasoning has been copied by many test developers, with adaptations appearing in recent versions of Wechsler's scales, the Stanford-Binet revisions, the Kaufman tests, and other batteries as well (e.g., Naglieri & Das, 1997; Woodcock et al., 2001). Figure 3.6 depicts an illustrative matrices item, which requires *Gf* to figure out the relationships among the abstract designs in the 3 × 3 matrix (this is a medium-difficulty example; see Figures 4.1 and 4.2 for easy and challenging items, respectively).

Figure 3.7 illustrates a Differential Ability Scales—Second Edition (DAS-II) task that uses numbers instead of designs to measure fluid reasoning (sequential and quantitative reasoning), and Figure 3.8 demonstrates KABC-II's Story Completion, which uses pictures to assess *Gf*.

The strategies needed to solve *Gf* items resemble Luria's notion of Block 3 planning and decision-making ability to a considerable extent, which is why the KABC-II scale for school-age children that measures high-level, complex, abstract reasoning is called Planning/*Gf*. In Story Completion, the child is shown a row of pictures that tells a story, but some of the pictures are

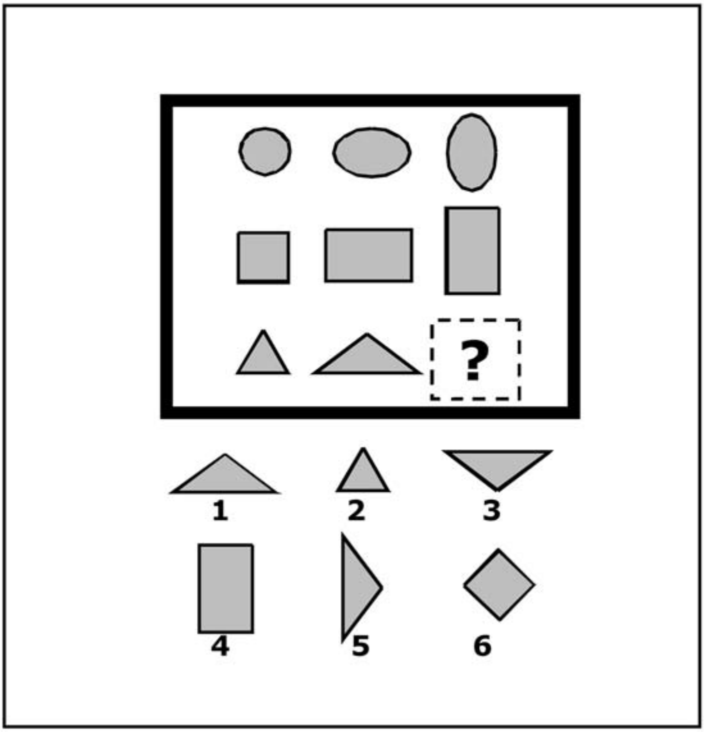

FIGURE 3.6 Example of a *Gf* matrices item (similar to items on a variety of matrices tests).
Note. The correct response is #5. From *Essentials of DAS-II Assessment*, by R. Dumont, J. O. Willis, and C. D. Elliott (p. 310, Figure 8.1), New York: Wiley. Reproduced with permission.

missing. The child has to complete the story by selecting pictures from an array of cards (see Figure 3.8). The *Gf* needed to select the appropriate pictures and insert them in their correct sequence in the story requires the Block 3 functions of developing a plan of action, generating hypotheses, making quick decisions, controlling impulses, and revising or rejecting previous plans as the task demands change.

Gc is measured by a variety of tasks, usually verbal. An example of a fairly pure *Gc* task is the Vocabulary subtest on the various

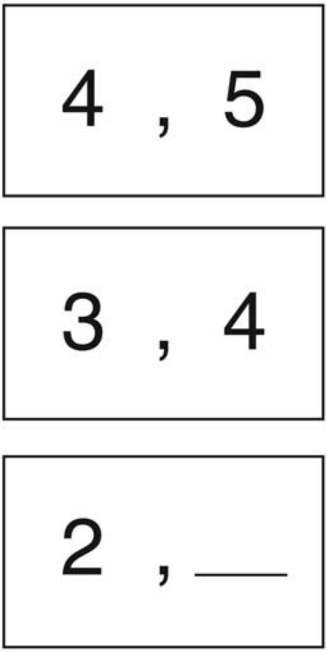

FIGURE 3.7 Example of a *Gf* subtest that uses numbers as stimuli (DAS-II Sequential and Quantitative Reasoning).
Note. Answer = 3. The child figures out how two pairs of numbers relate to each other and then applies the rule to discover the missing number in the incomplete pair. *Differential Ability Scales—Second Edition (DAS-II)*, by C. D. Elliott, 2007, San Antonio, TX: The Psychological Corporation. Copyright © 2007 by NCS Pearson, Inc. Reproduced with permission. All rights reserved.

editions of the Wechsler scales and Stanford-Binet ("What do we mean by *edifice*? What does *agitation* mean?"). But *Gc* tasks can also utilize pictorial stimuli, as illustrated by the KABC-II Expressive Vocabulary subtest (see Figure 3.9).

John Horn was Cattell's doctoral student. And just as Cattell expanded his mentor's theory of intelligence, so, too, did Horn. Horn and Cattell (1966, 1967) initially focused on the generation-old *Gf-Gc* dichotomy. But despite Cattell's (1963) devotion to *Gf* and *Gc*, Horn never really bought into the model. Almost from the beginning—indeed in his doctoral dissertation—Horn (1965)

FIGURE 3.8 Illustrative item for the Story Completion subtest (on the KABC-II Planning/*Gf* scale).
Notes. 1. Of the four pictures at the bottom, the first two on the left do not go with the story; they are distractors. The fourth picture from the left goes second in the story (showing the father letting go of the bike), and the remaining picture goes third in the story (showing the daughter riding by herself, just before the dog runs in front of her).
2. From *Kaufman Assessment Battery for Children, Second Edition (KABC-II)*, by A. S. Kaufman and N. L. Kaufman, 2004, Circle Pines, MN: American Guidance Service.

believed that the research supported more than just these two general abilities. He quickly identified four abilities in addition to *Gf* and *Gc* (Horn, 1965, 1968): short-term acquisition and retrieval (*Gsm*), long-term storage and retrieval (*Glr*), visual processing (*Gv*), and speed of processing (*Gs*).[2] That number would grow to 9 or 10 *broad abilities* by the mid-1990s (Horn, 1989; Horn & Noll, 1997).

The initial dichotomy had been expanded, but Horn did not consider any of the abilities to be more or less important than others. Although the theory continued to be called *Gf-Gc* theory, the 9 or 10 broad abilities were treated as equals, not as part of any type of hierarchy.

2. Different abbreviations and symbols have been used for various CHC abilities. The ones shown in parentheses are the ones currently used by most CHC theorists, not necessarily the original symbols.

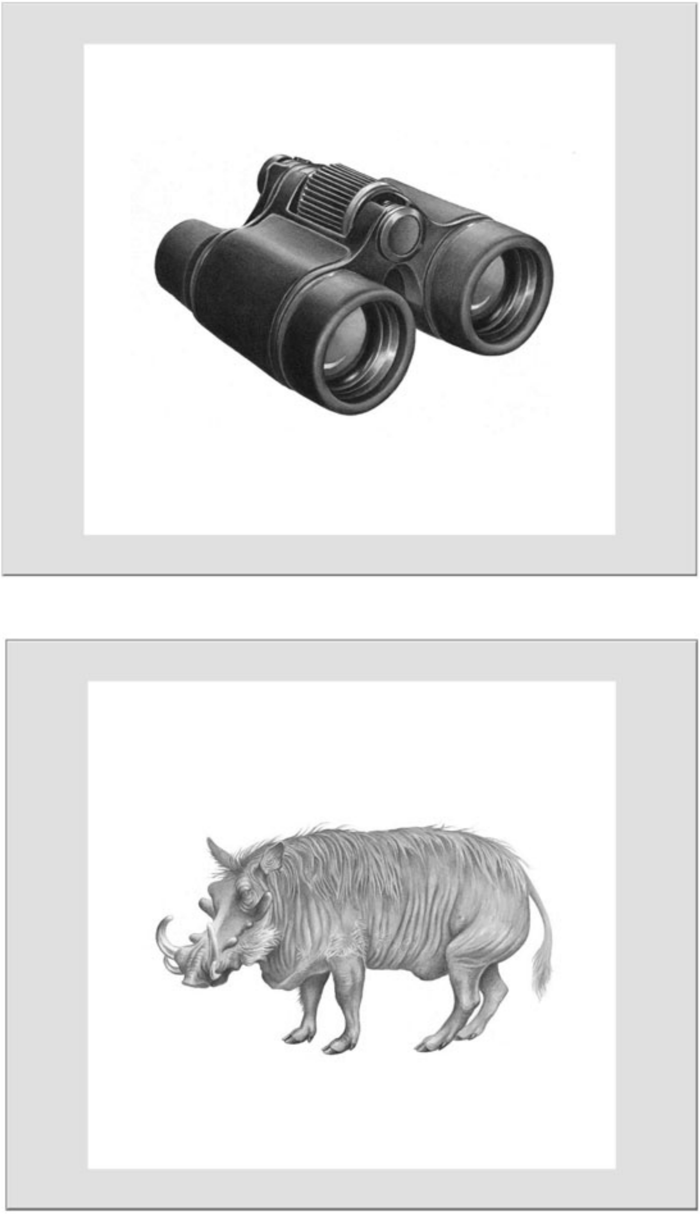

FIGURE 3.9 Illustrative items for the Expressive Vocabulary subtest (on the KABC-II Knowledge/*Gc* scale).

Note. Binoculars and warthog. From *Kaufman Assessment Battery for Children, Second Edition (KABC-II)*, by A. S. Kaufman and N. L. Kaufman, 2004, Circle Pines, MN: American Guidance Service.

John Carroll's Three-Stratum Theory

In contrast to Horn's egalitarian approach to cognitive abilities, John Carroll (1993, 1997) developed a hierarchical theory composed of three levels, or strata, of abilities:

- Stratum III (General), a Spearman-like *g*, which Carroll (1993, 1997) considered to be a valid and vital construct
- Stratum II (Broad), composed of 8 broad abilities that correspond closely to Horn's (1989) broad abilities and correspond roughly to Gardner's (1993) multiple intelligences (Carroll, 1997)
- Stratum I (Narrow), composed of about 70 fairly specific abilities, many of which indicate the person's "level of mastery, along a difficulty scale," "speed with which the individual performs tasks," or "rate of learning in learning and memory tasks" (Carroll, 1997, p. 124)

As my friend and colleague Mark Daniel (1997) said about Carroll's theory, "Never before has a psychometric-ability model been so firmly grounded in data" (p. 1043).

Horn's (1989) theory always focused on the broad abilities (Carroll's Stratum II), but Horn also discussed the more specific or narrow abilities as well. To Horn, Spearman's *g* (Stratum III of Carroll's model) had no place in any theory. It made him see red when other theorists defended it. Otherwise, the Carroll and Cattell-Horn theories were similar enough to warrant being merged into the new CHC theory.

The Merger of Theoretical Models to Form CHC Theory

The CHC model, the blend of the Cattell-Horn and Carroll theories, is a psychometric theory that rests on a large body of research accumulated over decades in literally thousands of empirical investigations. CHC owes a debt to Thurstone's (1938) pioneering primary mental abilities theory: "to a considerable extent, modern hierarchical theories derive from this theory" (Horn & Noll, 1997, p. 62).

Horn and Carroll agreed to merge their theories into a single model in the late 1990s, without fanfare, in a personal communication to Richard Woodcock in July 1999. But about a dozen years earlier, at a 1986 meeting in Dallas that included Horn, Carroll, and Woodcock, the intimate link between the Cattell-Horn theory and Carroll's comprehensive research was discovered. As Kevin McGrew (2005) recalled, "A collective 'Ah Ha!' engulfed the room as Carroll's WJ [Woodcock-Johnson] factor interpretation provided a meaningful link between the theoretical terminology of Horn and the concrete world of WJ tests" (p. 144). Though CHC theory would not be on the agenda for years to come, that 1986 meeting "was the flash point that resulted in *all* subsequent theory-to-practice bridging events leading to today's CHC theory and related assessment developments" (McGrew, 2005, p. 144).

CHC theory focuses on 10 broad abilities, which together define the range of the major human intellectual capacities, as determined by the research conducted by John Horn (1989) and his colleagues and by the intensive survey of literature assembled by John Carroll (1993). Each broad ability is subdivided into specific narrow abilities, which total about 70. The relationship between broad and narrow abilities is illustrated in Table 3.1 for crystallized intelligence (*Gc*) (Flanagan, Ortiz, & Alfonso, 2007, p. 281).

This table shows 12 narrow abilities—for example, listening ability, foreign language aptitude, and general science information—each measuring a different facet of *Gc;* taken together they demonstrate the depth and breadth of crystallized intelligence. Narrow abilities are important, but the linchpin of CHC theory is the array of broad abilities. It has never been clear whether Carroll's Stratum III (*g* or general ability) is part of CHC theory or not. The topic was rarely talked about while Horn and Carroll were alive because it was their one main bone of contention. To Carroll, *g* was a crucial and fundamental concept; to Horn it was

TABLE 3.1 DEFINITIONS OF CRYSTALLIZED INTELLIGENCE (*Gc*) NARROW STRATUM 1 ABILITIES

Stratum I (Narrow)	Definition
*Crystallized Intelligence (*Gc*)*	
Language development (LD)	General development, or the understanding of words, sentences, and paragraphs (*not* requiring reading), in spoken native language skills
Lexical knowledge (VL)	Extent of vocabulary that can be understood in terms of correct word meanings
Listening ability (LS)	Ability to listen to and comprehend oral communications
General (verbal) information (K0)	Range of general knowledge
Information about culture (K2)	Range of cultural knowledge (e.g., music, art)
General science information (K1)	Range of scientific knowledge (e.g., biology, physics, engineering, mechanics, electronics)
Geography achievement (A5)	Range of geographic knowledge
Communication ability (CM)	Ability to speak in "real-life" situations (e.g., lecture, group participation) in an adult-like manner
Oral production and fluency (OP)	Narrower or more specific oral communication skills than reflected by communication ability (CM)
Grammatical sensitivity (MY)	Knowledge or awareness of the grammatical features of the native language
Foreign language proficiency (KL)	Similar to language development (LD) but for a foreign language
Foreign language aptitude (LA)	Rate and ease of learning a new language

Note. From *Essentials of Cross-Battery Assessment*, 2nd ed. (p. 281, Table A2), by D. P. Flanagan, S. O. Ortiz, and V. C. Alfonso, 2007, New York: Wiley. Reproduced with permission.

anathema. So Stratum III has usually been ignored, and its role in CHC theory remains ambiguous (McGrew, 2005).

Regardless, broad abilities rule the roost, both from a theoretical perspective and for determining which scales constitute most of today's IQ tests. *Gf* and *Gc* have already been defined and illustrated. Here are capsules describing the remaining eight broad abilities.

Short-Term Memory (*Gsm*)

Gsm is a person's ability to take in and hold onto information, keep it in immediate awareness, and use it within a few seconds. "An example of *Gsm* is the ability to remember a telephone number long enough to dial it or the ability to retain a sequence of spoken directions long enough to complete the tasks specified in the directions" (Flanagan et al., 2007, p. 284). Word Order from the K-ABC and KABC-II (pointing to pictures named by the examiner in the order in which they were named), used earlier to illustrate sequential processing, is also a good example of a *Gsm* task. So too is Wechsler's Digit Span (Part 1—repeating numbers in the order in which they are spoken by the examiner; Part 2—repeating the numbers in the *reverse* order).

Processing Speed (*Gs*)

Gs is a person's ability "to fluently and automatically perform cognitive tasks, especially when under pressure to maintain focused attention and concentration" (Flanagan et al., 2007, p. 291). This Stratum II ability is illustrated by the Wechsler's Symbol Search subtest in Figure 3.10.

Auditory Processing (*Ga*)

Ga is a person's "ability to perceive, analyze, and synthesize patterns among auditory stimuli, and to discriminate subtle nuances in patterns of sound (e.g., complex musical structure) and speech when presented under distorted conditions" (Flanagan et al., 2007, p. 287). The WJ III measures *Ga* by several tasks, including incomplete words, for which the person hears a recording

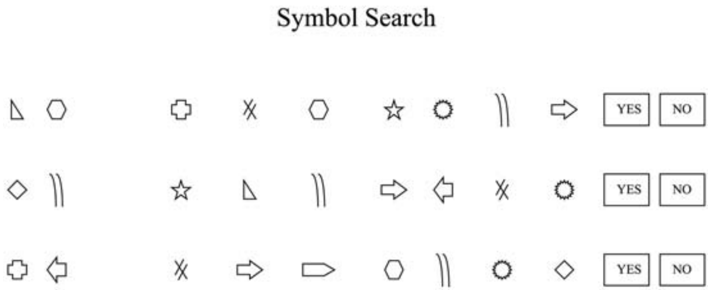

FIGURE 3.10 Illustration of a processing speed (*Gs*) subtest (Wechsler's Symbol Search).

Note. Symbol Search is a highly speeded task. The person has to look at the two symbols on the left side of each row and rapidly determine if either one of these target symbols appears in the array of symbols on the right. The person marks yes or no as quickly as possible. Of the three rows shown, only the third row is no. Simulated items similar to those in the Wechsler intelligence scales for adults and children. Copyright 1949, 1955, 1974, 1981, 1991, 1997, 1999 by NCS Pearson, Inc. Reproduced by permission. All rights reserved.

of words missing one or more phonemes and has to identify the complete word (e.g., "__eanut __utter" for *peanut butter;* or "__edroo__" for *bedroom.*

Visual Processing (*Gv*)

Gv is a person's "ability to generate, perceive, analyze, synthesize, store, retrieve, manipulate, transform, and think with visual patterns and stimuli (Lohman, 1994)" (Flanagan et al., 2007, p. 286). *Gv* is essentially the same thing as simultaneous processing, which was illustrated in Figure 3.1 with the Gestalt Closure subtest. Whereas Gestalt Closure depends mostly on perception and synthesis, some *Gv* tasks require the ability to use short-term memory (see Figure 3.11, which illustrates KABC-II Face Recognition, designed for preschool children), and some require visual-spatial reasoning, such as the KABC-II Block Counting subtest (see Figure 3.12). For Block Counting, visualization is needed to "see" the picture of the pile of blocks as a three-dimensional structure and reasoning is needed to figure out how many blocks are hidden or partially hidden.

FIGURE 3.11 Illustration of a visual processing (*Gv*) subtest that requires visual memory (KABC-II Face Recognition).
Note. From *Kaufman Assessment Battery for Children, Second Edition (KABC-II)*, by A. S. Kaufman and N. L. Kaufman, 2004, Circle Pines, MN: American Guidance Service. Copyright © 2004 by NCS Pearson, Inc. Reproduced with permission. All rights reserved.

Quantitative Thinking (*Gq*)

Gq is a person's "ability to use quantitative information and manipulate numeric symbols" (Flanagan et al., 2007, p. 282). It is math, an aspect of academic achievement. CHC theory lists it as a broad ability, but it's school achievement, not IQ.

Reading and Writing (*Grw*)

Grw is a person's "acquired store of knowledge that includes basic reading, reading fluency, and writing skills required for the comprehension of written language and the expression of thought via writing" (Flanagan et al., 2007, p. 283). Again, it's school achievement, not IQ.

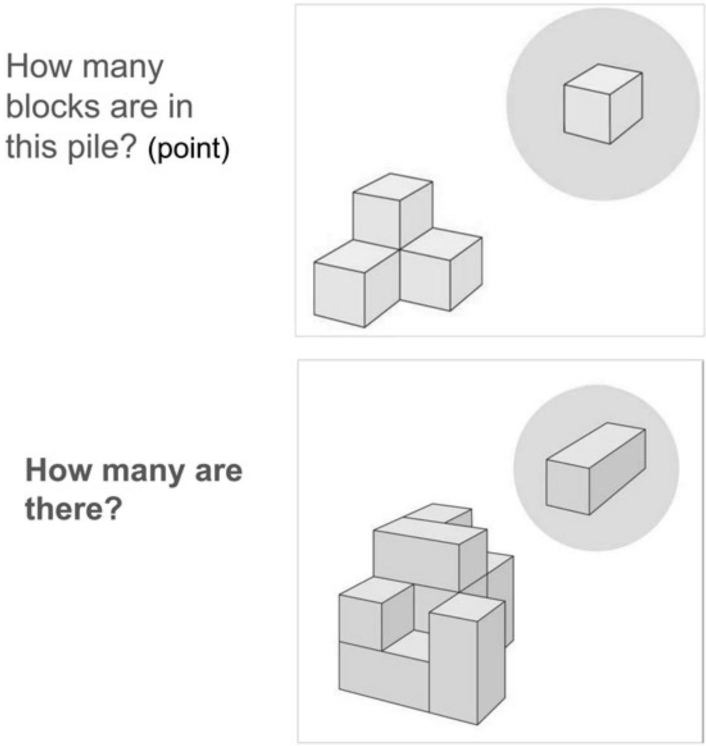

FIGURE 3.12 Illustration of a visual processing (*Gv*) subtest that requires visual-spatial reasoning (KABC-II Block Counting).

Note. Number of blocks: 4 (1st item) and 8 (2nd item). *Kaufman Assessment Battery for Children, Second Edition (KABC-II)*, by A. S. Kaufman and N. L. Kaufman, 2004, Circle Pines, MN: American Guidance Service.

Decision Speed/Reaction Time (*Gt*)

Gt is a person's quickness in reacting and making decisions, reflecting the immediacy of responding to a stimulus (measured in seconds or fractions of seconds), whereas *Gs* reflects rapid responding at intervals of 2 or 3 minutes (Flanagan et al., 2007). "None of the major IQ tests measure *Gt*, although Speed of Information Processing on the DAS and DAS-II—with several sets of very brief scanning tasks rather than one long one—may come closer to *Gt* than do other *Gs* tasks" (J. O. Willis, personal communication, November 2, 2008). Research measures of *Gt* are often included in the kind of reaction time experiments that stretch back to Galton's early sensory-motor tests. These tasks are usually used in investigations of Spearman's *g* theory of intelligence; surprisingly, these tasks correlate substantially with *g* when they involve both decision speed and reaction time (Jensen, 1998).

Long-Term Retrieval (*Glr*)

Glr is a person's ability to store information (either newly learned or acquired in the past) and efficiently retrieve the information from long-term memory. "*Gc, Gq,* and *Grw* represent *what* is stored in long-term memory, whereas *Glr* is the *efficiency* with which this information is initially stored in and later retrieved from long-term memory" (Flanagan et al., 2007, p. 289). *Gsm* measures immediate recall after a few seconds, while *Glr* begins "within a few minutes or hours of performing a task" (Flanagan et al., 2007, p. 289). Several of the *Glr* narrow abilities, such as naming facility, are associated with divergent-production from Guilford's theory, but creativity is virtually buried in the depths of CHC theory in contrast to the featured role it played in Guilford's conception of intelligence. In modern IQ tests, most notably the WJ III and KABC-II, *Glr* is measured primarily by paired-associate learning tasks, as in the KABC-II Rebus subtest (which was illustrated in Figure 3.5 with items from the similar KAIT Rebus Learning subtest). These learning tasks require the person to learn and retain the new information during a "teaching-and-learning" session that lasts for about 10 to 12 minutes. *Glr* over a longer time frame is also measured by delayed-recall tasks. After administering a few more subtests to the person, the examiners give a pop quiz, without warning, to see how much the person remembers. On the KABC-II, the interval is about 30 minutes; on the WJ III, the delayed versions of the learning tasks are given anywhere from 30 minutes to 8 days later.

Speaking of the WJ III, I have to admit that our KABC-II learning subtests are modeled after ingenious tests developed by Woodcock for the original WJ (Visual-Auditory Learning, which inspired Rebus) and for the WJ-R (Memory for Names, which led to Atlantis). Years ago, I asked Dr. Wechsler why he took subtests—sometimes exact test items—directly from the Binet or from the nonverbal tests developed during World War I. He smiled, and said, "There are only 9 commandments for test developers, not 10. The one that is missing is 'Thou shalt not steal.'" He was right. Woodcock's ideas were too good to ignore, because tests of learning ability translate directly to the classroom. Understanding how well children learn is usually the reason we test children in the first place. So we paid Dr. Woodcock the highest form of flattery—imitation! (And other test developers have returned the compliment to us by "borrowing" our novel ideas for tests like Riddles and Spatial Memory.) Imitation also has its benefits: Inclusion of well-researched item types allows examiners to draw on that research history when interpreting new tests.

Tests Built from *Gf-Gc* or CHC Theory

Ultimately, both the Cattell-Horn and Carroll models started from the same point—Spearman's (1904) *g*-factor theory—and ended up with remarkably consistent conclusions about the spectrum of human abilities. That consistency has formed the foundation for most contemporary IQ tests, and for the most prominent, research-based approach to the interpretation of all IQ tests from the CHC model: the cross-battery approach, developed by Dawn Flanagan and her colleagues (e.g., Flanagan & McGrew, 1997; Flanagan et al., 2007), which is rooted in Woodcock's (1990) seminal work. The cross-battery approach urges the selection of tasks from virtually all IQ tests, rather than relying on a single instrument, in order to assess a more complete array of broad and narrow abilities.

The first individually administered comprehensive tests of intelligence to be loosely grounded in *Gf-Gc* theory were the K-ABC (Kaufman & Kaufman, 1983) and the fourth edition of the Stanford-Binet (Binet-IV; Thorndike et al., 1986). As we stated way back when (Kaufman & Kaufman, 1983), "The Achievement Scale resembles closely the crystallized abilities, and the two Mental Processing scales together resemble the fluid abilities that characterize the Cattell-Horn theory of intelligence (Cattell, 1971; Horn, 1968; Horn & Cattell, 1966)" (p. 2). However, as already discussed, the K-ABC was rooted in neuropsychological theory and was only incidentally tied to *Gf-Gc* theory.

The Binet-IV offered a hierarchical model of intelligence (Thorndike et al., 1986): "This model had a general reasoning factor, *g*, at the top level. The second level consisted of three broad factors—crystallized abilities, fluid-analytic abilities, and short-term memory. The third level consisted of three more specific factors—verbal reasoning, quantitative reasoning, and abstract/visual reasoning" (p. 9).

The K-ABC did not do a very good job of translating *Gf-Gc* theory into practice. The K-ABC's separation of intelligence from achievement, which was done primarily for *practical* reasons concerning fairness to children from different ethnic groups, violated the basic premise that *Gf* and *Gc* were two types of intelligence. Furthermore, the K-ABC's measure of intelligence, said to measure *Gf*, really had only a few subtests that measured abstract reasoning ability.

And the Binet-IV did not fare any better. Thorndike and colleagues (1986) based the test on a blend of *g* theory and theories of multiple cognitive abilities, but they failed to disclose exactly which theories were most influential. They used Cattell-Horn terminology for two of their second-level abilities, and one can infer from their historical introduction to the manual that they were also influenced by Thurstone (1938); Guilford (1967); and Hunter, Schmidt, and Jackson (1982). From a *Gf-Gc* perspective, they missed the mark. They subdivided crystallized abilities into two scales: Verbal Reasoning and Quantitative Reasoning, even though the latter scale is known to be more closely aligned to *Gf* than *Gc*. Not surprisingly, the statistical method that identifies the abilities or constructs that underlie a battery of tests—factor analysis—did *not* support the meaningfulness of the Binet-IV scales (Reynolds, Kamphaus, & Rosenthal, 1988). As I wrote about the Binet-IV a few years after it was published (Kaufman, 1990, p. 608), "Had it not been for its venerated name, the new battery probably would have died a quick death, following at least one reviewer's proposal to heed a eulogy proposed previously for the old Binet: 'To the S-B IV, *requiescat in pace*: and so it should have stayed' (Reynolds, 1987, p. 141)."

The real hero in developing a *Gf-Gc*-based test of cognitive abilities was Dick Woodcock. His first comprehensive test battery, the WJ (Woodcock & Johnson, 1977), was deliberately nontheoretical; it was built to address the practical psychoeducational concerns of psychologists, special educators, and teachers, and it included an array of novel measures of intelligence. In a review I wrote of the test, I concluded that the WJ "is a mixture of extremes, possessing some outstanding qualities, yet hampered by glaring liabilities. . . . The [WJ] represents a monumental and creative effort by its authors" (Kaufman, 1985, p. 1762). One of its "glaring liabilities" was the total absence of a theoretical model, a liability that Woodcock walked the extra mile to address. He spent several years at the University of Southern California in order to study directly with John Horn and be mentored by the great man in the nuances of *Gf-Gc* theory. Horn excelled as a mentor (something I learned firsthand when my son James studied with him as an undergraduate at USC). And Dr. Woodcock excelled as Horn's student (so did James).

Woodcock revised his original test so thoroughly that the WJ was barely recognizable in its rebirth as the WJ-R. Woodcock retained, or modified, the original WJ tasks so long as they fitted nicely into *Gf-Gc* theory. The 1986 meeting I mentioned earlier that included Horn, Carroll, and Woodcock, the one that paved the way for future developments in *Gf-Gc* and CHC theory (McGrew, 2005), was part of the elaborate test-development process for the WJ-R. Overall, the revised, theory-based edition of the WJ measured seven of the broad abilities posited by Horn (1989) in his expansion and refinement of the original two-ability Cattell-Horn model. From personal conversations I had with Dr. Horn in the early 1990s, I found that he was clearly impressed with Woodcock's adept translation of theory to practice. Horn would surely have agreed with Esters, Ittenbach, and Han's (1997) review of the WJ-R, stating that "Quite possibly the best and purest example of *Gf-Gc* theory as operationally defined by an IQ test is the [WJ-R]" (p. 212). I marveled: "In particular, it includes fairly pure measures of *Gf* as well as true learning tasks (such as the *Glr* paired-associate subtests) that are basically excluded from Wechsler's system" (Kaufman, 2000b, p. 464). And I also relished Dick Woodcock's comment to me some years ago that my 1985 review of the WJ was a wake-up call that impelled him to action to seek out the best possible theory on which to build the WJ revision.

The WJ-R measured seven of Horn's broad abilities. In addition to *Gf,* referred to by Woodcock as fluid reasoning, and *Gc,* labeled comprehension-knowledge, the WJ-R provided reliable and valid measurement of long-term retrieval (*Glr*), short-term memory (*Gsm*), processing speed (*Gs*), auditory processing (*Ga*), and visual processing (*Gv*). Examiners who administered the complete WJ-R, including the tests of achievement, could also assess an eighth broad ability from Horn's model, quantitative thinking (*Gq*).

This same theoretical structure formed the foundation of the WJ III (Woodcock et al., 2001), but by the time this theory-based test was published, the Cattell-Horn and Carroll systems had been merged and CHC theory provided the theoretical underpinnings of the WJ III. The seven primary broad CHC factors, as they are called on the WJ III, are essentially the same as the WJ-R scales. In addition, administration of the WJ III achievement tests provides measurement of *Gq* and *Grw* (reading and writing). Therefore, the WJ III, in its entirety, measures 9 of the 10 primary broad factors that comprise Stratum II of the CHC model.

Extent of the Influence of CHC Theory on IQ Tests

What started more than a generation ago as an easy analogy with which to interpret Wechsler's Verbal and Performance IQ scales (Matarazzo, 1972) has grown to mammoth proportions in terms of its impact on contemporary IQ tests. As Horn expanded the *Gf-Gc* dichotomy to encompass many broad abilities, it became increasingly clear that Wechsler's Verbal IQ measured not only *Gc* but also *Gsm* and *Gq,* and that his Performance IQ measured more than *Gf,* providing measurement of *Gv* and *Gs;* indeed, Woodcock (1990) argued that older versions of Wechsler's scales measured *Gv* and not *Gf* at all.

That has all changed. New versions of the Wechsler scales include Matrix Reasoning and other subtests measuring *Gf* as well. The WAIS-IV includes three new subtests, as illustrated in Figure 3.13—Figure Weights (*Gf*), Visual Puzzles (*Gv*), and Cancellation (*Gs*). Like the KABC-II's Block Counting, Visual Puzzles is a good example of spatial reasoning. This new WAIS-IV subtest is similar to the Woodcock-Johnson's Spatial Relations subtest, and both, undoubtedly, were inspired by age-old paper formboard tasks that date back to the late 1920s (Roszkowski, 2001).

But even if some new Wechsler subtests have old roots, the look of the latest versions of Wechsler's scales is decidedly new. The two IQ scales (Verbal and Performance) have been replaced by four separate scales, each interpretable according to CHC theory (Flanagan & Kaufman, 2004, 2009). And many current IQ tests are built from CHC theory, including these:

- WJ III (Woodcock et al., 2001; see Table 3.2) for ages 2 to 95+ years
- Stanford-Binet-5 for ages 2 to 85+ years (Binet-5; Roid, 2003; see Table 3.3)
- DAS-II for ages 2½ to 17 years (Elliott, 2007; see Table 3.4)

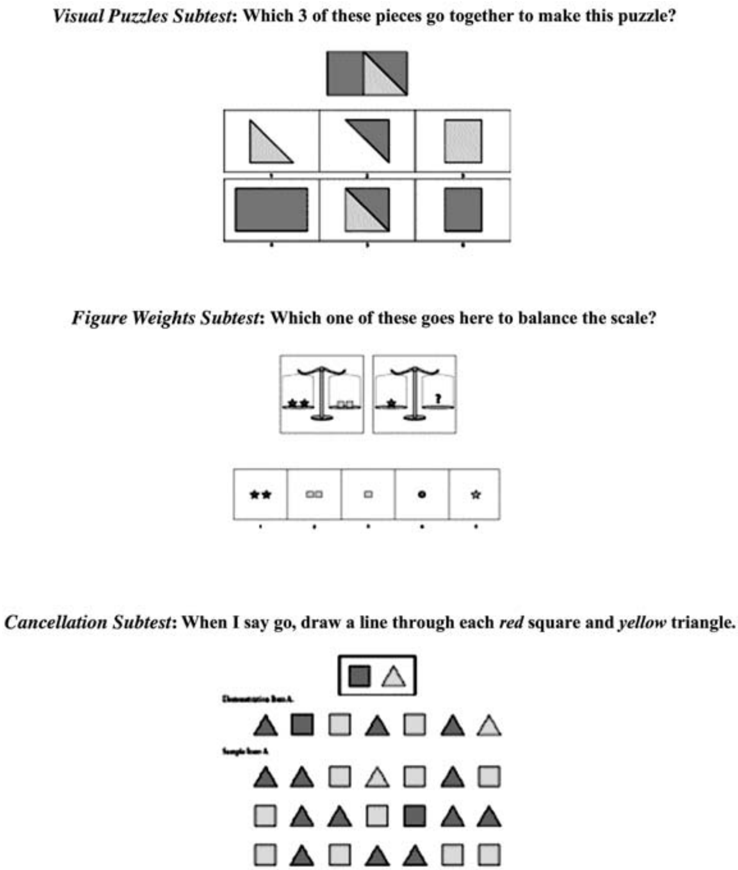

FIGURE 3.13 Three new WAIS-IV subtests.
Note. Sample items similar to items in the *Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV)*, by D. Wechsler, 2008, San Antonio, TX: The Psychological Corporation.

- Reynolds Intellectual Assessment Scales (RIAS; Reynolds & Kamphaus, 2003), which were developed to measure *Gf* and *Gc* efficiently and to provide a separate scale that assesses *Gsm*
- KABC-II for ages 3 to 18 years (Kaufman & Kaufman, 2004a; see Table 3.5, which shows its dual theoretical foundation)

TABLE 3.2 WOODCOCK-JOHNSON III (WJ III) FOR AGES 2–95+ YEARS

- Global Ability: General Intellectual Ability
- CHC Ability Factors:
  - Fluid reasoning (*Gf*)
  - Comprehension-knowledge (*Gc*)
  - Long-term retrieval (*Glr*)
  - Visual-spatial thinking (*Gv*)
  - Short-term memory (*Gsm*)
  - Auditory processing (*Ga*)
  - Processing speed (*Gs*)

TABLE 3.3 STANFORD-BINET-5 FOR AGES 2–85+ YEARS

- Global Ability: Full Scale IQ
- Factor Indexes:
  - Fluid reasoning (*Gf*)
  - Knowledge (*Gc*)
  - Quantitative reasoning (*Gq*)
  - Visual-spatial processing (*Gv*)
  - Working memory (*Gsm*)

TABLE 3.4 DIFFERENTIAL ABILITY SCALES–2ND ED. (DAS-II) FOR AGES 2½–17 YEARS

- Global Ability: General Conceptual Ability (GCA)
- Cluster Scores:
  - Verbal ability (*Gc*)
  - Nonverbal reasoning ability (*Gf*)
  - Spatial ability (*Gv*)
  - Processing speed (*Gs*)—*diagnostic*
  - Working memory (*Gsm*)—*diagnostic*

TABLE 3.5 **KAUFMAN ASSESSMENT BATTERY FOR CHILDREN—2ND ED. (KABC-II) FOR AGES 3–18 YEARS**

Luria Term	CHC Term	Name of KABC-II Scale
Learning ability	Long-term storage and retrieval (*Glr*)	Learning/*Glr*
Sequential processing	Short-term memory (*Gsm*)	Sequential/*Gsm*
Simultaneous processing	Visual processing (*Gv*)	Simultaneous/*Gv*
Planning ability	Fluid reasoning (*Gf*)	Planning/*Gf*
	Crystallized ability (*Gc*)	Knowledge/*Gc*
Mental Processing Index (MPI)	**Fluid-Crystallized Index (FCI)**	

Just as the WJ III Cognitive and Achievement tests were conormed, so too were the KABC-II and the Kaufman Test of Educational Achievement—Second Edition (KTEA-II; Kaufman & Kaufman, 2004c, 2005). The combination of Kaufman tests provides examiners with eight broad abilities. The KABC-II measures five (*Gf, Gc, Gv, Glr, Gsm*) and the KTEA-II measures quantitative thinking (*Gq*), reading and writing (*Grw*), and auditory processing (*Ga*), as well as additional *Glr* narrow abilities.

Indeed, most cognitive tasks can be viewed in different ways and are equally valid from one theoretical perspective as from another. That is one reason why we chose to base the KABC-II on the dual theoretical models of Luria and CHC. Another reason for the dual model is that Nadeen, an astute clinician, is immersed in the clinical nature of Luria's neuropsychological model, a theory that evolved from Luria's clinical work with neurological patients; in contrast, my psychometric, research-based, and statistical orientation is more in tune with the data-driven CHC theory.

For a thorough history of CHC theory, with an emphasis on the WJ-R and WJ III, see McGrew (2005). Also see Flanagan et al.

(2007) for a comprehensive application of CHC theory to the interpretation of all current IQ tests.

Extent of the Influence of IQ Tests on CHC Theory

I've already mentioned the 1986 meeting in Dallas attended by theorists, test authors, and the WJ-R test publisher. Now consider the meeting that took place in 1999 in Chapel Hill, North Carolina, organized by Riverside, the publisher of both the WJ III and the Binet-5. That meeting was attended by authors of the WJ III (Dick Woodcock, Kevin McGrew) and Binet-5 (Gale Roid), two theorists (John Horn and John Carroll), and staff members from Riverside. The goal was "to seek a common, more meaningful umbrella term that would recognize the strong structural similarities of their respective theoretical models, yet also recognize their differences" (McGrew, 2005, p. 149). The net result of that meeting was the merger of the Cattell-Horn and Carroll systems into CHC theory. Talk about the tail wagging the dog! What had begun back in the late 1970s and early 1980s as a search for the best theories on which to build an IQ test had come full circle: Two decades later, the needs of test publishers and test authors forged the theory that underlies almost all current-day IQ tests.

SO WHAT IS THE RIGHT NUMBER OF ABILITIES?

Theory has ultimately merged thoroughly with practice. It has infiltrated all IQ tests and dominated most. What is the "right" number of abilities for an IQ test to measure? Surely not the 1 posited by Spearman or the 120 or more that came with Guilford's territory. And not the 2 that were popular for so long when Wechsler's Verbal and Performance IQs pervaded schools, clinics, and the psychology literature—or the 2 of *g* that characterized the original Cattell-Horn dichotomy, or the 2 mental processes on which the K-ABC was built. Two abilities were not enough.

Wechsler's scales now feature four indexes (see Table 3.6), the same number as the PASS processes in the Luria-based CAS. Glancing over Tables 3.2 to 3.4, we see that the tests built from CHC theory measure five to seven abilities. The KABC-II (Table 3.5), founded on two theoretical models, measures either four or five depending on whether the CHC or Luria model is selected. The CHC model of the KABC-II yields scores on five abilities, whereas the Luria model measures four processes. The difference? We included the Knowledge/*Gc* scale in the CHC model, because *Gc* is such a key ingredient of *Gf-Gc* theory. But tests of factual knowledge and language ability are deliberately excluded from the Luria model, which emphasizes mental processing rather than acquired knowledge and is especially useful for the ethnically fair assessment of children from bilingual and bicultural backgrounds.

So, a contemporary answer to "What is the right number of abilities to measure?" is somewhere between four and seven. These numbers allow examiners to identify important areas of strength and weakness for each person tested. The four abilities measured by Wechsler's scales (Table 3.6)—verbal comprehension, perceptual reasoning, working memory, and processing speed—display distinctly different growth curves as adults travel the rocky road from young adulthood to old age (a hot topic

TABLE 3.6 WISC-IV FOR AGES 6–16 YEARS AND WAIS-IV FOR AGES 16–90 YEARS

- Global Ability: Full Scale IQ (FS-IQ)
- Factor Indexes:
  - Verbal Comprehension Index (VCI) (*Gc*)
  - Perceptual Reasoning Index (PRI) (*Gv-Gf*)
  - Working Memory Index (WMI) (*Gsm*)
  - Processing Speed Index (PSI) (*Gs*)

discussed in chapter 8). Also, the four to seven abilities and processes measured by current IQ tests are in lockstep with the federal definition of specific learning disabilities, which stipulates a disorder in a basic psychological process. But whether or not that disorder really needs to be measured is another story—a hot topic discussed in chapter 9.

DAVID WECHSLER'S LEGACY

As I've discussed at length in this chapter, theory-based tests began to appear throughout the decade of the 1980s, notably the K-ABC, Binet-IV, and WJ-R. All of these theory-based tests have been successful, as have the latest editions of each test, published in the early 2000s. They have met with worldwide success in some instances (e.g., Kaufman & Kaufman, 1993; Melchers & Preuß, 1991).

Nonetheless, Wechsler's scales for children, adolescents, and adults have withstood challenges by the theory-based tests. Though not specifically developed from CHC theory, Wechsler's modern-day tests were specifically revised in the 1990s and 2000s to incorporate CHC theory and state-of-the-art research on working memory and other executive functions. And the most popular interpretations of profiles yielded by Wechsler's children's and adult scales are decidedly CHC in origin (Flanagan & Kaufman, 2004, 2009; Kaufman & Lichtenberger, 2006; Keith et al., 2006). Theory-based tests such as the WJ III, KABC-II, RIAS, CAS, Binet-5, and DAS-II are of high quality and are frequently used, but they mainly serve as members of Wechsler's royal court. Make no mistake about it: More than seventy years after he published his first IQ test and nearly a century since several of his performance tasks were developed by World War I psychologists, the Wechsler scales are the most popular tests in the United States (Prifitera, Saklofske, Weiss, Rolfhus, & Holdnack, 2005) and throughout the world (Georgas, Weiss, van de Vijver, & Saklofske, 2003). David Wechsler is still the king.