

# INTELLIGENT TESTING

This is a book about the Wechsler Intelligence Scale for Children–Fifth Edition (WISC–V; Wechsler, 2014), which, as almost any graduate student in education or psychology knows, is an IQ-yielding intelligence test. But neither the IQ nor the concept of intelligence is the focus of the chapters that follow. The focus is the child, with interpretation of the WISC–V and communication of the test results in the context of the child’s particular background, behaviors, and approach to the test items as the main goals. Global scores are deemphasized, flexibility and insight on the part of the examiner are demanded, and the test is perceived as a dynamic helping agent rather than as an instrument for placement, labeling, or other types of academic oppression. In short, *intelligent testing* is the key, and the WISC–V is the vehicle.

The preceding paragraph introduced Alan’s 1979 *Intelligent Testing with the WISC–R* (A. S. Kaufman, 1979), with the single change being the substitution of WISC–V for WISC–R. That paragraph also summarized Alan’s beliefs about the value of intelligence tests in the 1990s when he wrote *Intelligent Testing with the WISC–III* (A. S. Kaufman, 1994a), even though the field of intellectual assessment had undergone turbulent changes during the 1980s and the “Kaufman method” had been the target of direct attack (e.g., McDermott, Fantuzzo, & Glutting, 1990). And it continues to summarize Alan’s current views, even while staring in the face of the post-Individuals with Disabilities Education Act (IDEA) tsunami of anti-IQ-testing sentiment that makes the controversies of the 1980s and 1990s seem like a mild family feud.

Further, that first paragraph is endorsed by Susie Raiford and Diane Coalson, even though these coauthors of *Intelligent Testing with the*

*WISC–V* were in elementary school when Alan first formulated the philosophy and they cut their teeth as fledgling psychologists on the 1994 edition. For over 15 years, Susie and Diane have been on the firing line and in the trenches as the heirs apparent to David Wechsler, with one or both serving as research directors for every U.S. publication of a Wechsler intelligence scale since 2000, including the WPPSI–III, WISC–IV, WISC–IV Integrated, WAIS–IV, WPPSI–IV, WISC–V, and WISC–V Integrated. Thus, they have weathered the variety of IQ storms as teammates and close colleagues throughout this time. At the core of Susie’s and Diane’s approaches to cognitive assessment (Raiford & Coalson, 2014) is an intelligent testing philosophy that extends back to David Wechsler, the clinician, but is simultaneously mindful of what assessment is all about in the middle of the second decade of the 21st century:

**Diane:** With every revision, my appreciation of Wechsler’s genius is again renewed. By distinguishing his definition of intelligence from the cognitive abilities he used to measure it, he avoided all of the pitfalls so many of us seem to experience when we align too closely with this theory or that assessment method. He knew his tests were, most importantly, *clinical* instruments, designed to measure key, but not all, aspects of an individual’s intelligence. Only in the hands of a skilled clinician (i.e., an intelligent tester), does the instrument become a powerful tool, yielding much more than a handful of scores. This is still true today.

**Susie:** Intelligent testers answer the question that is asked and, where necessary, provide the referral source with the answers to some other questions maybe they should have asked. They select measures that are engaging

to the child, efficient, and reliable; they assess all relevant areas of the child's functioning and avoid the irrelevant. Intelligent testers stay current on issues in the field (listservs), adapt on the fly as the picture of results become clearer, and use the most relevant and current theoretical models, neurocognitive research, and clinical research findings to understand the data as a whole.

**Diane:** Regardless of your choice of test kit or digital administration, your theoretical orientation to assessment and interpretation, or your approach to intervention, intelligent testing is an ongoing, dynamic process. It involves an active interchange between examiner and child, yielding rich, clinical information as it progresses. Modifications or additions to the assessment plan should be expected and welcomed, as clinical hypotheses are accepted or rejected.

**Susie:** Perhaps most importantly, intelligent testers present results as a clear description of the living, breathing child—as opposed to providing tables full of scores that are of no use in the real world—and translate results into recommendations that parents, teachers, and other associated professionals can use and that are effective.

**Diane:** Isn't Wechsler's way still the best way for intelligent testing? I think he would have been so interested in everything going on but quickly gotten back to how it helps the child (or adult).

In fact, David Wechsler—one of the first American clinical psychologists and among the first to open up a private practice—was likely the first intelligent tester within the field of clinical assessment. (Actually, there wasn't even a field of clinical assessment until Wechsler, single-handedly, converted the Binet-Terman tradition of *psychometric testing* to the field that still reigns supreme more than 75 years later.) Witness Wechsler's (1939) caution to clinicians when the first form of the Wechsler-Bellevue was published:

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)

Wechsler's elegant wisdom embodies clinical insight, humanism, and, without question, the epitome of *intelligent testing*.

Importantly, Wechsler viewed intelligence as a component of personality, as something inseparable from a person's affect. Drawing from Aristotle's perception of mental faculties, popularized by Kant, Lohman (1989) stated:

By this account, a complete theory of mind must explain not only the cognitive dimension but also the emotional and intentional dimensions as well. . . . Thus, one direction research on intelligence seems to be taking is to expand its horizons to include affective dimensions long recognized as central to intelligence (e.g., Wechsler, 1939) but rarely combined with the systematic study of the cognitive dimensions. (p. 360)

Putting that concept in simple structure is Dr. Wechsler's credo, spoken to Alan with various degrees of exasperation when he had to deal with Alan's psychometric tendencies: "The Wechsler scales are, first and foremost, *clinical instruments*."

But Dr. Wechsler was far more than a clinician who interpreted intelligence as an aspect of personality. As Alan wrote a few years ago:

Though I worked with Dr. Wechsler for nearly 5 years in the early 1970s during the process of revising the 1949 WISC and developing and standardizing the WISC-R—which was called the WISC (Rev.) in the manual's page proofs until a last-minute decision by an executive rewrote history—I never knew of his psychometric background. To me, Dr. Wechsler was the consummate clinician who deferred to my statistical expertise because I trained at Columbia with Robert Thorndike, Edward's son. I found out a few years later

about his work with psychometric pioneers just after World War I, but during the time Dr. Wechsler mentored me, he never let on about his statistical savvy. I wanted to include, directly in the 1974 WISC-R test manual (Wechsler, 1974a), the exploratory factor analyses of the WISC-R that I later published for normal children (Kaufman, 1975) and individuals with mental retardation (Van Hagen & Kaufman, 1975), and to have examiners compute three factor scores in addition to the three IQs, but he calmly said, “No, not yet; it isn’t time.” (A. S. Kaufman, 2013a, p. 225)

Our continued advocacy of an intelligent testing approach to IQ assessment—even tracing it back to the methods that David Wechsler favored almost a century ago—does not imply that we haven’t changed, because we have, and it doesn’t suggest that the field is static because it is as volatile and energetic and innovative as ever. The three of us retain our beliefs about how to assess intelligence in general and how to interpret the WISC-V in specific. The key is still intelligent testing, as opposed to the mindless testing that never quite disappears. But the context of the IQ construct, both societally and professionally, has altered with time.

## **IQ TESTING IN THE 1970s**

---

The field of intelligence testing is nearly unrecognizable as the field that Alan entered nearly a half century ago as a student of the brilliant-but-distant Robert L. Thorndike at Columbia University. Alan was hired by the Psychological Corporation’s Test Division in late 1968 as a young, idealistic, not-yet-dissertationed, psychologist. Sure the IQ test was at the center of heated controversy—hasn’t it always been?—but the issues and the antagonists were different, and the arguments were more emotional than empirical. When Alan was getting his feet wet in the early 1970s as Dorothea McCarthy’s and David Wechsler’s

right-hand person (though he is left-handed), he began to understand the depth of the feelings of the anti-testing people. At that time, the opponents of the IQ, and of the tests that served this unholy purpose, were mostly enraged about test bias, especially against African American children. They were social psychologists and African American psychologists and sociologists and civic leaders. Some words were tossed around—like “biased,” “unfair,” “middle class,” “discriminatory,” and “racist”—while other words were best tossed away, like “immutable,” “innate,” and “Jensen.”

Within the field of the clinical assessment of intelligence in the late 1970s, the WISC-R was virtually the *only* well-normed, psychometrically sound IQ test for children. “Stupid testing” was rampant at that time. Clinicians interpreted small differences between scaled scores as meaningful and tended to interpret “high” and “low” scores in isolation: A scaled score of 8 on Picture Completion meant that the child had trouble distinguishing essential from nonessential details, an 11 on Comprehension meant good social maturity. Psychoanalytic overinterpretation of Wechsler profiles was popular in the early 1950s: Failing easy Comprehension items conceivably reflected schizophrenia or psychotic depression; decrements in Information, contrasted with adequate Comprehension, indicated a hysterical reaction; increments in Picture Completion suggested a possible paranoid trend (Mayman, Schafer, & Rapaport, 1951; Rapaport, Gill, & Schafer, 1945–1946). Yet this type of nonsensical interpretation remained popular throughout the 1970s: Allison, Blatt, and Zimet (1968), for example, claimed that high Digit Symbol (Coding) and low Digit Span characterizes a person

who seems to be controlling strong and pressing anxiety by excessive activity . . . .  
When we find the reverse pattern, a high Digit Span and a low Digit Symbol, we are usually confronted with an essentially depressed person who is attempting to ward

off recognition of depressive affect perhaps in a hypomanic way, usually via denial, but not necessarily through acting and acting out behavior. (p. 32)

A major purpose of *Intelligent Testing with the WISC-R*—and of other landmark 1970s texts written by special educators (Bannatyne, 1971; Bush & Waugh, 1976; Myers & Hammill, 1976) and psychologists (Lutey, 1977; Matarazzo, 1972; Sattler, 1974)—was to impose some empirical order on profile interpretation; to make sensible inferences from the data with full awareness of errors of measurement; and to steer the field away from the psychiatric couch.

## IQ TESTING IN THE 1980s AND EARLY 1990s

---

When Alan revised his 1979 text on the WISC-R and wrote *Intelligent Testing with the WISC-III* (A. S. Kaufman, 1994a), the opposition to IQ tests was no longer so intensely focused on the unfairness of the tests to ethnic minorities. In the 1970s there was talk of “black intellectual genocide” and “the silent mugging of the black community” (R. L. Williams, 1974a, 1974b). In the 1980s and 1990s, the IQ testing opponents were no longer primarily from outside the field. Now many resided within the field: trainers of school psychologists, developers of new approaches to intellectual assessment, cognitive theorists, psychometricians, and neuropsychologists. The people who viewed the IQ test as an instrument of torture for minority group members were still around, but they spoke with quieter voices. By the early 1990s—and still true today—the critics offered few concessions, and the venom applied to everyone, regardless of socioeconomic or ethnic background.

Whereas the 1970s produced an array of innovative textbooks on IQ assessment, such as Matarazzo’s, Bannatyne’s, Sattler’s, and Kaufman’s original *Intelligent Testing*, the 1980s

witnessed the first group of individually administered tests built from the foundation of *theory*. Theory-based test construction started in the 1980s with the split-brain/Luria foundation of the Kaufman Assessment Battery for Children (K-ABC; A. S. Kaufman & Kaufman, 1983), the Cattell-Horn fluid-crystallized framework of the Stanford Binet Intelligence Scale—Fourth Edition (SB IV; Thorndike, Hagen, & Sattler, 1986) and the Woodcock-Johnson Tests of Cognitive Ability—Revised (WJ-R; Woodcock & Johnson, 1989). These newcomers on the block aroused both interest and controversy, but they did not knock the king of the hill—Wechsler’s scales—off the front pages.

There were two specific key targets of the 1980s and 1990s anti-IQ forces: subtest profile interpretation (especially the “Kaufman method”) and Wechsler’s children’s scales, notably the 1991 Wechsler Intelligence Scale for Children (WISC-III; Wechsler, 1991).

### Criticisms of Subtest Profile Interpretation

Some researchers and clinicians argued that using any type of subtest or profile interpretation was like taking illegal drugs: “Such approaches essentially violate primary principles guiding valid test interpretation” (McDermott, Fantuzzo, Glutting, Watkins, & Baggaley, 1992, p. 522); “we are compelled to advise that psychologists just say ‘no’ to subtest analysis” (McDermott et al., 1990, p. 299). These psychologists based their conclusions on a variety of psychometric analyses that they believed proved their points beyond dispute. They represented the new breed of anti-IQ testing professionals—with a link to the past—Hirshoren and Kavale (1976); but despite their strong words, they were perhaps the mildest of the species. They wanted to kick out subtest interpretation (a practice that Wechsler was devoted to and that the so-called Kaufman method endorsed), but they were okay

with keeping the IQ test and even Wechsler's three IQs.

First they put the Kaufman method securely under the blade of the guillotine:

Perhaps most popular among contemporary practices is the method of ipsative ability assessment advocated by Kaufman (1979) ... He cautioned practitioners not to overvalue IQ scores . . . . A major aspect of this interpretation process is the discovery of children's intellectual strengths and weaknesses by studying the magnitude and direction of each subtest score's deviation from a child's average subtest score. (McDermott et al., 1992, p. 506)

Then they damned his method with faint praise:

The Kaufman method is presently a common element in university curricula for preparing professional psychologists, with the ipsative procedure now generalized to many other ability tests. (p. 506)

And, finally, they interpreted their data with no mercy, letting the blade drop:

Thus we cannot recommend either ipsative or normative approaches for subtest interpretation. Such approaches essentially violate primary principles guiding valid test interpretation. (p. 522)

The *ipsative* interpretation that they criticized refers to an axiom that is implicit in any interpretive system that Kaufman has advocated from 1977 (A. S. Kaufman & Kaufman, 1977) through the present (A. S. Kaufman, 2013a; Lichtenberger & Kaufman, 2013): Identify the child's or adult's *relative* strengths and weaknesses (relative to the person's own level of ability, whether high or average or low).

The McDermott-Glutting critique of the Kaufman method was harsh, but you did not need to look too far to find stronger opposition. D. W. Macmann and Barnett (1994) shared the same psychometric tree as the McDermott-Glutting team, but they went farther out on the

limb. They used exploratory and confirmatory factor analysis to conclude that the Wechsler scales measured little more than *g*, or general intelligence. They weren't content to toss out subtest profile interpretation; Macmann and Barnett (1994) also wanted to discard the Verbal and Performance IQs because the separate factors that underlie these IQs were really nothing more than degraded versions of *g*. They then rode the steam of their empirical argument a little further and decided to chunk the Full Scale IQ (FSIQ) as well. They sought alternative types of assessment but saw no useful role for the measurement of *g* within their system.

McDermott and his colleagues shunned subtest analysis in favor of the global IQs. Macmann and Barnett first flushed the Verbal and Performance IQs and then tossed the FSIQ into the bowl for good measure. In fact, these traditions live on today in the middle of the second decade of the 21st century, with passionate fervor. Gary Canivez, Marley Watkins, and Joe Kush continue to argue that clinicians must say no to profile interpretation (Canivez & Kush, 2013; Canivez & Watkins, 2010a), including interpretation of the factor indexes that have replaced the Verbal and Performance IQs; but they show great respect for FSIQ and the *g* that it measures. Regarding WAIS-IV interpretation, for example, Canivez and Watkins (2010a) concluded from their analyses "that the WAIS-IV provides strong measurement of general intelligence, and clinical interpretation should be primarily at that level" (p. 827). By contrast, advocates of Response to Intervention (RTI) as the sole method for identifying children with specific learning disabilities (e.g., Gresham, 2002; VanDerHeyden & Burns, 2010) are entirely comfortable in relegating IQ tests to the historical relic section of museums.

Not all that much has changed since the early 1990s regarding antagonism against Kaufman's method of profile interpretation. Also contemporary in its flavor is the hold-no-prisoners approach to reviewing IQ tests. Wechsler's 1974 and 1991 versions of the WISC were not exempt.

## Criticisms of the WISC–R and the WISC–III

Witt and Gresham (1985) spoke metaphorically in their Buros test review:

The WISC–R is an anachronistic albatross which hangs gamely around the necks of applied psychologists . . . . Using the WISC–R to assess intelligence in light of the surge of information in [the fields of cognitive psychology and neuroscience] is analogous to applying Newtonian formulae to modern physics problems . . . . The WISC–R lacks treatment validity in that its use does not enhance remedial interventions for children who show specific academic skill deficiencies. In this sense, the WISC–R is biased for all children and for this reason should be replaced with assessment procedures which have greater treatment validity. (pp. 1716–1717)

Edwards and Edwards (1993) ended their very favorable WISC–III review by extending the metaphor: “Individuals who viewed the WISC–R as burdening our profession (Witt & Gresham, 1985) will probably see the WISC–III as nothing more than an albatross that has molted and grown a few new feathers” (p. 149).

Neuropsychologist Muriel Lezak (1988) also took her own potshots at IQ tests, especially Wechsler’s scales (“IQ: R.I.P.”). But she was not concerned about treatment validity and argued the opposite perspective of the proponents of *g*:

When the many and various neuropsychological observations elicited by so-called “intelligence” tests are lumped and leveled into a single IQ score—or even three—the product of this unholy conversion is a number that, in referring to everything, represents nothing . . . [W]e need to conceptualize [mental abilities] in all their multivariate complexity and report our examination findings in a profile of test scores. (pp. 352, 358)

From the vantage point of cognitive psychology, the news still wasn’t so good. John Carroll

(1993b) reviewed the WISC–III and rejected it on empirical grounds. Carroll resurrected Frank’s (1983) diatribe against “the Wechsler enterprise” and did not dispute Frank’s proclamation that it is time for Wechsler’s scales “to become extinct.” Carroll (1993b) concluded, “One can raise the question of whether the revisions and improvements introduced in the WISC–III justify a more favorable judgment of the validity and usefulness of this test” (p. 142). But Carroll was not condemning just the WISC–III; like many cognitive psychologists, he’d vote for extinction of all conventional intelligence tests.

Sternberg (1993) was kinder in his WISC–III review, stating “I do not share the view of some contemporary theorists that the conventional tests are worthless or worse” (p. 163). But he criticized the WISC–III for remaining too static (“Recycling is no longer the exclusive province of environmentalists,” p. 162). And, in his analogy of the WISC to *Rocky*, Sternberg said, “Eventually, we hope, Hollywood will stop recycling material and instead will retire Rocky in favor of a new shining light. Let’s also hope the same happens with the WISC series” (p. 164).

Directly within the field of clinical assessment, the theory-based Kaufman Assessment Battery for Children—Second Edition (KABC–II; A. S. Kaufman & Kaufman, 2004), WJ–R, and the SB–4 began to find supporters, but, on the whole, IQ testing remained a Wechsler establishment. Naglieri (1993) said, “[S]chool psychology in particular, and psychology in general, has relied too much on the Wechsler series and techniques that encourage overinterpretation of scale and subtest variation[.] . . . traditional IQ measures will need to be replaced by more modern ones” (pp. 14–15). The seriousness of these rumblings within the field of school psychology prompted Shaw, Swerdlik, and Laurent (1993) to warn that “the WISC–III could be rendered irrelevant in the schools in a short time” (p. 158).

Where do we stand today regarding the criticisms of 25 years ago leveled at Kaufman’s method of subtest profile interpretation and at

Wechsler's children's scales? What is our stance on intelligent testing with the latest version of the WISC as we approach 2020? The answers to those two questions frame the remainder of this chapter and, in fact, the remaining chapters of this book.

## IQ TESTING TODAY

---

Even though the critics of IQ testing in the 21st century are speaking virtually the same words spoken by the critics of yesteryear, the cognitive ability measures of today are not your grandparents' IQ tests. By now, astute readers will have noted the variety of terms for these measures (e.g., IQ test, cognitive ability test, intelligence test) and how they often are used interchangeably throughout this text. Regardless of the term used to describe a particular measure, they are all measures of cognitive abilities that have been shown to be critical aspects of intelligence. However, in the authors' opinion, it takes an intelligent tester to translate results from such an instrument into a meaningful estimate of an individual's *intelligence*. An adequate description of an individual's intelligence requires much more than a description of cognitive abilities—a recurring theme in intelligent testing today, as well as in the past. Relative to the psychologists of yesteryear, clinicians of today are armed with a plethora of reliable and valid cognitive ability measures as well as an increased understanding of the complex nature of intelligence.

The role of theory has escalated exponentially in the development and interpretation of cognitive ability tests during the past generation. Following the K-ABC, WJ-R, and SB-4 in the 1980s was a proliferation of instruments and techniques that continued nonstop during the 1990s and early 2000s. The 1990s featured Colin Elliott's (1990) Differential Abilities Scale (DAS)—an American version of his popular British Ability Scales (BAS; Elliott, 1983)—and Jack Naglieri's Cognitive Assessment System

(CAS; Naglieri & Das, 1997). The 2000s produced a flurry of high-quality revisions, each founded in well-researched and clinically important theory: the Woodcock-Johnson-Third Edition (WJ III; Woodcock, McGrew, & Mather, 2001), the Stanford-Binet Intelligence Scales, Fifth Edition (SB5; Roid, 2003), the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV; Wechsler, 2003), the Reynolds Intellectual Assessment Scales (RIAS; C. R. Reynolds & Kamphaus, 2003), the KABC-II (A. S. Kaufman & Kaufman, 2004), and the DAS-II (C. D. Elliott, 2007). The year 2014 was marked by publication of three major revisions to comprehensive cognitive ability measures: namely the CAS2 (Naglieri, Das, & Goldstein, 2014), the WJ IV (Schrank, McGrew, & Mather, 2014), and the WISC-V (Wechsler, 2014). These three cutting-edge tests are built on separate but clearly intertwining theoretical foundations.

### Applying Theory to Test Construction

Before describing the role of theory in test construction, it is helpful to be reminded of the definition and nature of theory, in general. As with *intelligence*, the term *theory* has many definitions, depending on the context and field of study. Kerlinger's (1986) oft-cited and generally accepted definition describes theory as “a set of interrelated constructs (concepts), definitions, and propositions that present a systematic view of phenomena specifying relations among variables, with the purpose of explaining and predicting the phenomena” (p. 9). Thus, theory provides important descriptive, organizational, and predictive functions. Theories are often represented by (and confused with) *models*, which are simplified graphical or figural representations of key theoretical concepts and relationships.

Since the time of Aristotle, theory has been contrasted with practice and appreciated as a close partner to research. The complex, interactive nature among theory, research, and

practice was eloquently described by Schneider and Flanagan (2015):

Research can be particularly cruel to theory, theory makes impossible demands of practice, and practice can go for long stretches acting as if research is not even in the room. Yet, they are family and they need each other. Research surprises theory with thoughtful gifts, theory gives sound advice when practice is troubled, and practice helps research to get out of the lab and meet people. (p. 317)

Despite the obvious necessity of theory, it is wise to remember some of its troublesome characteristics as it applies to test construction. Theories are never proven: They are ever-changing. Definitions (i.e., terminology and nomenclature) and relations among the concepts are continually revised according to ongoing research and practice. Examples of these ongoing terminology changes are evident in both the WISC-V (e.g., the Perceptual Reasoning Index has been separated into the Visual Spatial Index and the Fluid Reasoning Index) and the WJ IV (e.g., *Gsm* renamed as *Gwm*). Despite attempts to describe these changes in test manuals, they often cause confusion to users, especially during the transitional period after a newly released revision.

There are many types of theories related to cognitive ability and intelligence. Sternberg refers to five types of intelligence theories in the human intelligence entry he authored for *Encyclopædia Britannica* (2015), including psychometric theories (e.g., Cattell-Horn-Carroll [CHC] theory), cognitive theories (e.g., information-processing and working memory theories), cognitive contextual theories (e.g., Gardner's [1983] multiple intelligences theory), biological theories (e.g., neuroanatomical and behavioral genetics theories), and developmental theories (e.g., neurodevelopmental theory and Piaget's development theory). The sheer number of theory types (each with one or more theories as exemplars) attests to the fact that the study of intelligence and cognitive abilities continues

to be of critical importance across a number of related fields. Just as intelligent testing demands that clinicians stay abreast of theoretical changes and advances in areas of related research and clinical practice, intelligent test construction demands the same.

### Applying Theory to WISC-V Construction

As research directors for the most recent revisions of the Wechsler intelligence scales, Susie and Diane did not have the good fortune to work directly with Dr. Wechsler. The weight of this responsibility was keenly felt, especially considering the numerous assaults on IQ testing in general and, in particular, on the “antiquated” and “atheoretical” Wechsler scales. This perception has been difficult to overcome in some circles, despite substantial revisions to his scales in response to ongoing advances in intelligence theory, research, and clinical practice.

**Diane:** Wechsler was not atheoretical: He was all-theoretical. He was well aware that his tests were used by an increasingly diverse group of clinicians to address an ever-growing menagerie of referral questions. He was familiar with the intelligence theories of his time, as well as advances in psychometrics and relevant areas of research. Just as he was throughout his lifetime, I believe he would have been fascinated by and responsive to contemporary advances in psychometrics, intelligence theory, and neuropsychology, as well as the explosion of converging research in such areas as cognitive neuroscience, gene mapping, and brain imaging—but then quickly narrowed his efforts to incorporating in his scales those aspects that provided the most clinically relevant information for evaluating an *individual's* intelligence.

Confirmation of Wechsler's open-mindedness to competing intelligence theories was directly expressed in transcripts from a 1975 interview with David Wechsler conducted by Roger

Lennon and Jerome Doppelt of The Psychological Corporation at Wechsler's apartment in New York City. When asked to describe conversations regarding the definition of intelligence held at a 1921 symposium with such illustrious attendees as Thorndike, Terman, and Thurstone, Wechsler recalled:

the interesting thing is that there were 14 different opinions. So people got very scared. But I wasn't scared, I wasn't even hurt by it. It simply proved to me that intelligence is a multi something, it isn't one thing. And as you discovered, depending on your area of interest or specialization, you favor one or another definition ... I think they're all right, but not a single one suffices. (Wechsler, Doppelt, & Lennon, 1975, pp. 30–31)

At the time Dr. Wechsler developed his first intelligence scale in 1939, the primary theoretical controversy involved the historical debate between Spearman, a proponent of *g*, a general factor of intelligence, and Thorndike, who denounced *g* in favor of distinct types of intelligence. Dr. Wechsler's (1939) definition of intelligence, which remains unchanged in the WISC–V, specifically supports the idea that intelligence is composed of both global and distinct abilities. This idea is entirely consistent with contemporary CHC theory, which did not exist in its current form until almost 50 years after Wechsler penned his time-tested definition of intelligence as:

the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his [or her] 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. (p. 3)

It should not be surprising that the hierarchical WISC–V structure aligns closely with some of the broad and narrow domains of intelligence as categorized by CHC theory. Subtests with

long, preexisting histories on the Wechsler scales had previously been categorized according to the CHC model of intelligence (Flanagan & Kaufman, 2009; Flanagan, Ortiz, & Alfonso, 2013). Subtests were specifically revised (e.g., Digit Span), created (e.g., Matrix Reasoning), or replaced (e.g., Object Assembly replaced with Visual Puzzles) to improve or extend measurement of those CHC domains with the most clinically relevant evidence of validity for the numerous purposes of the test. Measures that look very similar to subtests selected for Wechsler's original (1939) intelligence scale (e.g., Information, Block Design, Digit Span) continue to appear in more recent intelligence measures with reports of closer ties to specific theory (e.g., KABC–II, CAS2, WJ IV).

Like Wechsler, the research directors responsible for the posthumous revisions of his intelligence scales are cautious about introducing revised or new content until its inclusion is supported by theory, research, and practice. The contributions of CHC theory cannot be understated: It is clearly the most comprehensive *structural* theory of cognitive abilities at the time of this writing. However, it is not the sole or primary driving force behind revisions of the Wechsler scales. To construct the most clinically useful measure of intelligence, it is, in our opinion, important to consider other theories related to intelligence and cognitive ability, especially those that are more *functional* in nature, including processing theories from the fields of cognitive psychology and neuropsychology, as well as various theories related to more specific aspects of cognitive ability, such as working memory, attention, and executive function. The importance of understanding how the components of intelligence function together was noted by Wechsler (1975):

[T]he attributes and factors of intelligence, like the elementary particles in physics, have at once collective and individual properties, that is, they appear to behave differently when alone from what they do when operating in concert. (p. 138)

A closer look at the three most recent publications of comprehensive intelligence measures reflects this shift in emphasis from structure to function—one that is moving beyond the impossible and unfruitful task of attempting to build a usable, completely comprehensive measure of cognitive abilities toward one that focuses on how these components are functionally interrelated when processing information and how this information can be used to develop effective interventions. Construction of the CAS2 (as well as its predecessor) was guided by the Luria-based PASS model of intellectual functioning, which focuses on measurement of cognitive processes (i.e., planning, attention, simultaneous processing, and successive processing) rather than specific broad or narrow cognitive abilities. A growing body of clinical evidence for the CAS and CAS2 suggests that an evaluation of cognitive processes may be sensitive to such conditions as attention-deficit/hyperactivity disorder (Naglieri, Goldstein, Iseman, & Schwebach, 2003; Van Luit, Kroesbergen, & Naglieri, 2005), in which the processing of information may be disrupted despite adequate cognitive abilities. Previous research using the CAS also supported interpretation from a CHC perspective (Keith, Kranzler, & Flanagan, 2001). Based on the relatively minor changes in the CAS2, it is likely that CHC-oriented clinicians will continue to be comfortable using the CAS2. The move from a structural to a more functional emphasis is also evident in the WJ IV (albeit more subtle than that of the CAS2 or WISC-V), which notes the influence of “other venues of research” on test development, including research from the field of neuroscience and on working memory (LaForte, McGrew, & Schrank, 2014, p. 2). This recent change to the WJ IV theoretical foundations makes it remarkably similar to that described in recent revisions of the Wechsler intelligence scales (i.e., WAIS-IV, WPPSI-IV, and WISC-V).

The most recent revisions of Wechsler’s intelligence scales have also been influenced by the exponential increase in research from the related

fields of cognitive neuroscience and functional brain imaging (see Chapter 10 for additional information). Additionally, intelligent test construction demands an ongoing evaluation of research related to the primary uses of the test (e.g., identification of specific learning disorder (SLD) or intellectual disability; developmental and socioeconomic risk factors for low cognitive ability or achievement; age, gender, and cultural differences in intelligence). Thus, the theoretical basis for development of the Wechsler intelligence scales (as well as the theoretical foundations for the WJ IV) would be more correctly termed a *theoretical framework*, with content reflecting aspects of both structural (e.g., CHC theory) and functional (e.g., neuropsychological processing theory) theories of intelligence or cognitive ability as well as current, relevant research related to theory and practice.

Anyone who develops psychological tests knows that construction is also guided by practical issues, such as psychometric qualities, examiner and examinee preference, testing time, and clinical utility. With so many interacting factors influencing development, test construction can be a humbling experience. Many subtests, scores, and other seemingly novel ideas must be abandoned on the road to publication, for any one or more of these reasons. It is tempting for intelligence test publishers, authors, developers, researchers, and even practitioners to emphasize the differences among the tests’ content or theoretical foundations rather than their similarities. We are all attempting to measure cognitive abilities, more specifically, those comprising intelligence. In addition to the wide variety of currently available intelligence measures, numerous other instruments are also available to appraise intelligence-related constructs, including executive function, attention, memory, and language development. Thus, the field of clinical assessment abounds with reliable and valid measures of intelligence and related cognitive abilities. Regardless of the underlying theory or theories that serve as the basis for test construction, or the reported nature of a test’s measured constructs,

it is the intelligent testers' responsibility to truly know the "tools" they are using.

### Applying Theory to Profile Interpretation

Just as impressive as theory-based test construction has been the emergence of theory-based profile interpretation during the past two decades. Most prominent have been the birth and expansion of the innovative cross-battery assessment technique, steeped both in CHC theory and the intelligent testing tradition (Flanagan, Alfonso, & Ortiz, 2012; Flanagan et al., 2013; McGrew & Flanagan, 1998); and the clinical and educational applications of neuropsychological processing theories to cognitive assessment (Kaplan, 1990; Korkman, Kirk, & Kemp, 1998, 2007; McCloskey, Whitaker, Murphy, & Rogers, 2012; Naglieri & Das, 1997; Naglieri et al., 2014).

Certainly theory was prominent from the time Alfred Binet first developed his test in Paris in 1905, almost synchronously with Spearman's (1904) emphasis on *g* theory. Opposite Spearman were the well-reputed, highly publicized "multiple abilities" research and theory by Thurstone (1938; Thurstone & Thurstone, 1941) and Guilford (1956, 1967). But these multiple ability theorists made their mark only in group-administered IQ tests, never managing to cross the clinical barrier and failing to make a dent in how clinicians interpreted the WISC and WAIS. Far more influential than Thurstone and Guilford was Jacob Cohen's groundbreaking factor-analytic research on Wechsler's scales (Cohen, 1952a, 1952b, 1957a, 1957b, 1959). Cohen shifted focus from subtest analysis to the interpretation of a few specific abilities; he demonstrated that the verbal comprehension and perceptual organization abilities did *not* correspond to Verbal and Performance IQs; and he emphasized the importance of a third factor, which he variously called memory or the infamous freedom from distractibility. As Alan wrote,

"Cohen's factor analyses in the 1950s—endorsed by Wechsler and integrated into his own clinical approach to interpretation—changed the way clinicians viewed subtest profiles and provided the foundation for present-day analyses" (A. S. Kaufman, 2013b, p. 228).

If theory-based test interpretation of Wechsler's scales began to take form with Cohen's empirical research, Matarazzo (1972) clearly gave wings to the movement by recognizing the analog between Wechsler's Verbal and Performance scales and the Cattell-Horn constructs of crystallized (*Gc*) and fluid (*Gf*) intelligence. That simple beginning led to more complex interpretation systems based in CHC theory, which underlies cross-battery assessment (Flanagan et al., 2013) and most contemporary IQ tests; and neuropsychological processing theories (McCloskey, 2009b; McCloskey et al., 2012), which have provided the foundation for other widely used tests (Korkman et al., 2007; Naglieri et al., 2014). And the blend of CHC theory, processing theories (e.g., neuropsychological processing theories; working memory, executive function, and fluid reasoning theories; as well as Edith Kaplan's process approach to interpretation), plus relevant research from related fields (e.g., cognitive neuroscience and clinical applications) together form the theoretical framework of the WAIS-IV, WPPSI-IV, and WISC-V (Wechsler, 2008, 2012, 2014; Weiss, Saklofske, Coalson, & Raiford, 2010).

Cohen's pioneering research was oddly prescient of the five CHC-based factors that state-of-the-art confirmatory factor analysis (CFA) has identified for the WISC-IV and WAIS-IV (Keith, Fine, Taub, Reynolds, & Kranzler, 2006; Niileksela, Reynolds, & Kaufman, 2012; Ward, Bergman, & Hebert, 2012; Weiss, Keith, Zhu, & Chen, 2013a, 2013b), and which correspond to the five indexes yielded by the WISC-V. Cohen's factors were: Verbal (Vocabulary, Similarities, Information, Comprehension), Nonverbal or Perceptual Organization (Object Assembly, Block Design, Mazes), Memory or Freedom from Distractibility (Arithmetic, Digit Span),

Picture Completion, and Coding/Digit Symbol. In CHC language, these factors measure *Gc*, *Gv*, *Gsm*, *Gf*, and *Gs*. In the WISC–V vernacular, the old Cohen factors reflect Verbal Comprehension, Visual Spatial, Working Memory, Fluid Reasoning, and Processing Speed. A bull’s-eye for probably the least appreciated pioneer in mental testing.

The field of clinical assessment has come full circle, forming a rising spiral that continues to ascend. Yet while this ascent in the field of IQ testing has been steady for 30 years and has achieved heights not even dreamed about in the past (when the hue and cry was to ban IQ tests, not improve them), the IQ test has never before been so easily disposed of by a segment of society. It seems unconscionable to eliminate IQ from the diagnosis of an SLD when the essence of SLD, conceptually, is a discrepancy between a child’s potential to learn and that child’s school achievement. Yes, state departments of education misused IQ tests, and they glorified formulas that had no place in psychology or education. But RTI is an overreaction to the problem. We are at a point in time that some states are RTI only, and many others are trying to establish sensible guidelines about determining a child’s pattern of strengths and weaknesses (PSW). Intelligent testing is all about PSW and putting that information about a child’s profile into action to determine the best educational interventions. The WISC is no longer *Rocky*; the WISC–IV was a giant step forward from its earlier editions, and the WISC–V has raised the bar once again.

## THE WISC–IV AND THE WISC–V

---

Unlike the WISC–R and the WISC–III, the WISC–IV abandoned the age-old Wechsler distinction between the Verbal and Performance IQs and relied solely on the FSIQ as a measure of *g*. Four factor-based indexes, two composed of Verbal subtests and two of Performance subtests, replaced the Verbal–Performance IQ discrepancy, and a new WISC age dawned.

The WISC–IV represented “the most substantial revision of any Wechsler scale to date... Many of the ways in which the instrument has been altered and restructured are considered strengths” (Flanagan & Kaufman, 2009, p. 202). The most notable strengths were inclusion of better measures of fluid reasoning and working memory, elimination of complex and unwieldy composites (like the Verbal and Performance IQs and the Freedom from Distractibility factor) that defied easy interpretation, and emphasis on constructs that were in lockstep with contemporary research and theory. Even the FSIQ, which traditionally was composed of the same basic 10 or 11 subtests that made up the Wechsler-Bellevue Form I (and every WISC or WAIS FSIQ since 1939) had a new look on the WISC–IV. “The FSIQ has changed dramatically in content and concept and barely resembles the FSIQ of previous WISCs. It includes only 5 of the traditional 10 subtests” (Flanagan & Kaufman, 2009, p. 30).

But if the WISC–IV left a lot of its historical roots behind, then the WISC–V is a game changer. Availability of digital administration format using Pearson’s Q-interactive™ (Qi) platform and an iPad-to-iPad Bluetooth connection affirms Wechsler’s posthumous entry into the digital-technology generation. A seven-subtest FSIQ. Five indexes that match five key CHC abilities. An ancillary global index, the five-subtest General Ability Index, which greatly reduces the working memory and processing speed demands from the *g* equation for children who, for example, have behavioral issues such as anxiety or distractibility or cognitive difficulties in these areas (e.g., many children with SLD). Four other ancillary index scores to measure quantitative reasoning, auditory working memory, nonverbal ability, and cognitive proficiency. Three complementary index scores to measure diverse aspects of the CHC ability Long-term Retrieval (*Glr*): the Naming Speed Index, the Symbol Translation Index, and the Storage and Retrieval Index. The Symbol Translation subtests, which resemble the Woodcock and

Kaufman measures of paired-associates learning, answer the question that has often been leveled at Wechsler's scales: *Why do the most popular IQ tests in the world fail to measure something as basic and essential as a person's ability to learn new material?*

And the list of subtests is Woodcockian and Reitanian in scope: 21 subtests, eight of them new to the WISC; 10 Kaplanesque process scores; 10 error scores (e.g., errors made on Coding while rapidly copying symbols). Picture Completion is gone (joining Wechsler's favorites, Object Assembly and Picture Arrangement, which disappeared when the WISC-IV was published). Clinicians will need to be intelligent, indeed, to interpret the WISC-V intelligently!

## **BASIC TENETS OF THE NEW INTELLIGENT TESTING APPROACH**

---

The tools are in place, whether clinicians are oriented toward the WJ IV, the CAS2, the WISC-V, or any other intelligence measure. The theories are in place. Wisdom abounds in the domains of cognitive assessment and educational intervention from a CHC framework (Flanagan et al., 2013; Mather & Wendling, 2012; Schneider & McGrew, 2012; Wendling & Mather, 2009) and from neuropsychological viewpoints (Fletcher-Janzen & Reynolds, 2008; Shaywitz & Shaywitz, 2013). Despite the movements to eliminate IQ tests from the diagnostic process, the ongoing attempts to integrate iPad technology with intelligence testing (e.g., Pearson's Q-i system or PAR's iConnect), and the inevitable time when someone important decides that clinicians are irrelevant, intelligent testing needs to stand tall for contemporary psychologists. The philosophy needs to be reconceived for the present, consistent with the mandates and implications of IDEA (McBride, Dumont, & Willis, 2011), taking into account the needs of unique learners (Mascolo, Alfonso, & Flanagan, 2014), and with an eye on the changing landscape of the future.

Like the theories that guide test construction and interpretation, the basic tenets of intelligent testing cannot remain static. They changed from Alan's 1979 WISC-R text to his 1994 WISC-III text and continued to evolve as the WISC and WAIS were transformed into their fourth editions (Flanagan & Kaufman, 2009; Lichtenberger & Kaufman, 2013). That evolution continues in this WISC-V text.

Certainly the leaps in theory-based models of test construction and profile interpretation rendered certain aspects of Alan's intelligent testing approaches for the WISC-R and WISC-III obsolete. Alan's notion of grouping subtests according to their "shared abilities"—when some of those abilities were armchair concoctions like "visual perception of complete meaningful stimuli" or "integrated brain functioning" or "culture-loaded knowledge"—is fanciful. So, too, is listing the so-called influences that affect children's performance on each subtest, such as "ability to respond when uncertain" or "alertness to environment" or "negativism" or "outside reading." These abilities and influences filled a gap during the last generation but are hopelessly naive and devoid of research support as we approach 2020. Joel Schneider (2013a) expressed these sentiments kindly:

It is important to group subtests by theory, not by intuition. As brilliant as Kaufman's (1994a) Intelligent Testing system was, there is little evidence that any of the hundreds of subtle influences on test performance he lists has ever been measured with any reliability ... Stick with a well-validated interpretive framework such as CHC theory. (p. 314)

Schneider (personal communication, October 20, 2014) expanded on his quote about the Kaufman intelligent testing system in an email to Alan:

As is often the case, I have a more nuanced opinion about that quote than a literal reading of my words would suggest . . . I believe that it is possible to go well beyond a cookbook interpretation of scores, but different cooks

will produce results of widely varying quality once they stray from the recipe. Since we have a profession in which we rarely experience any consequences for producing incorrect interpretations, it is hard to give clear advice on how to deviate from the cookbook. I think that *I can ... but then everyone else does too.*

While in graduate school, I made spreadsheets that automate your interpretative approach. The spreadsheets calculate the dozens of composite scores suggested by your 1994 *Intelligent Testing* book. I quickly came to appreciate how easy it was to get lost in the data. Using the spreadsheet, I found that a post-hoc discovery of which cognitive processes were shaping the cognitive profile was both too easy and too difficult. It was easy to find pairs of contrasting scores that were significantly different but it was difficult to find ways of convincing myself that I had found an explanation that was trustworthy.

My position is evolving on this issue. When I wrote the chapter (Schneider, 2013a) ... my own process was too prone to error and led to too many wild speculations. I needed to fly lower to the ground. Later I invested time and effort to producing methods that would give me a better balance of liberties and constraints. . . . These methods help me feel a little surefooted about my interpretations and a greater awareness of when I should be appropriately uncertain. Even so, I would like to learn about (or develop) much better methods than the ones I use now.

Schneider and Flanagan (2015) put the original intelligent testing approach to theory into an appropriate time capsule, in view of the great theoretical advancements in the 1990s and 2010s:

The early writings of Sattler (1974) and Kaufman (1979) were extremely persuasive in showing that it was worthwhile to use psychometrics to extract meaningful information from test scores beyond the *g* factor. Not only were Sattler and Kaufman persuasive and practical, their comprehensive and “intelligent” approach to intelligence test interpretation made the process interesting and meaningful to practitioners. Sattler’s and

Kaufman’s early writings were not purely psychometric, nor were they atheoretical, but compared to later efforts (including their own), they were less theoretically driven. (p. 320)

## CONTEMPORARY THOUGHTS ABOUT INTELLIGENT TESTING

---

Consistent with these sensible criticisms of the original intelligent testing method of profile interpretation, and with a continued search for improved methodologies, the authors sought to create an intelligent testing philosophy that was more compatible with the new wave of well-validated theoretical models, current research related to intelligence, and the contemporary assessment scene. Alan reached out to about a dozen highly respected experts in the field. He wrote in a group email dated October 7, 2014:

I am about to start writing the first chapter on the intelligent testing philosophy and I realize that I am no longer certain what best defines that philosophy in 2015, as the field of assessment is so in flux. It has been much influenced by RTI, PSW, blogs like the CHC and NASP listservs, and the field must now accommodate to the unknown potential of the computer’s impact. Pearson’s Q-Interactive is available for a number of their instruments, but many practitioners and school systems are still deciding when to turn digital.

I would very much like for each of you to send me sentences, paragraphs, bullet points, PowerPoint slides, or anything that communicates your notion of the most salient aspects of intelligent testing, both for today and “tomorrow.” Although the book is about the WISC–V, chapter 1 is a philosophy that applies just as well to the WJ IV or the KTEA–3 or any individual test of intelligence or achievement.

The personal communications of these experts were highly influential in refining and illustrating the specific tenets of the intelligent testing philosophy presented in this WISC–V text, modifying and expanding on previous sets

of guidelines (A. S. Kaufman, 1979, 1994a; Lichtenberger & Kaufman, 2013, pp. 119–127). Despite significant advances in the study of intelligence and cognitive abilities, these communications revealed that the intelligent testing philosophy of today is remarkably consistent with that described in the past, due to its child-centered focus. Thus, the tenet revisions are relatively minor, primarily consisting of clarifications based on interim research findings and an increased emphasis on the importance of linking theory, research, and practice to effective interventions. *The core concept of intelligent testing continues to rest on the notions that test interpretation is complex and subtle, that sound theory must underlie profile interpretation, that interventions should walk hand in hand with theory-based interpretation, and—bottom line—that the burden is on test users to be “better” than the tests they use.*

Concern about the present state of intelligence testing and uncertainty about its future produced cautionary words from some of the experts. (These unreferenced personal communications are from email communications with Alan between October 2014 and March 2015.)

**Ron Dumont**—The debate over IQ testing and its utility seems to me to be not about the tests themselves, but in how they are used or misused, and your book is addressing just those issues. I see many folks who give the tests but totally misuse them. It is unfortunate, but many don’t understand the basics of good, intelligent testing.

**Cecil Reynolds**—I am just not sure what to say about the intelligent testing philosophy at this point given the tremendous flux in the field. I think it will easily be 2 years or so to see where the field is trending—RTI has been a huge drag on everything in standardized assessment and neuropsychologists in private practice are very frustrated about their evaluations being essentially turned away as irrelevant to a lot of RTI-adopting school districts.

**Jack Naglieri**—I am convinced by recent papers showing WJ is a one-factor test, the lack of evidence for cross battery assessment, the

expansion of the number of subtests in WISC–V, the over-emphasis on subtest analysis, the illusion that subtests can be combined across tests without regard for differences in standardization samples, the view that someone can look at a subtest and decide what it measures, the misuse of factor analysis of WISC subtests old and new to decide the structure of intelligence, and the over-interpretation of tests from a “neuropsychological” perspective that our field has gone down a path that will not help children.

**Dan Miller**—I believe the field of school psychology is at a perilous crossroads due to the push towards computerized assessment. I must first say that I am a “techie” and I love all of my gadgets (e.g., tablet, smart phone, etc.) so I have a natural affinity for the idea of computerized assessment. However, I am concerned that the current implementation of computerized assessment will hurt the practice of psychology in general, and the practice of school psychology in particular . . . My first and foremost concern is that [in my opinion] the major test publishers have all adopted propriety software that work on tablets or cloud-based servers which are mutually exclusive to their own product lines and do not communicate with each other . . .

As an example, Pearson has developed their Q-interactive platform that allows practitioners to administer and score many of their assessment instruments using two iPads, one for the examiner and one for the examinee. This includes the ability to administer and score a complete WISC–V test of intelligence using the iPads. From a technological standpoint the Q-interactive product has a strong “cool factor” and ease of use. The WISC–V subtests look great on the iPad and the developers have done an excellent job of using the iPad features to facilitate administration and scoring . . . [However], the cost factor is a major concern for users of the Q-interactive system . . .

In summary, school psychologists run the risk of having their professional judgment further constrained because of budgetary constraints related to the costs of computerized assessment. School psychologists and school

neuropsychologists who are taught to use cross-battery assessment are the biggest risk of having their skills underutilized in the future.

**Diane Coalson**—With regards to the computer/digital transition on intelligent testing: Personally, I am still a little nervous about the digital transition. I think we need additional evidence of clinical validity for special groups. I am concerned that the field is prematurely enamored by the medium, similar to results from a recent study in which increased ratings of credibility resulted when bogus brain images were imbedded in a research article. Other research indicates that the mere mention of cognitive neuroscience (as in brain-training commercials) also increased participants' judgment of credibility. Despite such reservations, I believe that the integration of technology and psychological assessment ultimately will lead to substantial improvements in clinical utility and ease of use.

**Ron Dumont**—I think the future will include much more technology. The iPad versions are certainly taking steps to change how we administer tests, but I imagine a whole new set of test measures designed specifically for the technology. I imagine a test that has algorithms that branch to items depending upon the answer/response to each single item. This will create a fine-tuned assessment that is time efficient and very specific to the individual.

With these cautions in mind, with an eye on where the intelligent testing philosophy began, and with great reliance on the ideas and opinions articulated by brilliant clinicians, here are the latest tenets of intelligent testing.

### 1. Intelligent Testing with the WISC-V Requires Training, Experience, and Flexibility

This seemingly obvious statement represents a new addition to the intelligent testing tenets. The subject matter of this tenet was addressed in previous versions of the intelligent testing philosophy, but the number of expert comments on this

topic suggested it had earned its place as the first tenet of intelligent testing. Therefore, it seems most appropriate to begin this new tenet with a favorite quote by Meyer and colleagues that was submitted by Kevin McGrew in response to Alan's query about intelligent testing:

Tests do not think for themselves, nor do they directly communicate with patients. Like a stethoscope, a blood pressure gauge, or an MRI scan, a psychological test is a dumb tool, and the worth of the tool cannot be separated from the sophistication of the clinician who draws inferences from it and then communicates with patients and professionals. (Meyer et al., 2001, p. 153)

There's no question, individualizing WISC-V interpretation requires effort. It is fairly easy to look at the array of indexes and scaled scores and come up with some predictable statements about the child's general intellectual functioning and specific strengths and weaknesses (well, maybe not so easy with the bucket-load of scores yielded by the WISC-V). This type of cookbook interpretation is not compatible with the intelligent testing approach or intent of federal legislation, and has led to the insistence on accountability (in the form of treatment validity) by school psychology trainers, state education departments, and IDEA in general (McBride et al., 2011).

**John Willis**—Just as we claim to believe that each examinee is unique, each evaluation should be unique, driven not only by characteristics of the examinee but also by the examinee's circumstances and by referral questions. Intelligent testing is not just a matter of administering, scoring, and reporting the results of a test or battery of tests, but a thoughtful analysis of the findings within the broader context of the examinee's history and the referral questions. For me, intelligent testing demands heightened mindfulness in planning, conducting, interpreting, and reporting an evaluation.

**Joel Schneider**—A favorite quote from William Stern, who coined the term "IQ": Every individual is a singularity, a one-time

existing being, nowhere else and never before present . . . . In this sense, the individual is a limiting concept, toward which theoretical investigation strives but can never reach; it is, one could say, the asymptote of science.

A child-centered focus drives all aspects of intelligent testing. Individualization demands a flexible examiner, one who is not wedded to just one assessment approach that is applied indiscriminately to all children. If an interpretive approach relies strictly on one view of the world, no matter how theoretically or psychometrically defensible that view may be, it is doomed to fail for some children. The CHC approach, for example, does not deal as effectively with executive functions as does a neuropsychological processing approach. And none of the contemporary models applied to the development and interpretation of IQ tests handles creativity particularly well (J. C. Kaufman, Kaufman, & Lichtenberger, 2011). Concerns with “same-old, same-old” report writing were specifically noted as problematic:

*John Willis*—In addition to writing a few thousand evaluation reports and receiving feedback from parents, teachers, and adult examinees, I have occasion to read many evaluation reports written by graduate students and by practicing psychologists and other evaluators and to provide my own feedback to those examiners. Even when the testing, scoring, and transcriptions of scores have been competent, many examiners appear to have selected a consistent battery of tests for most or all purposes (perhaps with some variation to accommodate the age of the examinee) and to have settled on an unvarying format for presenting the results with little regard for the audience or the circumstances. That autopilot approach is neither intelligent nor mindful.

*Elaine Fletcher-Janzen*—I am going rogue right now with report writing style and organization—I believe that because psychologists have written the psychometric test results as they have done for years—the public, special education personnel, parents,

and legal professionals have come to think of us as “testers” and that the numeric test scores trump the clinical judgment of the clinician (who interprets all scores in relation to the whole evaluation—not in isolation)—and worse—that they are entitled to interpret the numerical scores and have just as much insight as the clinician. So . . . I believe that psychologists should write a narrative report (and leave tables of the numeric results in the back of the report).

*John Willis*—Too often, the history section of an evaluation report seems to be a careful, accurately typed, but thoughtless exercise in summarizing data without considering the implications of those data. The referral questions and intended outcomes need to be considered in the context of the examinee’s personal, familial, social, emotional, medical, educational, and vocational history.

Intelligent testing requires the adaptability to be eclectic. In this book, we present interpretation from several clinical and theoretical viewpoints and provide an array of approaches to report writing in the 17 illustrative case reports that are included in Part IV. We provide a similar array of intelligent (but widely varying and often contradictory) WISC–V test reviews in Part V of the book. Thorough knowledge of many techniques for interpreting the WISC–V is important, as is mastery of the core areas of psychology, since these are the inputs necessary for interpreting a child’s test profile. Of course, experienced clinicians and trainers can be flexible while staying within a single theoretical model, such as the neuropsychological processing approach (Fletcher-Janzen & Reynolds, 2008; McCloskey et al., 2012) or the CHC-based Cross-Battery Assessment approach (Flanagan et al., 2012; Flanagan, Alfonso, & Reynolds, 2013; Flanagan et al., 2013; Ortiz, Flanagan, & Alfonso, 2015; Schneider & Flanagan, 2015). New examiners, however, should explore a variety of models of profile assessment before determining a priori which methodology is right for them.

Regarding the notion that intelligent testing in the second decade of the 21st century means

the ability to respect different methodologies and divergent voices:

**George McCloskey**—*Intelligent Testing with the WISC-R* was the first book on test interpretation that I read during my internship year (December 1979). Your thoughts about test interpretation and your writing style have influenced everything that I have written since . . . . Regarding the presentation of diverse approaches in *Intelligent Testing with the WISC-V*, I think this is a good direction to take for the rebirth of the *Intelligent Testing* book given the availability of the *Essentials* series books for handling much of what you might have originally included in the spirit of the *WISC-R* and *WISC-III* editions. I like the fact that multiple perspectives will be provided on test interpretation; this harkens back to your original groundbreaking ideas in the *WISC-R* book about looking at the WISC from multiple interpretive perspectives and with new statistical techniques but with a big twist—a major focus on actual case studies showing how it is done. I know you had case studies in the earlier *Intelligent Testing* books, but they were not the major emphasis the way they are here in the *WISC-V* book. I believe that beyond the *WISC-R* book, I have refined most of my ideas about test interpretation and report writing from reading other clinician's reports (both good and bad) . . . . Thanks so much for including me on this book project; I am confident the content will provide food for thought to many practitioners and scholars.

Ideally, with experience, well-trained examiners will be able to shift from one approach to another to find the best explanations for the observed fluctuations in a particular child's profile and to infer cause-effect relationships between behaviors and test performance. Or, if they are most comfortable with a particular theoretical perspective, then they can be flexible within their preferred discipline. However, experience and good intentions are sometimes not enough in the real world.

**Joel Schneider**—I struggle with teaching assessment. I only have a semester to teach students about something that takes years to master. I usually give students the advice to fly low to the ground until they have been given excellent supervision and have had the experience of their initial hunches being proven wrong a few times so that they acquire the appropriate level of humility and prudence.

The approach to interpretation advocated in this book may be impractical for some practitioners in the field. School psychologists who practice in states that embrace IQ tests for SLD diagnosis (even if reluctantly) are often faced with long waiting lists and inordinate numbers of cases that must be tested each week; they cannot usually administer very many tests as supplements to the WISC-V. Psychologists or learning disabilities specialists in a clinic have to assess an urban bilingual Korean child referred for evaluation whether or not they speak Korean or have any knowledge whatsoever of the child's subculture. These realities force evaluations of school-age children to be less than ideal in some instances, but they should not be used as excuses for fostering inadequate test interpretation or other types of test abuse.

For new psychologists, including graduate students who are in the process of becoming intelligent testers, here are some tips for use before, during, and after the evaluation.

### **Before the Evaluation**

#### **John Willis**

- I require my students to actively solicit and quote referral questions from parents, teachers, specialists, and the examinee and then to copy and paste those questions into the Conclusions section of their report and attempt to answer each one (even if the answer is "I don't know").
- Referral "concerns" are helpful, but genuine, potentially answerable questions are even

better. The questions may, if we really think about them, inspire additional questions that we can add to the collection. The questions can then help drive our choice of evaluation instruments and procedures. For example, if the referral question is the reason the student has difficulty comprehending full chapters in science or social studies texts, then a brief reading comprehension test consisting of very short passages may rule out sentence-reading weakness as a cause of the chapter-reading problem but will not otherwise answer the question.

### **Nadeen Kaufman**

- Select tests to give that have the best chance of providing you with the information you need to both answer referral questions and devise a plan for treatment, amelioration, and other interventions.
- Make sure you are trained and experienced enough (even if it means role play to learn) on every test you give so you can be the most efficient and fluid tester.
- Always select the appropriate, current tests with the correct record forms completely filled out and scored accurately: Always imagine that these forms can be required to be court documents that you will need to defend as an examiner. Know the child's birth date before the day of testing, as well as detailed referral questions.
- Get as much of an idea as possible of *exactly* what is wanted by referral source—don't just do a test evaluation based on "updated status needed." On this point, Joel Schneider (2013a) remarked: "You would think that being mindful of the purpose of an assessment is so obvious that it need not be stated explicitly. Unfortunately, it has too often been my experience that I am asked to give a child an assessment and no one involved with the child has a well-articulated understanding of why an assessment is needed" (p. 322).

### **During the Evaluation**

#### **Nadeen Kaufman**

- The individual you see the day of testing may not reflect who you prepared the evaluation for; be adaptable and flexible enough to change your plans—even midway—so you can deal with the person in front of you emotionally, physically, and cognitively.
- Understand the importance of rapport, and work to maintain it throughout the entire evaluation. Attend to the child and respect him or her as a participant in this case study. An unhappy or partially cooperative examinee is making test interpretation open to more error.
- Only take breaks when needed; don't interrupt flow because it's longer than you thought appropriate. Take breaks, keep them short, and don't talk about the test during break. Younger children may need some physical action during the break, or bathrooming. Regain rapport once the evaluation has resumed. If you have observed any new information about the child's abilities during the break (e.g., she knew her way back through the winding hallways to lead the examiner to the distant waiting-room area), include it as observation or corroborating hypothesis information.
- Perceive the evaluation as a case study experiment,  $n = 1$ , where you are incorporating all the psychology, neurology, cognitive processing information, developmental theories, etc., as additional tools you take with you into the test session.

#### **Jennifer Mascolo** Target your observations:

- Observe how the child responds to *qualitative features* of the test (e.g., on a math test, how does the child respond to items that demand mental computation versus those that allow for use of paper and pencil?).
- *Contemplate the interaction between the child's performance and item difficulty* (e.g., at what point

does the child begin to experience difficulty—such as when the guided line structure is removed from a Block Design task).

- Consider the *reasoning* behind the child's responses (e.g., is a child selecting responses for a matrix reasoning task based on location of the option—such as selecting a multiple-choice response that is directly below the “blank” space of the matrix?).
- Evaluate the *child's verbalizations* during testing (e.g., are they productive comments reflecting attempts to problem solve, are they bids for help from the examiner, are they self-evaluative remarks?).
- Note the *child's ability to shift within and between tasks* (e.g., Are children frequently self-correcting/changing responses? Do they want to refer back to check answers? Do they move on when prompted?).
- As clinicians, we need to actively consider these and other behaviors and maintain a close watch over them and record them (either manually or mentally) so that we can begin to see what themes emerge and can begin to build a picture of how this child likely responds and/or reacts to his or her learning environment.

### **After the Evaluation**

#### **Nadeen Kaufman**

- Don't let numbers make conclusions for you. Keep the gestalt in mind, and see the test scores as good evidence from one source of information. Factor in the child's behaviors, environment, and so on, to help readers understand the test scores more accurately, including the error inherent in measurement. Look for consistencies and inconsistencies across tests and within tests. Go outside the tasks and include behavioral responses to verbal or nonverbal activities and make comparisons.
- Be prompt in getting information back to referral sources and family. Things are requested for a need and things change; you don't need

to be a part of their problems when you can help solve their problems.

#### **Joel Schneider**

- When writing case reports, make every sentence worth reading.

## **2. The WISC–V Measures What the Individual Has Learned *and* Predicts Future Learning**

Subtests measure what the individual has learned. This is a point stated simply, but elaborated cogently, by Wesman (1968) in his seminal article on intelligent testing. The content of all tasks, whether they measure a CHC ability or neuropsychological process, is learned within a culture. The learning may take place formally in the school, casually in the home, or incidentally through everyday life. As a measure of past learning, the IQ test is best thought of as a kind of achievement test, not as a simple measure of aptitude. That contention is supported by the interesting finding that the *g* measured by intelligence tests is essentially the same *g* that underlies tests of academic achievement (reading, math, and writing), across all childhood and adolescent age groups studied (S. B. Kaufman, Reynolds, Liu, Kaufman, & McGrew, 2012). Thus, the strong relationships between cognitive tests and achievement measures (Naglieri & Bornstein, 2003) are inevitable. Plain and simple, cognitive tests measure prior learning.

But not so fast. Cognitive tests also have ample, long-standing evidence of clinical utility for predicting future learning in terms of academic achievement. As Schneider (2013a) articulated, the issue of aptitude-achievement is more complex, circular, and, ultimately, depends on the way the tests are used:

Most the time, aptitude is assessed by measuring abilities that are considered to be necessary precursors of achievement. For example, children who understand speech have greater aptitude for reading comprehension

than do children who do not understand speech. Such precursors may themselves be a form of achievement . . . . If we use a test to measure current mastery of a culturally valued ability, it is an achievement test. If we use a test to explain or forecast mastery of a culturally valued ability, it is an aptitude test. IQ tests are primarily used as aptitude tests. However, an inspection of the contents of most IQ tests reveals that many test items could be repurposed as items in an achievement test (e.g., vocabulary, general knowledge, and mental arithmetic items). Sometimes the normal roles of reading tests and IQ tests are reversed, such as when neuropsychologists estimate loss of function following a brain injury by comparing current IQ to performance on a word-reading test. (pp. 286–287)

Theories and research related to cognitive processing also describe a more circular relationship between crystallized and fluid intelligence. According to this line of research, learning occurs when the novel content of fluid reasoning becomes more crystallized, with transfer to long-term memory. The now-crystallized intelligence then serves to strengthen future fluid reasoning ability through increased associations between information in long-term storage for use in subsequent situations requiring fluid reasoning with novel content. Related investigations of the neuroanatomical and neurophysiological correlates of cognitive abilities seem to be zeroing in on this transitional process in the brain (Barbey, Colom, Paul, & Grafman, 2014; Hunt & Jaeggi, 2013).

Although the predictive validity of cognitive test scores is strong, the relationship between cognitive test scores and school achievement need not indicate a predetermined fate for an individual. That is, if results from a cognitive test are appropriately interpreted and translated into helpful recommendations, then positive change in academic achievement may occur, thereby changing one's IQ-determined "destiny."

The interaction between learning potential and availability of learning experiences is

too complex to ponder for any given person or subculture, making the whole genetics–environment issue of theoretical value but limited in terms of interpreting an individual's test profile. Issues of heredity versus environment and the validity of the IQ construct have been meaningful for understanding the multifaceted intelligence construct (A. S. Kaufman, 2009, chap. 6), and more recent efforts to evaluate the interaction of heritability and environmental influences on intelligence may shed additional light on this complex subject (Z. Wang, Katz, & Shaw, 2014). Regardless, the accumulating research helps test developers, practitioners, and theoreticians appreciate the relationships between genetic and environmental influences on the development of intellectual abilities as well as their interacting effects. The IQ tests are vehicles for the research, serving as essential sources of group data for use in the scientific study of these topics.

### 3. The WISC–V Subtests Are Samples of Behavior and Are Not Exhaustive

This tenet is relevant to the generalizability of the test findings. Because test results usually are obtained within 1 or 2 hours and include samples of behavior from a select set of tasks, caution needs to be exercised in generalizing the results to other behaviors in different circumstances. Even though there are 21 separate WISC–V subtests, and an array of subtests that compose the WJ IV, CAS2, DAS–II, and the Kaufman tests, the sum of these parts does not reflect the essential ingredients of intelligence whose mastery implies some type of ultimate life achievement. They, like tasks developed originally by Binet, are more or less arbitrary samples of behavior, even if they are driven by CHC or neuropsychological processing theory or by other research and theories related to cognitive structure, development, and function. Teaching people how to solve similarities, assemble blocks to match abstract designs, or repeat randomly presented digits in sequential order will not make them

smarter in any broad or generalizable way. What we are able to infer from the person's success on the tasks and style of responding to them is important; the specific, unique aspect of intellect that each subtest measures is of minimal consequence.

*John Willis*—Thoughtful consideration of referral questions, intended outcomes, and the examinee's history should be used to guide the selection of instruments and procedures. The referral may include specific desired outcomes, such as a *DSM-5* diagnosis, an IDEA disability identification, a recommendation for an educational or therapeutic placement, determination of legal competence, or prescription for a specific remedial reading program. The intended outcomes can then help drive the choice of evaluation instruments and procedures. A long history of difficulties in reading and spelling should inspire us to include adequate tests of word recognition, phonetic word attack, phonology, rapid naming, reading vocabulary, reading comprehension, oral vocabulary, and listening comprehension.

We can select formal or informal interview techniques, questionnaires, and rating scales; observations; and formal and informal tests that take into account what is already known about the examinee and that can be expected to provide the requested information. If we find ourselves administering the same procedures over and over again, we are probably not engaging in intelligent testing.

True to form, John concluded with a witty remembrance:

History may even guide the choice of examiners. I am a short, fat, homely, older man. When I used to work with a tall, slim, beautiful, young woman, we had the opportunity to accommodate examinees' fears and prejudices.

Limitations in the selection of tasks necessarily mean that one should be cautious in generalizing the results to circumstances that are removed from the one-on-one assessment of a

finite number of skills and processing strategies. Theoretical models of intelligence should be used as a guide for selecting cognitive tasks to administer during an evaluation, and examiners should not be wedded to a single instrument, even one as popular as the WISC-V. The need to use multiple measures was humorously emphasized by Kevin McGrew in his response to Alan's request for expert views on intelligent testing: "Moses may have gone to the mountain to get the Ten Commandments, but where is it written that David Wechsler went to the mountain to get the 12 subtests?"

To the degree that supplementary tasks and tests are chosen to meet theoretical and research-based constructs, the array of cognitive subtests chosen for a comprehensive test battery is systematic and not arbitrary. Dawn Flanagan's cross-battery assessment provides a user-friendly methodology for ensuring that any comprehensive evaluation of a person's intelligence measures a diversity of key broad and narrow CHC abilities (Flanagan et al., 2013). Jack Naglieri's PASS model (Naglieri et al., 2014), Sally Shaywitz's approach to diagnosing dyslexia (Shaywitz & Shaywitz, 2013), and George McCloskey's interpretive system via process analysis (McCloskey, 2009b; McCloskey, Whitaker, Murphy, & Rogers, 2012) are also useful frameworks for choosing the best supplementary measures to administer.

Nonetheless, even when using cross-battery assessment or a thorough neuropsychological processing analysis, neither a global IQ nor a comprehensive profile of CHC abilities should be interpreted as an estimate of a person's "total" or "complete" level of intellectual functioning. Examination of one's individual cognitive strengths and weaknesses obtained from IQ test data is more fruitful when combined with supportive data from other samples of behavior such as those data obtained from supplemental measures. These measures might include behavioral assessment, personality assessment, neuropsychological assessment, adaptive behavior assessment, and even informal assessment of

abilities that are not easily tested by standardized instruments, such as measures of the creative and practical intelligence components of Sternberg's triarchic theory of successful human intelligence (J. C. Kaufman, in press; Sternberg, Kaufman, & Grigorenko, 2008).

#### 4. The WISC–V Assesses Cognitive Functioning Under Fixed Experimental Conditions

*Jennifer Mascolo*—In my view, intelligent testing, as it applies to today and tomorrow, involves the clinician maintaining very *active, purposeful involvement* in the assessment process. While it will remain absolutely essential to familiarize oneself with key administration and scoring features of new and revised tests in any medium (e.g., paper-and-pencil or digital), these basic requirements of assessment must be mastered quickly so that the clinician can focus his or her attention on the child's interaction with the task at hand.

The standardized procedures for administration and scoring of the WISC–V help ensure objectivity in evaluating a child, but they sacrifice the in-depth understanding of a youngster's cognitive processing that may be obtained from a technique such as Jean Piaget's probing *methode clinique*, Reuven Feuerstein's test-teach-test dynamic assessment approach (Tzuriel, 2000), Lev Vygotsky's zone of proximal development (Rutland & Campbell, 1996), or Edith Kaplan's (1990) process approach. The rigidity of test instructions, the use of materials such as a stopwatch, and the recording of most words spoken by a child add to the artificial nature of the situation and make the standardized intelligence test scores comparable to data obtained in a psychological experiment. Do *not* deviate from required administration procedures or add nonpermissible probes to elicit a correct response from the child (except when testing the limits). Strict adherence to standardized procedures is essential, for otherwise the obtained scores—derived from

normative data collected painstakingly—are utterly meaningless. But interpretation is another matter, one that demands an awareness of the limitations of the standardized procedures so as to make the most sense out of the numerical scores. That is the time for a clinician's acumen and flexibility to be displayed.

Consider the finding by Hardy, Welcher, Mellits, and Kagen (1976) that urban children really “know” the answers to some WISC questions they get wrong, based on a testing-the-limits procedure. Their conclusion that “a standardized test, in this instance the WISC, may not be a valid estimate of the intellectual capabilities of inner-city children” (p. 50) follows logically only if the intelligence test is viewed as a criterion-referenced measure rather than as a sampling of abilities assessed under carefully specified conditions. Testing the limits on a subtest often can give valuable insight into the reasons for failure or confusion, so long as this flexible, supplemental testing occurs after the score has been recorded under appropriate conditions. Realization of the experimental nature of the testing process will prevent an examiner or researcher from interpreting a child's IQs as evidence of maximum performance or capacity.

*Jennifer Mascolo*—I firmly believe that with the drive to link assessment results to intervention, we must begin to meaningfully evaluate the child's response to testing of limits in the context of our evaluations. A very basic example might be a child who is administered an audio-recorded sound blending task and has difficulty. An informal testing of limits might involve forgoing the audio recording and, instead, having the child focus on the examiner as he or she is “saying” the sounds to be blended. If the testing of limits reveals that the provision of visual support (e.g., being able to “see” the speaker) improves performance, this recommendation can be carried over to the child's real instructional setting.

In an experiment, the empirical results are of limited value until they are interpreted and discussed in the context of pertinent theory and

research by a knowledgeable clinician. By the same token, the empirical outcomes of an IQ test are often meaningless until put into context by the examiner. Interpreting test scores in the context of observed behaviors can aid in the appropriate interpretation of the scores. For example, when an adolescent's oppositionality during testing has led to a low level of motivation on timed tasks, this behavior is crucial to understanding that the obtained scores may be a gross underestimate of the adolescent's abilities. When an examiner is able to relate observations of the child's behaviors in the testing situation to the profile of obtained scores (e.g., by noting that the child's anxiety disrupted test performance on all timed tasks), two things occur: (a) The examiner learns important information about the child that can be translated to practical educational suggestions, thereby enhancing the value of the intelligence test; and (b) The actual IQs earned by the child may represent gross underestimates of his or her real intellectual abilities.

In general, standardized test scores are valuable because they provide evidence of a child's mental functioning under a known set of conditions and permit comparison with youngsters of a comparable age. The value of the scores increases when the examiner functions as a true experimenter and tries to determine *why* the child earned the particular profile revealed on the record form; the scores become harmful when they are unquestioningly interpreted as valid indicators of intellectual functioning and are misconstrued as evidence of the child's maximum or even typical performance. For example, a person with excellent visual-spatial and manipulative skills might perform slowly and ineffectively on Block Design because of anxiety caused by the time pressure; or a person with a good commonsense understanding of social situations (coupled with limited word knowledge) may fail several Comprehension items because of a failure to understand some of the key words used in the questions. It is tempting to give credit to a design solved "just 2 or 3 seconds overtime" or to simplify the

wording of a question that the person "certainly knows the answer to." But the good examiner will resist these temptations, knowing that the people in the reference group did not receive such help.

Truly, standardized administration and scoring makes it imperative for examiners to function as scientist-practitioners and conduct an experiment with  $n = 1$  every time they administer a clinical test of intelligence. Only a competent clinician, armed with cutting-edge theory and research findings, can make sense of a person's profile of test scores. Intelligent testers must be exceptional clinicians to establish and maintain rapport and to weave the standardized administration into a natural, pleasant interchange between examiner and subject. Clinical skills are also essential when observing and interpreting a person's myriad behaviors during the examination and when integrating all available information and data to give meaning to the myriad numbers that are yielded by every administration of the WISC-V.

### 5. The WISC-V Is Optimally Useful When It Is Interpreted from a Theoretical Basis

The theoretical models that began to influence the development of IQ tests in the 1980s and continue to drive both test construction and profile interpretation more than a generation later have reshaped the field of cognitive assessment. CHC theory, neuropsychological processing models, neurodevelopmental, and a variety of other approaches have been researched extensively and applied intelligently to the clinical assessment of mental abilities (Flanagan & Harrison, 2012; Lichtenberger & Kaufman, 2013; Schneider, 2013a).

The need to interpret the WISC-V from theory is axiomatic, is illustrated throughout this book, and requires no elaboration here. Of great interest, though, is the interface between the growth of theory and the development of

theory-based IQ tests. Cause and effect is hard to determine.

Still awed by a momentous 1986 meeting in Dallas, *Kevin McGrew* said

... that this moment—a moment where the interests and wisdom of a leading applied test developer (Woodcock), the leading proponent of Cattell-Horn *Gf-Gc* theory (Horn), and one of the preeminent educational psychologists and scholars of the factor analysis of human abilities (Carroll) intersected—was the *flash point that resulted in all subsequent theory-to-practice bridging events that led to today's CHC theory and related assessment developments*. A fortuitous set of events had resulted in the psychometric stars aligning themselves in perfect position to lead the way for most all subsequent CHC assessment related developments.

Further, Alan has written about a second crucial meeting that took place in 1999 in Chapel Hill:

That meeting was attended by authors of the WJ III (Dick Woodcock, Kevin McGrew) and Binet-5 [SB5] (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. (A. S. Kaufman, 2009, p. 99)

No matter how much a theory such as the CHC theory has been validated and used to inform test development, profile interpretation, and educational interventions, Schneider and

Flanagan (2015) remind us that the individual, not the theory, is the bottom line:

Although classifying tests can be fun, it is a stale enterprise when it becomes an end in itself. It is easy to lose sight of the fact that whether a test measures lexical knowledge, perceptual speed, memory span, and the like cannot be our ultimate concern. What matters is what each test can tell us about individuals. If we do not know what a low score on the WJ III Visual Closure subtest means for the future well-being of an individual, the test's place in a taxonomy is of minor importance. (p. 335)

As previously noted in this chapter and elaborated in the following chapters, the study of intelligence today spans several fields, including clinical psychology, school psychology, cognitive psychology, neuropsychology, cognitive neuroscience, brain imaging, and gene mapping. Increasingly, experts from these different disciplines are joining forces to conduct collaborative research with the common goal of improving outcomes for individuals experiencing difficulties with mental functioning. In light of this trend, it is very likely that the next “meeting of the minds” will require a larger round table and a longer invitation list to forge the next important steps of uniting theory, research, and practice toward that goal. We sincerely hope that is the case.

## 6. Hypotheses Generated from WISC-V Profiles Should Be Supported with Data from Multiple Sources

Test score profiles are optimally meaningful when interpreted in the context of known background information, observed behaviors, approach to each problem-solving task, and scores on related tasks. Virtually any examiner can figure out that the WISC-V Verbal Comprehension Index is not a very good measure of *Gc* for a child raised in a foreign culture, a child who understands Spanish or Tagalog far better than English, or a child with a hearing

impairment; that the WISC–V Processing Speed Index or Block Design (*Gv*) scaled score does not measure its designated CHC ability very well for a child with cerebral palsy or a visual handicap. Intelligent testers must try to deduce when one or more subtests or indexes may be an invalid measure of a child’s intellectual functioning for more subtle reasons: distractibility, subcultural differences in language or custom, emotional content of the items, suspected or known lesions in specific regions of the brain, fatigue, boredom, extreme shyness, bizarre thought processes, inconsistent effort, and the like.

**John Willis**—If we have been asked to measure a person’s intelligence, then we should not use an intelligence test that, for that person, measures abilities and disabilities that are not related to that individual’s intelligence. Many intelligence tests include, for example, subtests that require motor speed and precision, adequate visual acuity, good auditory acuity and perception, and lifelong exposure to United States culture. If none of those issues represents a disability or disadvantage for the examinee, then those subtests may serve as valid measures of intelligence. However, if one or more of those issues is a disability or disadvantage for the examinee, then the test measures that issue, not intelligence. Similarly, a test of oral reading fluency does not measure the automaticity of reading decoding, recognition, and comprehension for a student with a severe stutter. We should not copy down the information that the examinee has convergence insufficiency or severe fine-motor coordination weaknesses and then measure intelligence with a score that includes such subtests as the Wechsler Coding or Woodcock-Johnson Cross Out.

As samples of behavior obtained under conditions that resemble psychological experiments, test scores can mislead just as easily as they can lead. An investigation of peaks and valleys in WISC–V profiles might yield a relatively high Quantitative Reasoning Index and a low Processing Speed Index. That profile

suggests hypotheses of good quantitative reasoning ability, a measure of a CHC narrow ability within the *Gf* domain, and poor processing speed (*Gs*). Even though this combination of strong and weak areas is based on reliable, theory-based indexes, clinicians cannot automatically assume that they have identified the child’s precise strength and weakness. The labels given each index must be thought of as *hypotheses*, nothing more. And these hypotheses must be translated to educational interventions for any child referred for a learning problem (or they must help pinpoint processing deficits) or they are of limited value. Experts elaborated further on the important role of intelligent testing in evaluating a child’s pattern of cognitive strengths and weaknesses:

**Nancy Mather**—Essentially, intelligent testing involves an appreciation that the test instruments are only tools that help a clinician validate the specific factors that enhance and impede performance. By obtaining a deeper, more nuanced understanding of both a person’s strengths and weaknesses, an evaluator can determine appropriate individualized, targeted interventions and therapies. A skilled clinician uses the results from intelligent testing to weave together a tapestry that depicts all important facets of performance, addresses the individual’s current challenges and needs, and presents solutions designed to enhance a person’s life.

Another important aspect of intelligent testing is the uncovering and documentation of an individual’s areas of strength on the various abilities and factors that facilitate performance. Howard Gardner (1999) clearly made this point when he advised, “We shouldn’t ask how smart you are, but rather how are you smart?” The results from intelligent testing can be used to help an individual deepen self-understanding, consider or reconsider scholastic and vocational opportunities, and increase appreciation of his or her own unique talents and abilities.

**Ron Dumont**—The use of IQ testing has changed over the years and continues to evolve. It needs to be viewed, in my opinion, as a

snapshot of differing abilities (maybe the IDEA psychological processes) or the CHC broad abilities. These abilities all have complete test batteries available to assess the separate abilities, and no one seems to be damning or downplaying those tests' importance. The trouble seems to come because modern IQ tests attempt to incorporate the separate abilities all into the "global aggregate." We have to resist the thought that that single score is some sort of be-all and end-all number.

I believe IQ testing has an important place in the diagnosis of certain issues, learning problems included. It isn't the overall global score that makes much sense, but the underlying skills assessed.

*Elaine Fletcher-Janzen*—I must admit that there is less need for an emphasis on the FSIQ and much more need for determining strengths and weaknesses that must be deconstructed and validated with further testing. The "further testing" does not rely on what test publishers purport the tests to measure but more on the clinician's understanding of the processes needed to succeed on the test at various levels of difficulty throughout the test—the process approach if you will. The "further testing" does not rely on arbitrary rules, such as needing two subtests in any given processing area. The analyses go until the clinician understands the construct boundaries—and also the context of how the process develops in different settings (e.g., testing room with one-on-one examiner versus classroom with teacher and 25 students).

How should one best interpret a particular profile of WISC-V indexes? Are the supposed strengths or weaknesses repeatable, or a one-time phenomenon? Both of these questions can be answered, with varying degrees of confidence, by using multiple sources of information and diverse pieces of data. Some of these bits of evidence come from diligent observation during the WISC-V administration and careful scrutiny of response patterns; others come from the background information provided by referral sources, including previous test data, and from the administration of additional subtests and tests.

*Nancy Mather*—In the case of an individual suspected of having a reading disability, the evaluator would want to consider:

(a) background information, including a family history of reading difficulties; (b) educational history, including prior services, tutoring, and accommodations that the student received; (c) performance on different facets of reading performance, including word reading accuracy, reading rate, and comprehension; (d) performance on relevant oral language measures (e.g., listening comprehension) versus their performance on reading measures; (e) performance on higher-level cognitive tasks (e.g., language and reasoning measures) and how these abilities compare to performance on lower-level more perceptual tasks (e.g., processing speed, working memory, and phonological awareness); and (f) qualitative information, such as the reader's ability to pronounce multisyllabic words, as well as the types of errors he or she makes, such as having difficulty pronouncing medial vowel sounds. After this information is gathered and analyzed, the evaluator then attempts to determine the exact nature of the reading problem in order to recommend appropriate accommodations and interventions.

Consider the child with WISC-inspired hypotheses of weak processing speed and strong quantitative reasoning (an aspect of the broader domain of fluid reasoning). The processing speed hypothesis is based on a WISC-V grouping of two subtests, Coding and Symbol Search. Those low scores might be due to poor processing speed. However, they also might be due to poor visual-motor coordination (an aspect of output), visual-spatial ability, visual perception (an aspect of input); or to low motivation, reflectiveness, distractibility, or obsessiveness; and so forth. What was the child like during the entire evaluation? Did he or she handle the pencil and blocks awkwardly and appear uncoordinated? Was the child unusually reflective or unmotivated? Was it possibly an attentional problem? What picture was painted of the child by parents or teachers? Did the child score low on other tests

of processing speed given sometime in the past, or administered as part of the present evaluation (the secondary Cancellation subtest or even the complementary measures of Naming Speed)? What about the error scores? Did the child make rotation errors on Coding and Symbol Search (if so, the low score might be due to a visual-perceptual input problem)?

Similarly, good quantitative reasoning ability would automatically be hypothesized from high scores on Figure Weights and Arithmetic, the two WISC–V subtests that compose this ancillary index. But to truly test this hypothesis, examiners should check out the child’s performance on Coding B (for children ages 8 and above), because that task also involves numbers, and on Letter–Number Sequencing. They might also examine the child’s scores on Items 18 to 40 on the Information subtest: Five of these items require numerical answers or knowledge of number concepts—did the child or adolescent pass these number items while failing items of comparable difficulty (i.e., items just before or after the number items)? Children with good quantitative reasoning ability also might have unusually good longest spans on Digit Span Backward and Sequencing, perhaps relative to that of Forward, because the ability to repeat numbers *backward* or in *sequence*—unlike the more mindless task of mimicking a forward span—is aided by a child’s facility in handling numbers. The WISC–V process scores are helpful in this regard, providing scaled scores on each separate component of Digit Span and offering invaluable contrast scaled scores to permit direct comparisons (Wechsler, 2014, Appendix C).

Other checks on hypotheses of good quantitative reasoning might come from scores on Mathematics achievement subtests in group or individual batteries that were included on the child’s record or administered as part of the evaluation; report card grades in arithmetic; statements by a referral source, or by the child, that he or she is good at math; performance on cognitive or achievement subtests from other batteries that demand quantitative reasoning, such

as Sequential and Quantitative Reasoning on the DAS–II (C. D. Elliott, 2007), Number Series on the WJ IV Cognitive, Applied Problems and Number Matrices on the WJ IV Achievement, and KTEA–3 Math Concepts & Applications (A. S. Kaufman & Kaufman, 2014).

## 7. Results from WISC–V Profiles Must Be Tied to Educational Interventions

In the olden days of assessment, profile interpretation might have stopped with an incisive understanding of the child’s pattern of strengths and weaknesses, but that is no longer the case. If the RTI movement and the passage of IDEA have taught us anything, it is that examiners must translate the test profile to action, especially empirically based educational interventions (McBride et al., 2011). The mandate for making meaningful recommendations that will potentially change a child’s life warrants inclusion of this aspect as a new tenet of the intelligent testing philosophy. The increased emphasis on linking practice to intervention was addressed by almost every expert, attesting to the importance of its inclusion in the intelligent testing philosophy of today:

*Nancy Mather*—Intelligent testing needs to lead to intelligent decision making. Typically, the central purpose of an assessment is to derive solutions for some type of concern or problem. Thus, the results when combined with other relevant data and observations must be used to diagnose specific conditions, as well as address and hopefully answer the referral question. As Cruickshank (1977) advised: “Diagnosis must take *second* place to instruction, and must be made a *tool of instruction*, not an end in itself” (p. 193). Essentially, intelligent testing involves an appreciation that the test instruments are only tools that help a clinician validate the specific factors that enhance and impede performance. By obtaining a deeper, more nuanced understanding of both a person’s strengths and weaknesses, an evaluator can determine appropriate individualized, targeted interventions and therapies.

*Jennifer Mascolo*—If a child is consistently shutting down or becoming tentative in his or her responding after corrective feedback in an assessment situation, and continues to demonstrate this reaction/response throughout tests involving feedback, we have at least initial support that corrective feedback in learning situations needs to be carefully designed so as to not “close off” the child’s attention to the task at hand. In building this picture, I believe that we can use such information to inform instructional efforts.

*Nancy Mather*—Intelligent testing is not synonymous with the interpretation of an intelligence test; it is the process of using test instruments as tools for helping to understand an individual’s unique abilities and then developing recommendations that will result in positive outcomes. The essential focus is upon problem solving for each individual who is evaluated. In many cases, the results from intelligent testing are used to confirm or negate the existence of a specific disability. Intelligent testing, however, goes way beyond the identification of disabilities, or compliance with eligibility requirements that only require the simple calculation of scores. As aptly noted by Willis and Dumont (2002), the determination of a disability involves more than “an exercise in arithmetic” (p. 173). In order to make an accurate diagnosis, an evaluator considers information from a variety of sources, including actual classroom performance, educational history, and behavioral observations. The results then provide important information that helps the evaluator explain the nature and severity of a problem and then determine the most appropriate treatment options.

*Jennifer Mascolo*—We need to begin to evaluate response to strategy instruction. For example, when we offer the child additional real-time support or model ways to re-approach a task, does the child’s performance improve? The degree of responsiveness can inform intervention efforts and potentially inform the level of support that might be required (e.g., a highly responsive child might need basic strategy instruction, whereas one with more significant difficulties

might require more intensive remediation and/or accommodations).

So, in short, I think that targeted observations that focus on the child, the interaction of child and task, and the child’s response to testing of limits/instructional scenarios are all important to consider and represent “intelligent testing” in the context of assessment.

*Nancy Mather*—The goals of intelligent testing are to identify the person’s strengths and weaknesses, attempt to understand the relationship between and among his or her abilities, and then translate these findings into meaningful intervention plans. Thus, the crux of intelligent testing is the translation of an individual’s test results into meaningful recommendations that will produce positive outcomes. Dr. Herman Hall, regarded as an expert in the interpretation of standardized tests, demonstrated how the interpretation of test results could directly help educators, parents, and students find success in their lives. In describing Hall’s insights, Shapiro (2004) explained: “the diagnosis could lead to the implementation of a program to help the child—whether formally through a tutor, quietly through a teacher’s enhanced awareness of what would work best for the child, indirectly through the ebbing of parents’ anxiety or simply by the student gaining a better understanding of himself as a learner, and to some degree, as a person. All of these great things could happen.” (pp. 15–16)

And where do we see intelligent testing heading in the future? Here are some closing predictions from experts of today:

*Alan Kaufman* (in 2009, regarding the future of IQ testing)—If IQ tests virtually disappear from the school scene in many districts throughout the nation and in some states altogether, how big a dent will that make in IQ test use? Big, in terms of *quantity*, little in terms of *quality*.

In fact, the quality of test use will increase. The identification of children with SLD has provided a prominent role for IQ tests for more than 30 years. But because of the discrepancy formula that loomed over every

child referred for possible SLD, IQ tests were often given for the wrong reasons (to plug a number into a formula) by the wrong people (those who found no use for IQ tests *except* to plug into a formula). That will stop . . . . But most states, I believe, that opt to use IQ tests will do so to identify the child's pattern of strengths and weaknesses. That approach will help identify a processing disorder as well as cognitive strengths for the purpose of individualizing educational planning. And that is intelligent testing. (pp. 296–297)

**Ron Dumont**—I see tests being held to higher standards. I see many new tests continually changing to incorporate the science. How long did it take the Wechsler scale to finally divide the Performance/Perceptual Reasoning factor into *Gv* and *Gf*? There seems to be expectations from users for more validity evidence. This will continue I believe.

**Diane Coalson**—To take things a step further in the future, I think we need well-designed research that integrates improved digital assessments, brain-imaging techniques, “brain-training” interventions, with real-life results to see what type of intervention actually helps (i.e., improved function indicated in a brain scan may not correlate well with improved performance in a real-life task).

I also think that we now know enough about some of the brain's processing system maps to hypothesize about additional problem areas a child may have, based on ongoing test

results—something a digital environment may do with more ease than hand scoring. For example, the brain areas indicated in working memory, fluid reasoning, and executive function overlap. Can a combined knowledge of function in these areas lead us to more targeted assessment (something like subtraction-based hypothesis testing) and subsequent interventions?

**Joel Schneider and Dawn Flanagan** (2015)—It is likely that cognitive ability test interpretation will be directly incorporated into academic progress monitoring. That is, traditional cognitive ability assessments and the Response to Intervention (RTI) approach will become integrated into a coherent and unified interpretive framework. Complex prediction equations involving cognitive and academic abilities will include important covariates such as past performance, time on task, task persistence, and quality of instruction. This framework is likely to be increasingly informed by dynamic brain imaging techniques and well-developed cognitive information processing models. That is, we will be able to observe information processing deficits and abnormalities in real time as evaluatees perform academic tasks. Furthermore, we will be able to monitor directly whether interventions succeed in normalizing the processing deficits we identify. (p. 337)

We eagerly await the future events that will support or refute these predictions.