One

INTRODUCTION AND OVERVIEW

There are more individually administered tests of intelligence and IQ available today than were available at any other time in the history of psychological assessment and applied measurement. Despite all the innovations and exemplary quantitative and qualitative characteristics of new and recently revised intelligence tests, the Wechsler scales continue to reign supreme. In fact, the Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV)—like its predecessor, the WISC-III—has quickly become the most widely used measure of intelligence the world over. Because the latest edition of the WISC represented the most substantial revision of any Wechsler scale to date, we developed, in the first edition of this book, an interpretive system for the WISC-IV that was quite different from Wechsler interpretive systems of the past (e.g., Flanagan, McGrew, & Ortiz, 2000; Kaufman & Lichtenberger, 1999). For example, the elimination of the Verbal and Performance IQs required us to reconceptualize previous systems completely. Also, the proliferation of anti-profile research and writing, primarily by Glutting, Watkins, and colleagues, and the anti-profile sentiment that currently characterizes the field, impelled us to deal with the interpretive system not just as an empirical, logical, and theoretical endeavor, but also as a controversial topic. Finally, the nature of the contemporary scene, which has undergone substantial changes in test usage based on the wording of the Individuals with Disabilities Education Act (IDEA) legislation in 2004 and its attendant regulations in 2006, forced us to think outside of the box with an eye toward the future. Thus, the first edition of this book provided a psychometrically and theoretically defensible system of interpreting the WISC-IV and we believe we achieved our goal of anticipating what *best practices* in the use of the Wechsler scales would be in the coming decade (Flanagan, Ortiz, Alfonso, & Dynda, 2008). Our main reasons for this second edition were to update the research that has been conducted with the WISC-IV since the first edition of this book was published in 2004; to provide more detailed information on how to link WISC-IV assessment results to research-based interventions; to extend our interpretive system to include an interesting new cluster, the Cognitive Processing Index or CPI (Weiss, Saklofske,

Schwartz, Prifitera, & Courville, 2006); to demonstrate how the WISC-IV Integrated can complement information gleaned from the WISC-IV; and to include a CD-ROM with a software program that automates our interpretive system. Note that the CD-ROM also contains all the Appendixes to this book. Each appendix may be downloaded for your convenience.

Similar to our previous writings on the Wechsler scales, our main objective was to provide a comprehensive and user-friendly reference for those who use the WISC-IV. This book was developed specifically for those who test children between the ages of 6 and 16 and wish to learn the *essentials* of WISC-IV assessment and interpretation in a direct and systematic manner. The main topics included in this book are administration, scoring, interpretation, and clinical application of the WISC-IV. In addition, this book highlights the most salient strengths and limitations of this instrument. Throughout the book, important information and key points are highlighted in Rapid Reference, Caution, and Don't Forget boxes. In addition, tables and figures are used to summarize critical information and to explain important concepts and procedures, respectively. Finally, each chapter contains a set of Test Yourself questions that are designed to help you consolidate what you have read. We believe you will find the information contained in this book quite useful for the competent practice of WISC-IV administration, scoring, and interpretation.

This chapter provides a brief overview of historical and contemporary views of the Wechsler scales as well as a brief historical account of Wechsler scale interpretation. In addition, the WISC-IV is described and its most salient features are highlighted. Finally, a brief summary of the controversy surrounding profile interpretation with the Wechsler scales is provided, followed by a comprehensive rationale for the interpretive method described in this book.

HISTORICAL AND CONTEMPORARY VIEWS OF THE WECHSLER SCALES

Within the field of psychological assessment, the clinical and psychometric features of the Wechsler intelligence scales have propelled these instruments to positions of dominance and popularity unrivaled in the history of intellectual assessment (Alfonso et al., 2000; Flanagan et al., 2000; Kaufman, 2003; Kaufman, Flanagan, Alfonso, & Mascolo, 2006). The concepts, methods, and procedures inherent in the design of the Wechsler scales have been so influential that they have guided much of the test development and research in the field for more than a half century (Flanagan et al.). Virtually every reviewer of these scales, including those who have voiced significant concerns about them, has acknowledged the monumental impact that they have had on scientific inquiry into the nature of

human intelligence and the structure of cognitive abilities. For example, despite the critical content and tone of their "Just Say No" to Wechsler subtest analysis article, McDermott, Fantuzzo, and Glutting (1990) assert their "deep respect for most of the Wechsler heritage" by stating that "were we to say everything we might about the Wechsler scales and their contributions to research and practice, by far our comments would be quite positive" (p. 291).

Likewise, Kamphaus (1993) observed that praise flows from the pages of most reviews that have been written about the Wechsler scales. Kaufman's (1994b) review, entitled "King WISC the Third Assumes the Throne," is a good example of the Wechsler scales' unrivaled position of authority and dominance in the field (Flanagan et al., 2001). Although the strengths of the Wechsler scales have always outweighed their weaknesses, critics have identified some salient limitations of these instruments, particularly as they apply to their adherence to contemporary theory and research (e.g., Braden, 1995; Flanagan et al., 2000, 2008; Little, 1992; Kaufman et al., 2006; McGrew, 1994; Shaw, Swerdlik, & Laurent, 1993; Sternberg, 1993; Witt & Gresham, 1985). Nevertheless, it remains clear that when viewed from an historical perspective, the importance, influence, and contribution of David Wechsler's scales to the science of intellectual assessment can be neither disputed nor diminished. The following paragraphs provide historical information about the nature of the Wechsler scales and summarize important developments that have occurred over several decades in attempts to derive meaning from the Wechsler IQs and scaled scores.

BRIEF HISTORY OF INTELLIGENCE TEST DEVELOPMENT

Interest in testing intelligence developed in the latter half of the 19th century. Sir Francis Galton developed the first comprehensive test of intelligence (Kaufman, 2000b) and is regarded as the father of the testing movement. Galton theorized that because people take in information through their senses, the most intelligent people must have the best developed senses; his interest was in studying gifted people. Galton's scientific background led him to develop tasks that he could measure with accuracy. These were sensory and motor tasks, and although they were highly reliable, they proved ultimately to have limited validity as measures of the complex construct of intelligence.

Alfred Binet and his colleagues developed tasks to measure the intelligence of children within the Paris public schools shortly after the end of the 19th century (Binet & Simon, 1905). In Binet's view, simple tasks like Galton's did not discriminate between adults and children and were not sufficiently complex to measure human intellect. In contrast to Galton's sensorimotor tasks, Binet's were primarily language oriented, emphasizing judgment, memory, comprehension, and reasoning. In the

1908 revision of his scale, Binet (Binet & Simon, 1908) included age levels ranging from 3 to 13 years; in its next revision in 1911, the Binet-Simon scale was extended to age 15 and included five ungraded adult tests (Kaufman, 1990a).

The Binet-Simon scale was adapted and translated for use in the United States by Lewis Terman (1916). Binet's test was also adapted by other Americans (e.g., Goddard, Kuhlmann, Wallin, and Yerkes). Many of the adaptations of Binet's test were of virtual word-for-word translations; however, Terman had both the foresight to adapt the French test to American culture and the insight and patience to obtain a careful standardization sample of American children and adolescents (Kaufman, 2000b). Terman's Stanford-Binet and its revisions (Terman & Merrill, 1937, 1960) led the field as the most popular IQ tests in the United States for nearly 40 years. The latest edition of the Stanford-Binet—the Stanford-Binet Intelligence Scales–Fifth Edition (SB5; Roid, 2003)—is a testament to its continued popularity and longevity in the field of intellectual assessment.

The assessment of children expanded rapidly to the assessment of adults when the United States entered World War I in 1917 (Anastasi & Urbina, 1997). The military needed a method by which to select officers and place recruits, so Arthur Otis (one of Terman's graduate students) helped to develop a group-administered IQ test that had verbal content quite similar to that of Stanford-Binet tasks. This was called the Army Alpha. A group-administered test consisting of nonverbal items (Army Beta) was developed to assess immigrants who spoke little English. Ultimately, army psychologists developed the individually administered Army Performance Scale Examination to assess those who simply could not be tested validly on the group-administered Alpha or Beta tests (or who were suspected of malingering). Many of the nonverbal tasks included in the Beta and the individual examination had names (e.g., Picture Completion, Picture Arrangement, Digit Symbol, Mazes) that may look familiar to psychologists today.

David Wechsler became an important contributor to the field of assessment in the mid-1930s. Wechsler's approach combined his strong clinical skills and statistical training (he studied under Charles Spearman and Karl Pearson in England) with his extensive experience in testing, which he gained as a World War I examiner. The direction that Wechsler took gave equal weight to the Stanford-Binet/Army Alpha system (Verbal Scale) and to the Performance Scale Examination/Army Beta system (Performance Scale). The focus that Wechsler had in creating his battery was one of obtaining dynamic clinical information from a set of tasks. This focus went well beyond the earlier use of tests simply as psychometric tools. The first in the Wechsler series of tests was the Wechsler-Bellevue Intelligence Scale (Wechsler, 1939). In 1946, Form II of the

Wechsler-Bellevue was developed, and the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949) was a subsequent downward extension of Form II that covered the age range of 5 to 15 years. Ultimately, the WISC became one of the most frequently used tests in the measurement of intellectual functioning (Stott & Ball, 1965). Although the practice of using tests designed for school-age children in assessing preschoolers was criticized because of the level of difficulty for very young children, the downward extension of such tests was not uncommon prior to the development of tests specifically for children under age 5 (Kelley & Surbeck, 1991).

The primary focus of the testing movement until the 1960s was the assessment of children in public school and adults entering the military (Parker, 1981). However, in the 1960s the U.S. federal government's increasing involvement in education spurred growth in the testing of preschool children. The development of government programs such as Head Start focused attention on the need for effective program evaluation and the adequacy of preschool assessment instruments (Kelley & Surbeck, 1991). In 1967, the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) was developed as a downward extension of certain WISC subtests but provided simpler items and an appropriate age-standardization sample. However, because the WPPSI accommodated the narrow 4:0- to 6:5-year age range, it failed to meet the needs of program evaluations because most new programs were for ages 3 to 5 years.

Public Law 94-142, the Education for All Handicapped Children Act of 1975, played an important role in the continued development of cognitive assessment instruments. This law and subsequent legislation (IDEA of 1991, IDEA Amendments in 1997, and IDEA of 2004) included provisions that required an individualized education program (IEP) for each disabled child (Sattler, 2001). A key feature of the development of the IEP is the evaluation and diagnosis of the child's level of functioning. Thus, these laws directly affected the continued development of standardized tests such as the WPPSI and WISC. The WISC has had three revisions (1974, 1991, 2003), and the WPPSI has had two (1989, 2002). The WISC-IV is the great-great-grandchild of the 1946 Wechsler-Bellevue Form II; it is also a cousin of the Wechsler Adult Intelligence Scale–Third Edition (WAIS-III), which traces its lineage to Form I of the Wechsler-Bellevue. Figure 1.1 shows the history of the Wechsler scales.

In addition to the Wechsler scales and SB5, the Woodcock-Johnson Test of Cognitive Ability (originally published in 1977) is in its third edition (WJ III; Woodcock, McGrew, & Mather, 2001); the Kaufman Assessment Battery for Children (K-ABC; published in 1983) is in its second edition (KABC-II; Kaufman & Kaufman, 2004a); and the Differential Abilities Scale (DAS; Elliott,

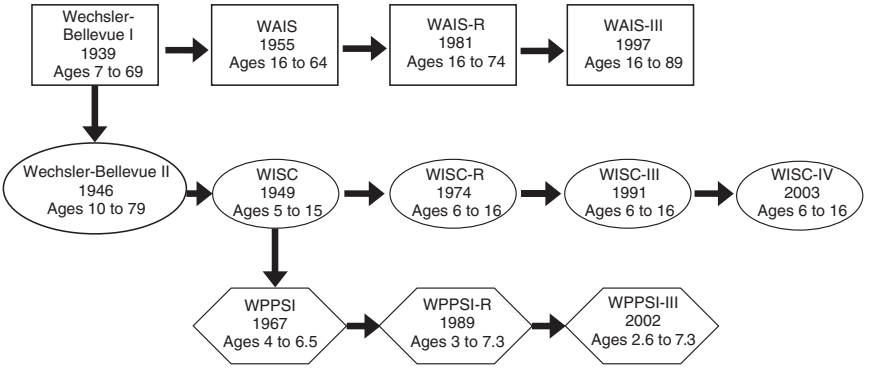

Figure 1.1 History of Wechsler Intelligence Scales

Note: WPPSI = Wechsler Preschool and Primary Scale of Intelligence; WISC = Wechsler Intelligence Scale for Children; WAIS = Wechsler Adult Intelligence Scale. From A. S. Kaufman & E. O. Lichtenberger, *Essentials of WISC-III and WPPSI-R Assessment.* Copyright © 2000. John Wiley & Sons, Inc. This material is used by permission of John Wiley & Sons, Inc.

1991) is in its second edition (DAS-II; Elliott, 2007). Other intelligence tests that have joined the contemporary scene include the Cognitive Assessment System (CAS; Naglieri & Das, 1997), the Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, 1997), and the Reynolds Intellectual Ability Scale (RIAS; Reynolds & Kamphaus, 2003). What is most striking about recently revised and new tests of intelligence is their generally close alliance with theory, particularly the Cattell-Horn-Carroll (CHC) theory. (See Appendix A on the CD-ROM for detailed definitions of the CHC abilities and Appendix B on the CD-ROM for a list of major intelligence tests and the CHC abilities they measure.) For a complete discussion of contemporary intelligence tests and their underlying theoretical models, see Flanagan and Harrison (2005).

BRIEF HISTORY OF INTELLIGENCE TEST INTERPRETATION

Randy Kamphaus and his colleagues provided a detailed historical account of the many approaches that have been used to interpret an individual's performance on the Wechsler scales (Kamphaus, Petoskey, & Morgan, 1997; Kamphaus, Winsor, Rowe, & Kim, 2005). These authors describe the history of intelligence test interpretation in terms of four *waves:* (a) quantification of general level; (b) clinical profile analysis; (c) psychometric profile analysis; and (d) application of theory to

DON'T FORGET

Origin of WISC-IV Subtests

Verbal Comprehension Index (VCI)	**Historical Source of Subtest**
Vocabulary	Stanford-Binet
Similarities	Stanford-Binet
Comprehension	Stanford-Binet/Army Alpha
(Information)	Army Alpha
(Word Reasoning)	Kaplan's Word Context Test (Werner & Kaplan, 1950)
Perceptual Reasoning Index (PRI)	**Historical Source of Subtest**
Block Design	Kohs (1923)
Matrix Reasoning	Raven's Progressive Matrices (1938)
Picture Concepts	Novel task developed by The Psychological Corporation
(Picture Completion)	Army Beta/Army Performance Scale Examination
Working Memory Index (WMI)	**Historical Source of Subtest**
Digit Span	Stanford-Binet
Letter-Number Sequencing	Gold, Carpenter, Randolph, Goldberg, and Weinberger (1997)
(Arithmetic)	Stanford-Binet/Army Alpha
Processing Speed Index (PRI)	**Historical Source of Subtest**
Coding	Army Beta/Army Performance Scale Examination
Symbol Search	Schneider and Shiffrin (1977) and S. Sternberg (1966)
(Cancellation)	Diller et al. (1974), Moran and Mefford (1959), and Talland and Schwab (1964)

Source: From A. S. Kaufman & E. O. Lichtenberger, *Essentials of WISC-III and WPPSI-R Assessment.*

Note: Supplementary subtests appear in parentheses.

intelligence test interpretation. Kamphaus and colleagues' organizational framework is used here to demonstrate the evolution of Wechsler test interpretation.

The First Wave: Quantification of General Level

Intelligence tests, particularly the Stanford-Binet, were used widely because they offered an objective method of differentiating groups of people on the basis of their general intelligence. According to Kamphaus and colleagues (1997; Kamphaus et al., 2005), this represented the first wave of intelligence test interpretation and was driven by practical considerations regarding the need to classify individuals into separate groups.

During the first wave, the omnibus IQ was the focus of intelligence test interpretation. The prevalent influence of Spearman's *g* theory of intelligence and the age-based Stanford-Binet scale, coupled with the fact that factor-analytic and other psychometric methods were not yet available for investigating multiple cognitive abilities, contributed to the almost exclusive use of global IQ for classification purposes. Hence, a number of classification systems were proposed for organizing individuals according to their global IQs.

Early classification systems included labels that corresponded to medical and legal terms, such as *idiot*, *imbecile*, and *moron*. Although the Wechsler scales did not contribute to the early classification efforts during most of the first wave of test interpretation, Wechsler eventually made his contribution. Specifically, he proposed a classification system that relied less on evaluative labels (although it still contained the terms *defective* and *borderline*) and more on meaningful deviations from the mean, reflecting the "prevalence of certain intelligence levels in the country at that time" (Kamphaus et al., 1997, p.35). With some refinements over the years, interpretation of intelligence tests continues to be based on this type of classification system. That is, distinctions are still made between individuals who are mentally retarded and gifted, for example. Our classification categories are quite different from earlier classification systems, as you will see in Chapter 4.

It appears that Wechsler accepted the prevailing ideas regarding *g* and the conceptualization of intelligence as a global entity, consistent with those already put forth by Terman, Binet, Spearman, and others (Reynolds & Kaufman, 1990), when he offered his own definition of intelligence. According to Wechsler (1939), *intelligence* is "the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment" (p. 3). He concluded that this definition "avoids singling out any ability, however esteemed (e.g., abstract reasoning), as crucial or overwhelmingly important" and implies that any one intelligence subtest is readily interchangeable with another (p. 3).

The Second Wave: Clinical Profile Analysis

Kamphaus and colleagues (1997; Kamphaus et al., 2005) identified the second wave of interpretation as *clinical profile analysis* and stated that the publication of the Wechsler-Bellevue (W-B; Wechsler, 1939) was pivotal in spawning this approach to interpretation. Clinical profile analysis was a method designed to go beyond global IQ and interpret more specific aspects of an individual's cognitive capabilities through the analysis of patterns of subtest scaled scores.

The Wechsler-Bellevue Intelligence Scale, Form I (W-B I), published in 1939 (an alternate form—the W-B II—was published in 1946), represented an approach to intellectual assessment of adults that was clearly differentiated from other instruments available at that time (e.g., the Binet scales). The W-B was composed of 11 separate subtests, including Information, Comprehension, Arithmetic, Digit Span, Similarities, Vocabulary, Picture Completion, Picture Arrangement, Block Design, Digit Symbol, and Coding. (The Vocabulary subtest was an alternate for W-B I.)

Perhaps the most notable feature introduced with the W-B, which advanced interpretation beyond classification of global IQ, was the grouping of subtests into Verbal and Performance composites. The Verbal-Performance dichotomy represented an organizational structure that was based on the notion that intelligence could be expressed and measured through both verbal and nonverbal communication modalities. To clarify the Verbal-Performance distinction, Wechsler asserted that this dichotomy "does not imply that these are the only abilities involved in the tests. Nor does it presume that there are different kinds of intelligence, e.g., verbal, manipulative, etc. It merely implies that these are different ways in which intelligence may manifest itself" (Wechsler, 1958, p. 64).

Another important feature pioneered in the W-B revolved around the construction and organization of subtests. At the time, the Binet scale was ordered and administered sequentially according to developmental age, irrespective of the task. In contrast, Wechsler utilized only 11 subtests, each scored by points rather than age, and each with sufficient range of item difficulties to encompass the entire age range of the scale.

In his writings, Wechsler often shifted between conceptualizing intelligence as either a singular entity (the first wave) or a collection of specific mental abilities. At times he appeared to encourage the practice of subtest-level interpretation, suggesting that each subtest measured a relatively distinct cognitive ability (McDermott et al., 1990). To many, this position appeared to contradict his prior attempts not to equate general intelligence with the sum of separate cognitive or intellectual abilities. This shift in viewpoint may have been responsible, in part, for the development of interpretive methods such as profile analysis (Flanagan et al., 2001).

Without a doubt, the innovations found in the W-B were impressive, practical, and in many ways superior to other intelligence tests available in 1939. More importantly, the structure and organization of the W-B scale provided the impetus for Rapaport, Gill, and Schafer's (1945–1946) innovative approaches to test interpretation, which included an attempt to understand the meaning behind the shape of a person's profile of scores. According to Kamphaus and colleagues (Kamphaus et al., 1997; Kamphaus et al., 2005), a new method of test interpretation had developed under the assumption that "patterns of high and low subtest scores could presumably reveal diagnostic and psychotherapeutic considerations" (Kamphaus et al., 1997, p. 36). Thus, during the second wave of intelligence test interpretation, the W-B (1939) was the focal point from which a variety of interpretive procedures were developed for deriving diagnostic and prescriptive meaning from the shape of subtest profiles and the difference between Verbal and Performance IQs.

In addition to the scope of Rapaport and colleagues' (1945–1946) diagnostic suggestions, their approach to understanding profile shape led to a flurry of investigations that sought to identify the psychological functions underlying an infinite number of profile patterns and their relationships to each other. Perhaps as a consequence of the clinical appeal of Rapaport and colleagues' approach, Wechsler (1944) helped to relegate general-level assessment to the back burner while increasing the heat on clinical profile analysis.

The search for meaning in subtest profiles and IQ differences was applied to the WISC (Wechsler, 1949), a downward extension of the W-B II. The WISC was composed of the same 11 subtests used in the W-B II but was modified to assess intellectual functioning in children within the age range of 5 to 15 years. Subtests were grouped into the verbal and performance categories, as they were in the W-B II, with Information, Comprehension, Arithmetic, Digit Span, Similarities, and Vocabulary composing the Verbal Scale and Picture Completion, Picture Arrangement, Block Design, Digit Symbol, and Coding composing the Performance Scale. The WISC provided scaled scores for each subtest and yielded the same composites as the W-B II: Full Scale IQ (FSIQ), Verbal IQ (VIQ), and Performance IQ (PIQ).

Although the search for diagnostic meaning in subtest profiles and IQ differences was a more sophisticated approach to intelligence test interpretation as compared to the interpretive method of the first wave, it also created methodological problems. For example, with enough practice, just about any astute clinician could provide a seemingly rational interpretation of an obtained profile to fit the known functional patterns of the examinee. Nonetheless, analysis of profile shape and IQ differences did not result in diagnostic validity for the WISC. The next wave in intelligence

test interpretation sought to address the methodological flaws in the clinical profile analysis method (Kamphaus et al., 1997; Kamphaus et al., 2005).

The Third Wave: Psychometric Profile Analysis

In 1955, the original W-B was revised and updated and its new name—Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955)—was aligned with the existing juvenile version (i.e., WISC). Major changes and revisions included (a) incorporating Forms I and II of the W-B into a single scale with a broader range of item difficulties; (b) realigning the target age range to include ages 16 years and older (which eliminated overlap with the WISC, creating a larger and more representative norm sample); and (c) refining the subtests to improve reliability.

Within this general time period, technological developments in the form of computers and readily accessible statistical software packages to assist with intelligence test interpretation provided the impetus for what Kamphaus and colleagues (1997; Kamphaus et al., 2005) called the *third wave* of interpretation—*psychometric profile analysis*. The work of Cohen (1959), which was based primarily on the WISC and the then-new WAIS (Wechsler, 1955), sharply criticized the clinical profile analysis tradition that defined the second wave. For example, Cohen's factor-analytic procedures revealed a viable three-factor solution for the WAIS that challenged the dichotomous Verbal-Performance model and remained the de facto standard for the Wechsler scales for decades and for the WISC, in particular, until its third and fourth editions. The labels used by Cohen for the three Wechsler factors that emerged in his factor analysis of the WISC subtests (i.e., Verbal Comprehension, Perceptual Organization, Freedom from Distractibility) were the names of the Indexes on two subsequent editions of this test (WISC-R and WISC-III), spanning more than 2 decades.

By examining and removing the variance shared between subtests, Cohen demonstrated that the majority of Wechsler subtests had very poor *specificity* (i.e., reliable, specific variance). Thus, the frequent clinical practice of interpreting individual subtests as reliable measures of a *presumed* construct was not supported. Kamphaus and colleagues (1997; Kamphaus et al., 2005) summarize Cohen's significant contributions, which largely defined the third wave of test interpretation, as threefold: (a) empirical support for the FSIQ based on analysis of shared variance between subtests; (b) development of the three-factor solution for interpretation of the Wechsler scales; and (c) revelation of limited subtest specificity, questioning individual subtest interpretation.

The most vigorous and elegant application of psychometric profile analysis to intelligence test interpretation occurred with the revision of the venerable WISC

as the Wechsler Intelligence Scale for Children–Revised (WISC-R; Wechsler, 1974). Briefly, the WISC-R utilized a larger, more representative norm sample than its predecessor; included more contemporary-looking graphics and updated items; eliminated content that was differentially familiar to specific groups; and included improved scoring and administration procedures. "Armed with the WISC-R, Kaufman (1979) articulated the essence of the psychometric profile approach to intelligence test interpretation in his seminal book, *Intelligent Testing with the WISC-R* (which was superseded by *Intelligent Testing with the WISC-III;* Kaufman, 1994)" (Flanagan et al., 2000, p. 6).

Kaufman emphasized flexibility in interpretation and provided a logical and systematic approach that utilized principles from measurement theory (Flanagan & Alfonso, 2000). His approach was more complex than previous ones and required the examiner to have a greater level of psychometric expertise than might ordinarily be possessed by the average psychologist (Flanagan et al., 2000). Anastasi (1988) lauded and recognized that "the basic approach described by Kaufman undoubtedly represents a major contribution to the clinical use of intelligence tests. Nevertheless, it should be recognized that its implementation requires a sophisticated clinician who is well informed in several fields of psychology" (p. 484).

In some respects, publication of Kaufman's work can be viewed as an indictment against the poorly reasoned and unsubstantiated interpretation of the Wechsler scales that had sprung up in the second wave (clinical profile analysis; Flanagan et al., 2000). Kaufman's ultimate message centered on the notion that interpretation of Wechsler intelligence test performance must be conducted with a higher than usual degree of psychometric precision and based on credible and dependable evidence, rather than merely the clinical lore that surrounded earlier interpretive methods.

Despite the enormous body of literature that has mounted over the years regarding profile analysis of the Wechsler scales, this form of interpretation, even when upgraded with the rigor of psychometrics, has been regarded as a perilous endeavor primarily because it lacks empirical support and is not grounded in a well-validated theory of intelligence. With over 75 different profile types discussed in a variety of areas, including neuropsychology, personality, learning disabilities, and juvenile delinquency (McDermott et al., 1990), there is considerable temptation to believe that the findings of this type of analysis alone are reliable. Nevertheless, many studies (e.g., Hale, 1979; Hale & Landino, 1981; Hale & Saxe, 1983) have demonstrated consistently that "profile and scatter analysis is not defensible" (Kavale & Forness, 1984, p. 136; also see Glutting, McDermott, Watkins, Kush, & Konold, 1997). In a meta-analysis of 119 studies of the WISC-R subtest data, Mueller, Dennis, and Short (1986) concluded that using profile analysis with

the WISC-R in an attempt to differentiate various diagnostic groups is clearly not warranted. More recent evaluations regarding the merits of profile analysis have produced similar results (e.g., Borsuk, Watkins, & Canivez, 2006; Glutting, McDermott, & Konold, 1997; Glutting, McDermott, Watkins, et al., 1997; Kamphaus, 1993; McDermott, Fantuzzo, Glutting, Watkins, & Baggaley, 1992; Watkins & Kush, 1994). The nature of the controversy surrounding clinical profile analysis is discussed later in this chapter.

The Fourth Wave: Application of Theory

Although the third wave of intelligence test interpretation did not meet with great success in terms of establishing validity evidence for profile analysis, the psychometric approach provided the foundation necessary to catapult to the fourth and present wave of intelligence test interpretation, described by Kamphaus and colleagues (1997; Kamphaus et al., 2005) as *application of theory*. The need to integrate theory and research in the intelligence test interpretation process was articulated best by Kaufman (1979). Specifically, Kaufman commented that problems with intelligence test interpretation can be attributed largely to the lack of a specific theoretical base to guide this practice. He suggested that it was possible to enhance interpretation significantly by reorganizing subtests into clusters specified by a particular theory. In essence, the end of the third wave of intelligence test interpretation and the beginning of the fourth wave was marked by Kaufman's pleas for practitioners to ground their interpretations in theory, as well as by his efforts to demonstrate the importance of linking intellectual measurement tools to empirically supported and well-established conceptualizations of human cognitive abilities (Flanagan et al., 2000, 2008).

Despite efforts to meld theory with intelligence test development and interpretation, the WISC-III (Wechsler, 1991), published nearly 2 decades after the WISC-R (Wechsler, 1974), failed to ride the fourth, *theoretical* wave of test interpretation. That is, the third edition of the WISC did not change substantially from its predecessor and was not overtly linked to theory. Changes to the basic structure, item content, and organization of the WISC-III were relatively minimal, with the most obvious changes being cosmetic. However, the WISC-III did introduce one new subtest (Symbol Search) and four new Indexes—namely Verbal Comprehension (VC), Perceptual Organization (PO), Freedom from Distractibility (FD), and Processing Speed (PS)—to supplement the subtest scaled scores and the FSIQ, VIQ, and PIQ. As with the WISC-R, Kaufman provided a systematic approach to interpreting the WISC-III in a manner that emphasized psychometric rigor and theory-based methods (Kaufman, 1994; Kaufman & Lichtenberger, 2000).

Similar to Kaufman's efforts to narrow the theory-practice gap in intelligence test development and interpretation, Flanagan and colleagues (Flanagan & Ortiz, 2001; Flanagan, Ortiz, & Alfonso, 2007; Flanagan, Ortiz, Alfonso, & Mascolo, 2006; Flanagan et al., 2000; McGrew & Flanagan, 1998) developed a method of assessment and interpretation called the *Cross-Battery approach* and applied it to the Wechsler scales and other major intelligence tests. This method is grounded in CHC theory and provides a series of steps and guidelines that are designed to ensure that science and practice are closely linked in the measurement and interpretation of cognitive abilities. According to McGrew (2005), the Cross-Battery approach infused CHC theory into the minds of assessment practitioners and university training programs, regardless of their choice of favorite intelligence battery. Kaufman's (2001) description of the Cross-Battery approach as an interpretive method that (a) has "research as its foundation," (b) "add[ed] theory to psychometrics," and (c) "improve[d] the quality of the psychometric assessment of intelligence" (p. xv) is consistent with Kamphaus's (1997; Kamphaus et al., 2005) fourth wave of intelligence test interpretation (i.e., application to theory).

Despite the availability of theory-based systems for interpreting the WISC-III (and other intelligence tests), the inertia of tradition was strong, leading many practitioners to continue using interpretive methods of the second and third waves (Alfonso et al., 2000). A few critics, however, did not succumb and instead evaluated this latest version of the WISC according to the most current and dependable evidence of science. These reviews were not positive and their conclusions were remarkably similar—the newly published WISC-III was *outdated*. According to Kamphaus (1993), "the Wechsler-III's history is also its greatest liability. Much has been learned about children's cognitive development since the conceptualization of the Wechsler scales, and yet few of these findings have been incorporated into revisions" (p. 156). Similarly, Shaw, Swerdlik, and Laurent (1993) concluded that, "despite more than 50 years of advancement of theories of intelligence, the Wechsler philosophy of intelligence, written in 1939, remains the guiding principle of the WISC-III. . . . [T]he latest incarnation of David Wechsler's test may be nothing more than a new and improved dinosaur."

Notwithstanding initial criticisms, the several years that followed the publication of the WISC-III can be described as *the calm before the storm*. That is, the WISC-III remained the dominant intelligence test for use with children aged 6 to 16 with little more in the way of critical analysis and review. With the advent of the 21st century, however, the CHC storm hit and has not changed its course to date. In the past 8 years, revisions of four major intelligence tests were published, each one having CHC theory at its base (i.e., WJ III, SB5, KABC-II, DAS-II).

Never before in the history of intelligence testing has a single theory (indeed any theory) played so prominent a role in test development and interpretation. Amidst the publication of these CHC-based instruments was the publication of the WISC-IV. Was it structurally different from the WISC-III? Did it have theory at its base? These questions will be answered in the paragraphs that follow; suffice it to say that the WISC-IV represents the most significant revision of any Wechsler scale in the history of the Wechsler lineage, primarily because of its closer alliance with theory. A brief timeline of the revisions to the Wechsler scales, from the mid-1940s to the present day, and their correspondence to interpretive approaches, is located in Figure 1.2.

Although we have associated our own methods of Wechsler scale interpretation with the fourth wave—application to theory—our methods continue to be criticized because they include an intraindividual analysis component. We believe these criticisms are largely unfounded, primarily because our methods have not been critiqued as a whole; rather Watkins and colleagues have critiqued only one aspect of our systems—intraindividual analysis—and concluded that because their research shows that ipsative subtest scores are less reliable and less stable than normative subtest scores, any conclusions that are drawn from ipsative analysis are unsupported. Notwithstanding the problems with this conclusion, our current interpretive approaches do not involve subtest-level analysis. The intraindividual analysis component of our interpretive approaches focuses on cluster-level, not subtest-level, analysis (Flanagan & Kaufman, 2004; Flanagan & Ortiz, 2001; Kaufman & Kaufman, 2004a). Because there is continued debate about the utility of intraindividual analysis, especially as it applies to Wechsler test interpretation, the following section provides a brief review of the most salient debate issues as well as a justification for the interpretive approach we continue to advocate in this new edition, found in Chapter 4.

THE CONTINUING DEBATE ABOUT THE UTILITY OF INTRAINDIVIDUAL (IPSATIVE) ANALYSIS

Since the early 1990s, Glutting, McDermott, and colleagues "have used their research as an obstacle for clinicians, as purveyors of gloom-and-doom for anyone foolish enough to engage in profile interpretation" (Kaufman, 2000a, p. xv). These researchers have shown that ipsative scores have poor reliability, are not stable over time, and do not add anything to the prediction of achievement after *g* (or general intelligence) is accounted for. Thus, Glutting and colleagues believe that ipsative analysis has virtually no utility with regard to (a) understanding a child's unique pattern of cognitive strengths and weaknesses or (b) aiding in

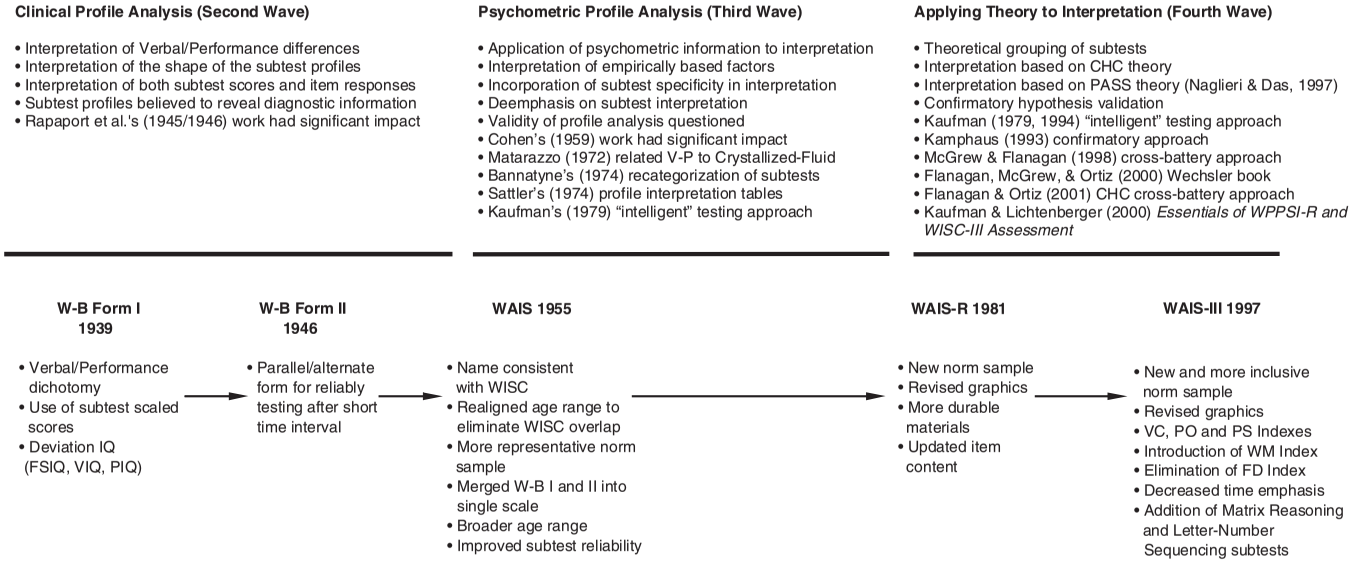

Figure 1.2 Timeline of Revisions to Wechsler Scales and Corresponding Interpretive Methods

Source: Wechsler Intelligence Scales for Children–Fourth Edition. Copyright © 2004 by Harcourt Assessment, Inc. Reproduced by permission. All rights reserved.

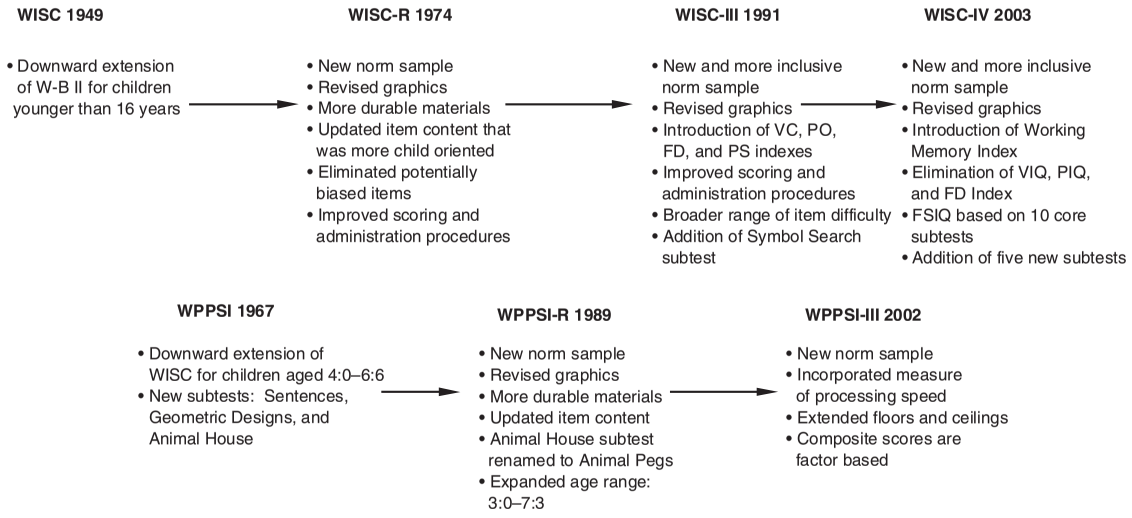

Figure 1.2 (Continued)

Note: The first wave of interpretation (quantification of general level) is omitted from this figure due to space limitations and the fact that the publication of the first Wechsler Scale did not occur until near the end of that wave. W-B = Wechsler-Bellevue; FSIQ = Full Scale IQ; VIQ = Verbal IQ; PIQ = Performance IQ; VC = Verbal Comprehension; PO = Perceptual Organization; FD = Freedom from Distractibility; PS = Processing Speed; WMI = Working Memory Index. See Figure 1.1 note for other abbreviations.

developing educational interventions. It is beyond the scope of this chapter to provide a detailed discussion of the numerous arguments that have been made for and against ipsative analysis in the past decade. Therefore, we only comment briefly on the whole of Glutting and colleagues' research and then describe how our interpretive method, which includes (but by no means is defined by) intraindividual analysis, differs substantially from previous interpretive methods.

In much of their writing, Glutting and colleagues have assumed incorrectly that all cognitive abilities represent enduring traits and, therefore, ought to remain stable over time. They further assume that interpretations of test data are made in a vacuum—that data from multiple sources, no matter how compelling, cannot influence the findings generated from an ipsative analysis of scores from a single intelligence battery. Furthermore, the method of test interpretation initially developed by Kaufman (1979) remains the focus of Glutting and colleagues' research, despite the fact that it has changed considerably in recent years (Flanagan & Kaufman, 2004; Flanagan & Ortiz, 2001; Flanagan et al., 2006, 2007; Kaufman & Lichtenberger, 2006 ; Kaufman, Lichtenberger, Fletcher-Janzen, & Kaufman, 2005). Interestingly, these changes reflect, in part, the research of Glutting and colleagues (e.g., McDermott et al., 1992). Perhaps most disturbing is the fact that these researchers continue their cries of "Just Say No" to *any* type of interpretation of test scores beyond a global IQ, and they offer *no* recommendations regarding how clinicians can make sense out of an individual's scaled score profile (e.g., Borsuk et al., 2006; Oh, Glutting, Watkins, Youngstrom, & McDermott, 2004).

We, on the other hand, recognize the onerous task facing clinicians in their daily work of identifying the presumptive cause of a child's learning difficulties. Hence we provide clinicians with guidance in the test interpretation process that is based on theory, research, psychometrics, and clinical experience. What Glutting and colleagues have yet to realize is that our interpretive method extends far beyond the identification of intraindividual (or ipsative) strengths and weaknesses.

Despite its inherent flaws, we believe that intraindividual analysis has not fared well because it historically has not been grounded in contemporary theory and research and it has not been linked to psychometrically defensible procedures for interpretation (Flanagan & Ortiz, 2001). When theory and research are used to guide interpretation and when psychometrically defensible interpretive procedures are employed, *some* of the limitations of the intraindividual approach are circumvented, resulting in the derivation of useful information. Indeed, when an interpretive approach is grounded in contemporary theory and research, practitioners are in a much better position to draw clear and useful conclusions from the data (Carroll, 1998; Daniel, 1997; Flanagan et al., 2006, 2007, 2008; Kamphaus, 1993; Kamphaus et al., 1997; Keith, 1988).

The findings of an intraindividual analysis are not the end of the interpretation process, but only the beginning. We do find many flaws with the purely empirical approach that Glutting and colleagues have used to evaluate the traditional approach to profile interpretation. Nonetheless, we have taken quite seriously many of the criticisms of a purely ipsative method of profile analysis that have appeared in the literature in articles by Watkins, Glutting, and their colleagues (e.g., Borsuk et al., 2006; McDermott et al., 1992; Oh et al., 2004). Indeed, one of us (DPF) has been frankly critical of ipsative analysis that ignores normative analysis (Flanagan & Ortiz, 2002a, 2002b). We have relied on all of these criticisms to modify and enhance our interpretive method. Following are a few of the most salient ways in which we and our colleagues have attempted to improve the practice of ipsative analysis (Flanagan & Kaufman, 2004; Flanagan & Ortiz, 2001; Kaufman & Kaufman, 2004).

First, we recommend interpreting test data within the context of a well-validated theory. Use of the CHC theory of the structure of cognitive abilities is commonplace in test construction and interpretation because it is the best-supported theory within the psychometric tradition (Daniel, 1997; Flanagan & Harrison, 2005; Flanagan & Ortiz, 2001; McGrew, 2005). Without knowledge of theory and an understanding of its research base, there is virtually no information available to inform interpretation.

Second, we recommend using composites or clusters, rather than subtests, in intraindividual analysis. Additionally, the clusters that are used in the analysis must represent *unitary* abilities, meaning that the magnitude of the difference between the highest and lowest scores in the cluster is not statistically significant ($p < .01$; see Chapter 4 for an explanation). Furthermore, the clusters that are included in the interpretive analysis should represent basic primary factors in mental organization (e.g., visual processing, short-term memory). When the variance that is common to all clusters (as opposed to subtests) is removed during ipsatization, *proportionately more reliable variance remains.* And it is precisely this shared, reliable variance that we believe ought to be interpreted because it represents the construct that was intended to be measured by the cluster. For example, when the following clusters are ipsatized—Fluid Reasoning (*Gf*), Crystallized Intelligence (*Gc*), Short-Term Memory (*Gsm*), Visual Processing (*Gv*), and Long-Term Storage and Retrieval (*Glr*)—the variance that is common to all of them (presumably *g*) is removed, leaving the variance that is shared by the two or more tests that compose each cluster. That is, if the *Gf* cluster emerged as a significant relative weakness, then our interpretation would focus on what is common to the *Gf* tests (viz., reasoning). The number of research investigations examining the relationship between broad CHC clusters and various outcome criteria

(e.g., academic achievement) provide important validation evidence that may be used to inform the interpretive process (Evans, Floyd, McGrew, & Leforgee, 2002; Flanagan, 2000; Floyd, Evans, & McGrew, 2003; Floyd, Keith, Taub, & McGrew, 2007; McGrew, Flanagan, Keith, & Vanderwood, 1997; Reeve, 2004; Taub, Keith, Floyd, & McGrew, 2008; Vanderwood, McGrew, Flanagan, & Keith, 2002). Much less corresponding validity evidence is available to support traditional ipsative (subtest) analysis.

Third, we believe that a common pitfall in the intraindividual approach to interpretation is the failure to examine the scores associated with an identified *relative weakness* in comparison to most people. That is, if a relative weakness revealed through ipsative analysis falls well within the average range of functioning compared to most people, then its clinical meaningfulness is called into question. For example, despite presumptions of disability, average ability is achieved by most people and most people are not disabled. Therefore, a relative weakness that falls in the average range of ability compared to same-age peers will suggest a different interpretation than a relative weakness that falls in the deficient range of functioning relative to most people.

Fourth, we believe that the lack of stability in an individual's scaled score profile over an extended period of time (e.g., the 3 years spanning initial evaluation and reevaluation) is not unusual, let alone a significant flaw of intraindividual analysis. A great deal happens in 3 years: the effects of intervention. Developmental changes. Regression to the mean. Changes in what some subtests measure at different ages. The group data that have been analyzed by Glutting and colleagues do not have implications for the individual method of profile interpretation that we advocate. The strengths and weaknesses that we believe might have useful applications for developing educational interventions are based on cognitive functioning at a particular point in time. They need to be cross-validated at that time to verify that any supposed cognitive strengths or weaknesses are consistent with the wealth of observational, referral, background, and other-test data that are available for each child who is evaluated. Only then will those data-based findings inform diagnosis and be useful in developing interventions to help the child.

The simple finding that reevaluation data at age 13 do not support the stability of children's data-based strengths and weaknesses at age 10 says *nothing* about the validity of the intraindividual interpretive approach. If one's blood pressure is "high" when assessed in January and is "normal" when assessed 3 months later, does this suggest that the physician's categories (e.g., high, normal, low) are unreliable? Does it suggest that the blood-pressure monitor is unreliable? Or does it suggest that the medication prescribed to reduce the individual's blood pressure was effective?

Despite the pains taken to elevate the use of ipsative analysis to a more respectable level, by linking it to normative analysis and recommending that only unitary, theoretically derived clusters be used, one undeniable fact remains. The intraindividual analysis does not diagnose—clinicians do. Clinicians, like medical doctors, will not cease to compare scores, nor should they:

> Would one want a physician, for example, not to look at patterns of test results just because they in and of themselves do not diagnose a disorder? Would you tell a physician not to take your blood pressure and heart rate and compare them because these two scores in and of themselves do not differentially diagnose kidney disease from heart disease? (Prifitera, Weiss, & Saklofske, 1998, p. 6)

Comparing scores from tests, whether psychological or medical, is a necessary component of any test interpretation process. Why? We believe it is because comparing scores assists in making diagnoses when such comparisons are made using psychometric information (e.g., base-rate data) as well as numerous other sources of data, as mentioned previously (e.g., Ackerman & Dykman, 1995; Flanagan et al., 2007; Hale & Fiorello, 2004; Hale, Fiorello, Kavanagh, Hoeppner, & Gaither, 2001). The learning disability literature appears to support our contention. For example, the *double-deficit hypothesis* states that individuals with reading disability have two main deficits relative to their abilities in other cognitive areas, including phonological processing and rate, or rapid automatized naming (e.g., Wolf & Bowers, 2000). Moreover, in an evaluation of subtypes of reading disability, Morris and colleagues (1998) found that phonological processing, verbal short-term memory and rate (or rapid automatized naming) represented the most common profile, meaning that these three abilities were significantly lower for individuals with reading disability as compared to their performance on other measures of ability. Similarly, other researchers have argued for profile analysis beyond the factor or Index level (e.g., Flanagan et al., 2007; Kramer, 1993; Nyden, Billstedt, Hjelmquist, & Gillberg, 2001), stating that important data would be lost if analysis ceased at the global ability level.

Indeed, this is not the first place that the flaws of the purely empirical approaches advocated by Glutting, McDermott, Watkins, Canivez, and others have been articulated, especially regarding the power of their group-data methodology for dismissing individual-data assessment. Anastasi and Urbina (1997) state,

> One problem with several of the negative reviews of Kaufman's approach is that they seem to assume that clinicians will use it to make decisions based solely on the magnitude of scores and score differences. While it is true that

> the mechanical application of profile analysis techniques can be very misleading, this assumption is quite contrary to what Kaufman recommends, as well as to the principles of sound assessment practice. (p. 513)

The next and final section of this chapter provides specific information about the WISC-IV from a qualitative, quantitative, and theoretical perspective.

DESCRIPTION OF THE WISC-IV

Several issues prompted the revision of the WISC-III. These issues are detailed clearly in the *WISC-IV Technical and Interpretive Manual* (The Psychological Corporation, 2003, pp. 5–18). Table 1.1 provides general information about the WISC-IV. In addition, Rapid Reference 1.1 lists the key features of the WISC-IV, and Rapid Reference 1.2 lists the most salient changes from the WISC-III to WISC-IV. Finally, Rapid References 1.3 and 1.4 include the CHC broad and narrow ability classifications of the WISC-IV subtests.

Although you will recognize many traditional WISC subtests on the WISC-IV, you will also find five new ones. The WISC-IV has a total of 15 subtests—10 core-battery subtests and 5 supplemental subtests. Table 1.2 lists and describes each WISC-IV subtest.

Structure of the WISC-IV

The WISC-IV has been modified in terms of its overall structure. Figure 1.3 depicts the theoretical and scoring structure of the WISC-IV as reported in the *WISC-IV Technical and Interpretive Manual* (The Psychological Corporation, 2003). Several structural changes from the WISC-III are noteworthy.

- The VCI is now composed of three subtests rather than four.
- Information is now a supplemental subtest.
- The POI has been renamed the PRI. In addition to Block Design, the PRI is composed of two new subtests, Matrix Reasoning and Picture Concepts, which are primarily measures of fluid reasoning. Fluid reasoning tasks are important as they have little dependence on cultural and educational background (Burns & O'Leary, 2004). Picture Completion is now a supplemental subtest. Object Assembly, Picture Arrangement, and Mazes have been dropped, all of which primarily measured visual processing. Picture Arrangement and Object Assembly were heavily dependent on bonus points for quick responses and presumably were dropped to put less emphasis on response time (Kaufman, Flanagan, Alfonso, & Mascolo, 2006).

Table 1.1 The WISC-IV at a Glance

GENERAL INFORMATION	
Author	David Wechsler (1896–1981)
Publication Date(s)	1949, 1974, 1991, 2003
Age Range	6:0 to 16:11
Administration Time	65 to 80 minutes
Qualification of Examiners	Graduate- or professional-level training in psychological assessment
Publisher	Pearson Assessments/The Psychological Corporation 19500 Bulverde Road San Antonio, TX 78259 Ordering Phone No. 1-800-211-8378 http://pearsonassess.com
Price WISC-IV™ Basic Kit	Includes Administration and Scoring Manual, Technical and Interpretive Manual, Stimulus Book 1, Record Form (pkg. of 25), Response Booklet 1 (Coding and Symbol Search; pkg. of 25), Response Booklet 2 (Cancellation; pkg. of 25), Blocks, Symbol Search Scoring Template, Coding Scoring Template, and Cancellation Scoring Templates. $950.00 (in box) or $1,006.00 (in hard- or soft-sided cases) **WISC-IV™ Scoring Assistant® $228.00** **WISC-IV™ Writer™ $462.00**
COMPOSITE MEASURE INFORMATION	
Global Ability	Full Scale IQ (FSIQ)
Lower Order Composites	Verbal Comprehension Index (VCI) Perceptual Reasoning Index (PRI) Working Memory Index (WMI) Processing Speed Index (PSI)
SCORE INFORMATION	
Available Scores	Standard Scaled Percentile Age Equivalent
Range of Standard Scores for Total Test Composite	40–160 (ages 6:0 to 16:11)

(*continued*)

Table 1.1 (Continued)

NORMING INFORMATION	
Standardization Sample Size	2,200
Sample Collection Dates	Aug. 2001–Oct. 2002
Average Number per Age Interval	200
Age Blocks in Norm Table	4 months (ages 6:0 to 16:11)
Demographic Variables	Age Gender (male, female) Geographic region (four regions) Race/ethnicity (White; African American; Hispanic; Asian; other) Socioeconomic status (parental education)
Types of Validity Evidence in Test Manual	Test content Response processes Internal structure Relationships with other variables Consequences of testing

Rapid Reference 1.1

Key Features Listed in the *WISC-IV Administration and Scoring Manual* (Wechsler, 2003)

- Includes several process scores that may enhance its clinical utility (see Chapters 6 and 7 for a discussion)
- Special group studies designed to improve its clinical utility
- Statistical linkage with measures of achievement (e.g., WIAT-II)
- Includes supplemental tests for core battery tests
- Provides computer scoring and interpretive profiling report
- Ability-Achievement discrepancy analysis available for FSIQ, VCI, and PRI with WIAT-II
- Wechsler Abbreviated Scale of Intelligence (WASI) prediction table (WASI FSIQ-4 and predicted WISC-IV FSIQ range at 68% and 90% confidence interval)
- Twelve subtests on WISC-III yielded four Indexes; 10 subtests on WISC-IV yield four Indexes
- Two manuals included in kit (Administration and Scoring; Technical and Interpretive)

Rapid Reference 1.2

Changes from the WISC-III to the WISC-IV

- Structural foundation updated to include measures of *Gf* and additional measures of *Gsm* (i.e., Letter-Number Sequencing) and *Gs* (i.e., Cancellation)
- Scoring criteria modified to be more straightforward
- Picture Arrangement, Object Assembly, and Mazes deleted (to reduce emphasis on time)
- Items added to improve floors and ceilings of subtests
- Instructions to examiners more understandable
- Artwork updated to be more attractive and engaging to children
- Increased developmental appropriateness (instructions modified; teaching, sample, and/or practice items for each subtest)
- Norms updated
- Outdated items replaced
- Manual expanded to include interpretation guidelines and more extensive validity information
- Weight of kit reduced by elimination of most manipulatives
- Arithmetic and Information moved to supplemental status
- Five new subtests added: Word Reasoning, Matrix Reasoning, Picture Concepts, Letter-Number Sequencing, and Cancellation
- VIQ and PIQ dropped
- FSIQ modified substantially to include only 5 of the 10 traditional Full Scale subtests
- Freedom from Distractibility (FD) Index replaced with a Working Memory Index
- Perceptual Organization Index (POI) renamed Perceptual Reasoning Index (PRI)
- Stimulus book has been changed so that the pages are turned toward the child
- Increased use of queries and prompts to improve children's understanding of the task
- WISC-IV record form includes an analysis page that can be used to calculate a child's relative strengths and weaknesses

Source: Information in this table is from the *WISC-IV Technical and Interpretive Manual* (The Psychological Corporation, 2003; Burns et al., 2004; and Kaufman et al., 2006).

Rapid Reference 1.3

WISC-IV Classifications

Subtest	Broad Ability Classifications Based on CFA of WISC-IV Standardization Data[a]	Broad and Narrow Ability Classifications Based on Expert Consensus[b]	
1. Block Design	***Gv***	***Gv***	**Spatial Relations**
2. Similarities	***Gc***	***Gc***	**Language Development**
			Lexical Knowledge
3. Digit Span	***Gsm***	***Gsm***	**Memory Span**
			Working Memory
4. Picture Concepts	***Gf***	***Gf***	**Induction**
		Gc	General Information
5. Coding	***Gs***	***Gs***	**Rate of Test-Taking**
6. Vocabulary	***Gc***	***Gc***	**Lexical Knowledge**
7. Letter-Number Sequencing	***Gsm***	***Gsm***	**Working Memory**
8. Matrix Reasoning	***Gf,*** *Gv*	***Gf***	**Induction and General Sequential Reasoning**

9. Comprehension	*Gc*	***Gc***	**General Information**
10. Symbol Search	***Gs,*** *Gv*	***Gs***	**Perceptual Speed**
			Rate of Test-Taking
11. Picture Completion	***Gv,*** *Gc*	***Gc***	**General Information**
		Gv	Flexibility of Closure
12. Cancellation	***Gs***	***Gs***	**Perceptual Speed**
			Rate of Test-Taking
13. Information	***Gc***	***Gc***	**General Information**
14. Arithmetic	***Gf* (especially older children)**	***Gq***	**Math Achievement**
	Gsm (especially younger children)	*Gf*	Quantitative Reasoning
15. Word Reasoning	***Gc***	***Gc***	**Lexical Knowledge**
		Gf	Induction

Note: Primary classifications appear in bold type. Secondary classifications appear in regular type. CFA = Confirmatory Factor Analysis.

[a]Keith, Fine, Taub, Reynolds, and Kranzler (2006).

[b]Alfonso, Flanagan, and Radwan (2005).

Rapid Reference 1.4

The Psychological Corporation's a Posteriori WISC-IV CHC Classifications

Subtest	Broad Ability Classifications of the WISC-IV Subtests (TPC®)[a]
Block Design	*Gv*
Similarities	*Gf*
Digit Span	*Gsm*
Picture Concepts	*Gf*
Coding	*Gs*
Vocabulary	*Gc, Glr*
Letter-Number Sequencing	*Gsm*
Matrix Reasoning	*Gf*
Comprehension	*Gc*[b]
Symbol Search	*Gs*
Picture Completion	*Gv*
Cancellation	*Gs*
Information	*Gc, Glr*
Arithmetic	*Gq, Gsm*
Word Reasoning	*Gf*

Note: TPC® = The Psychological Corporation.

[a]CHC constructs corresponding to WISC-IV Indexes were provided by The Psychological Corporation® after the publication of the WISC-IV and were obtained from a list of "WISC-IV Frequently Asked Questions (FAQs)" appearing on the Harcourt website.

[b]A classification for the WISC-IV Comprehension subtest was not available from the Harcourt website. The *Gc* classification denoted for the WISC-IV Comprehension subtest was based on previous classifications (e.g., Flanagan et al., 2000).

- Information and Arithmetic were moved to supplemental status, reducing the emphasis of the WISC-IV on school achievement.
- The FD Index has been renamed the WMI. The WMI is composed of Digit Span and the new Letter-Number Sequencing subtest. Arithmetic, which was formerly part of the FD Index, is now a supplemental subtest, minimizing the influence of math achievement on WMI (Kaufman et al., 2006).
- The PSI remains unchanged. However, a new speed-of-processing test—Cancellation—was added as a supplemental subtest.
- The Verbal IQ (VIQ) and Performance IQ (PIQ) were dropped. This change probably reflects the greatest change in interpretation of Wechsler scales. The VIQ-PIQ discrepancy was overused and its

Table 1.2 WISC-IV Subtest Definitions

Subtest	Description
1. Block Design (BD)	The examinee is required to replicate a set of modeled or printed two-dimensional geometric patterns using red-and-white blocks within a specified time limit.
2. Similarities (SI)	The examinee is required to describe how two words that represent common objects or concepts are similar.
3. Digit Span (DS)	On Digit Span Forward, the examinee is required to repeat numbers verbatim as stated by the examiner. On Digit Span Backward, the examinee is required to repeat numbers in the reverse order as stated by the examiner.
4. Picture Concepts (PCn)	The examinee is required to choose one picture, from among two or three rows of pictures presented, to form a group with a common characteristic.
5. Coding (CD)	The examinee is required to copy symbols that are paired with either geometric shapes or numbers using a key within a specified time limit.
6. Vocabulary (VC)	The examinee is required to name pictures or provide definitions for words.
7. Letter-Number Sequencing (LN)	The examinee is read a number and letter sequence and is required to recall numbers in ascending order and letters in alphabetical order.
8. Matrix Reasoning (MR)	The examinee is required to complete the missing portion of a picture matrix by selecting one of five response options.
9. Comprehension (CO)	The examinee is required to answer a series of questions based on his or her understanding of general principles and social situations.
10. Symbol Search (SS)	The examinee is required to scan a search group and indicate the presence or absence of a target symbol(s) within a specified time limit.

(*continued*)

Table 1.2 (Continued)

Subtest	Description
11. *Picture Completion (PCm)*	The examinee is required to view a picture and name the essential missing part of the picture within a specified time limit.
12. *Cancellation (CA)*	The examinee is required to scan both a random and a nonrandom arrangement of pictures and mark target pictures within a specified time limit.
13. *Information (IN)*	The examinee is required to answer questions that address a wide range of general-knowledge topics.
14. *Arithmetic (AR)*	The examinee is required to mentally solve a variety of orally presented arithmetic problems within a specified time limit.
15. *Word Reasoning (WR)*	The examinee is required to identify a common concept being described by a series of clues.

Note: Subtests printed in italics are supplemental.

meaningfulness and clinical utility were never made clear in the literature (Kaufman & Lichtenberger, 2006; Kaufman et al., 2006).

- The four Indexes are derived from 10 subtests rather than 12.
- 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: Similarities, Comprehension, Vocabulary, Block Design, and Coding. Among the five new Full Scale subtests, three are from the WMI and PSI.

The *WISC-IV Technical and Interpretive Manual* (The Psychological Corporation, 2003) provides a series of exploratory and confirmatory factor analyses that offer support for the factor structure of the test, depicted in Figure 1.3. Specifically, four factors underlie the WISC-IV—namely Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed. The structural validity of the WISC-IV is discussed further in the following paragraphs.

Standardization and Psychometric Properties of the WISC-IV

Standardization

The WISC-IV was standardized on a sample of 2,200 children who were chosen to match closely the 2002 U.S. Census data on the variables of age, gender,

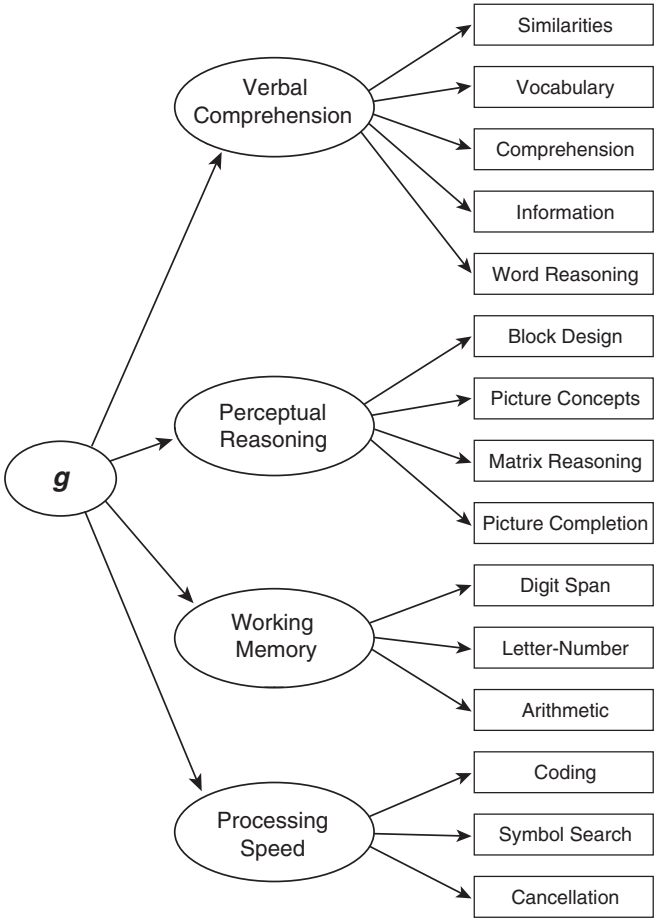

Figure 1.3 The Organization of the WISC-IV

geographic region, ethnicity, and socioeconomic status (SES; parental education). The standardization sample was divided into 11 age groups, each composed of 200 children. The sample was split equally between boys and girls (see Table 1.1).

The WISC-IV has also been adapted and standardized in Canada, the United Kingdom, France, Australia, and Germany; is currently in standardization in Japan (T. Ishikuma, personal communication, July 19, 2008); and will undoubtedly continue to be published throughout the world in many languages and cultures, as was its predecessor, the WISC-III (Georgas, Weiss, van de Vijver, & Saklofske, 2003). The test is the same in Canada as in the United States except for three questions in the Arithmetic subtest that were changed to imperial units of

measurements. Standardization norms for Canada can be found on the Harcourt Assessment website.

Reliability

The reliability of the WISC-IV is presented in its *Technical and Interpretive Manual* (The Psychological Corporation, 2003, Table 4.1, p. 34) and is summarized in Rapid Reference 1.5. The average internal consistency coefficients are 0.94 for VCI, 0.92 for PRI, .92 for WMI, .88 for PSI, and 0.97 for FSIQ. Internal consistency values for individual subtests across all ages ranged from 0.72 for Coding (for ages 6 and 7) to .94 for Vocabulary (for age 15). The median internal consistency values for the individual subtests ranged from .79 (Symbol Search, Cancellation) to .90 (Letter-Number Sequencing).

The WISC-IV is a stable instrument with average test–retest coefficients (corrected for variability of the sample) of 0.93, 0.89, 0.89, 0.86, and 0.93 for the VCI, PRI, WMI, PSI, and FSIQ, respectively (The Psychological Corporation, 2003, Table 4.4, p. 40). Rapid Reference 1.6 shows 1-month practice effects (gains from test to retest) for the WISC-IV Indexes and FSIQ for three separate age groups (i.e., 6–7, 8–11, and 12–16) and the overall sample. In general, practice effects are largest for ages 6 to 7 and become smaller with increasing age. As may be seen in Rapid Reference 1.6, average FSIQ gains dropped from about 8 points (ages 6–7) to 6 points (ages 8–11) to 4 points (ages 12–16). Rapid Reference 1.7 shows the WISC-IV subtests that demonstrated relatively large gains from test to retest. For ages 6 to 7, Coding and Symbol Search showed the largest gains, while Picture Completion showed the largest gains at ages 8 to 16. Other interesting facts about 1-month practice effects on the WISC-IV are found in Rapid Reference 1.8.

G-Loadings

G-loadings are an important indicator of the degree to which a subtest measures general intelligence. Additionally, *g*-loadings aid in determining the extent to which a single subtest score can be expected to vary from other scores within a profile. The WISC-IV subtest *g*-loadings are provided in Appendix C on the CD at the back of this book. Table C.1 in Appendix C provides WISC-IV subtest *g*-loadings by age groups and overall sample. These *g*-loadings represent the unrotated loadings on the first factor using the principle factor-analysis method. This method assumes that *g* influences the subtests indirectly through its relationship with the four factors. Table C.1 shows that the VCI subtests generally have the highest *g*-loadings at every age, followed by the PRI, WMI, and PSI subtests. Arithmetic, however, has *g*-loadings that are more consistent with the VCI subtest loadings as compared to the WMI core battery subtests. Table C.2 in Appendix C includes *g*-loadings

Rapid Reference 1.5

Average Reliability Coefficients of WISC-IV Subtests, Process Scores, and Composite Scales, Based on Total Sample

	Overall Reliability[a]
Subtest	
Block Design	.86
Similarities	.86
Digit Span	.87
Picture Concepts	.82
Coding	.85
Vocabulary	.89
Letter-Number Sequencing	.90
Matrix Reasoning	.89
Comprehension	.81
Symbol Search	.79
Picture Completion	.84
Cancellation	.79
Information	.86
Arithmetic	.88
Word Reasoning	.80
Process Score	
Block Design No Time Bonus	.84
Digit Span Forward	.83
Digit Span Backward	.80
Cancellation Random	.70
Cancellation Structured	.75
Composite Scale	
Verbal Comprehension Index	.94
Perceptual Reasoning Index	.92
Working Memory Index	.92
Processing Speed Index	.88
Full Scale	.97

Source: Information in this table was reproduced from the *WISC-IV Technical and Interpretive Manual* (The Psychological Corporation, 2003).

[a]Average reliability coefficients were calculated with Fisher's *z* transformation.

Rapid Reference 1.6

One-Month Practice Effects for the WISC-IV Indexes and Full-Scale IQ (Total *N* = 243)

Scale	Ages 6–7	Ages 8–11	Ages 12–16	All Ages
VCI	**+3.4**	**+2.2**	**+1.7**	**+2.1**
	(.31 SD)	(.20 SD)	(.14 SD)	(.18 SD)
PRI	**+6.4**	**+4.2**	**+5.4**	**+5.2**
	(.46 SD)	(.34 SD)	(.38 SD)	(.39 SD)
WMI	**+4.7**	**+2.8**	**+1.6**	**+2.6**
	(.33 SD)	(.22 SD)	(.12 SD)	(.20 SD)
PSI	**+10.9**	**+8.2**	**+4.7**	**+7.1**
	(.72 SD)	(.60 SD)	(.35 SD)	(.51 SD)
FSIQ	**+8.3**	**+5.8**	**+4.3**	**+5.6**
	(.62 SD)	(.53 SD)	(.34 SD)	(.46 SD)

Source: Data are from *WISC-IV Technical and Interpretive Manual* (The Psychological Corporation, 2003, Table 4.4).

Note: Intervals ranged from 13 to 63 days, with a mean of 32 days.

for the overall sample from the last column in Table C.1 alongside *g*-loadings based on confirmatory factor analysis (CFA) using a nested factors model. This latter method assumes that each subtest has a distinct and direct relationship with both *g* and a broad ability (factor; Keith, 2006). Therefore, the *g*-loadings in the second column of Table C.2 were derived in a manner more consistent with the factor and scoring structure of the WISC-IV. Table C.2 shows that subtest *g*-loadings are generally consistent across methods, with two exceptions—both Word Reasoning and Comprehension had high *g*-loadings (.70 or greater) based on the principal factor-analysis method and medium *g*-loadings (.51 to .69) based on the CFA (nested factors) method. These *g*-loadings may be useful in generating hypotheses about fluctuations in a child's scaled score profile.

Floors, Ceilings, and Item Gradients

The floors and ceilings for all WISC-IV subtests are excellent, indicating that scaled scores greater than 2 SDs above and 2 SDs below the mean may be obtained on

Rapid Reference 1.7

One-Month Practice Effects for the Separate WISC-IV Scaled Scores: Subtests with Relatively Large Gains from Test to Retest

Ages 6–7	Ages 8–11	Ages 12–16
Coding (+0.65 SD)	Picture Completion (+0.68 SD)	Picture Completion (+0.58 SD)
Symbol Search (+0.62 SD)	Symbol Search (+0.52 SD)	Cancellation (+0.44 SD)
Picture Completion (+0.58 SD)	Picture Concepts (+0.52 SD)	Coding (+0.40 SD)
Arithmetic (+0.57 SD)	Cancellation (+0.47 SD)	Block Design (+0.40 SD)
Picture Concepts (+0.50 SD)	Block Design (+0.40 SD)	Picture Concepts (+0.35 SD)
Block Design (+0.45 SD)		
Similarities (+0.45 SD)		
Word Reasoning (+0.42 SD)		
Letter-Number Sequencing (+0.39 SD)		

Source: Data are from *WISC-IV Technical and Interpretive Manual* (The Psychological Corporation, 2003, Table 4.4).

Note: Relatively large gains are defined as at least 0.33 SD (a gain from test to retest of approximately 1.0 scaled score point, depending on the precise SDs at each age). Gains are listed by the magnitude of the gain for each age group. Intervals ranged from 13 to 63 days, with a mean of 32 days.

Rapid Reference 1.8

Interesting Facts about One-Month Practice Effects on the WISC-IV

- WISC-IV practice effects (gains from test to retest) are largest for ages 6 to 7 and become smaller with increasing age. Average FSIQ gains dropped from about 8 points (ages 6–7) to 6 points (ages 8–11) to 4 points (ages 12–16). See Rapid Reference 1.6.
- The age-related changes in practice effects held for VCI, WMI, and PSI, but not for PRI. The PRI, which measures the *performance* abilities that traditionally yield the largest practice effects, averaged test–retest gains of about 5 points across the age range (see Rapid Reference 1.6).
- Despite the very large practice effect of 11 points (.72 SD) for ages 6 to 7 on PSI, this age group showed no practice effect at all on Cancellation, the supplemental Processing Speed subtest. In contrast, Cancellation produced among the largest practice effects for ages 8 to 16 (effect sizes of about 0.45 SD; see Rapid Reference 1.7).
- Arithmetic and Letter-Number Sequencing, both measures of Working Memory, had substantial practice effects at ages 6 to 7 (see Rapid Reference 1.7), but yielded little or no gains for all other age groups.
- Picture Completion had by far the largest practice effect for all ages combined (0.60 SD). It joins Picture Concepts and Block Design as the only WISC-IV subtests to yield relatively large test–retest gains for each age group studied: 6 to 7, 8 to 11, and 12 to 16 (see Rapid Reference 1.7).
- Practice effects for Digits Forward and Digits Backward varied as a function of age. For ages 6 to 11, test–retest gains were larger for Digits Backward (effect size of 0.19 SD versus 0.12 SD for Digits Forward). For ages 12 to 13, gains were about equal for Digits Forward and Digits Backward. For ages 14 to 16, test–retest gains were larger for Digits Forward (effect size of 0.29 SD versus 0.11 SD for Digits Backward).

all subtests at all ages. Therefore, the WISC-IV may be used confidently as part of an evaluation for the identification of individuals who are functioning in either the gifted or mentally retarded ranges of functioning, respectively. Item gradients refer to the spacing between items on a subtest. The item gradients for the WISC-IV subtest range from good to excellent across the age range of the test. In fact, the only item gradient violation occurred at age 6. Thus, the spacing between items on the WISC-IV subtests is generally small enough to allow for reliable discrimination between individuals on the latent trait measured by the subtest.

Structural Validity

As stated previously, the structural validity of the WISC-IV is supported by the factor-analytic studies described in the *WISC-IV Technical and Interpretive Manual* (The Psychological Corporation, 2003; see Figure 1.3 in this chapter). However, the manual did not provide information about the stability or invariance of this factor structure across age. In addition, because The Psychological Corporation did not provide factor loadings and factor correlations for the confirmatory factor analyses presented in the manual, additional analyses were needed to clarify the nature of the cognitive constructs measured by the test.

Keith and colleagues (2006) investigated whether the WISC-IV measured the same constructs across its 11-year age span, as well as the nature of those constructs using the WISC-IV standardization data. Results of their analyses indicated that the WISC-IV measures the same constructs across the age range of the test. These constructs are represented by the large ovals in Figure 1.3. However, according to Keith and colleagues, the factor structure of the WISC-IV (depicted in Figure 1.3) is not a good explanation of the constructs measured by the test. Rather, based on a comparison of theory-derived alternative models with the one depicted in Figure 1.3, Keith and colleagues found that a factor structure more consistent with CHC theory provided a better fit to the WISC-IV standardization data. See Appendix A for detailed definitions of the CHC abilities.

According to Keith and colleagues (2006), the WISC-IV measures Crystallized Ability (*Gc*), Visual Processing (*Gv*), Fluid Reasoning (*Gf*), Short-Term Memory (*Gsm*), and Processing Speed (*Gs*). These findings are depicted in Figure 1.4 and are consistent with the results of a recently conducted content-validity study of the WISC-IV, based on CHC theory, that used an expert consensus format (Alfonso, Flanagan, & Radwan, 2005). Rapid Reference 1.3 summarizes the results of the studies conducted by Keith and colleagues (2006) and Alfonso and colleagues (2005). Although The Psychological Corporation identified four factors to describe the constructs underlying the WISC-IV, Rapid Reference 1.3 shows that Keith and colleagues and Alfonso and colleagues found five. In addition, the results of these latter two studies were consistent, with the exception of the CHC abilities presumed to underlie the Arithmetic subtest. Keith and colleagues described this test as *Gf* and *Gsm,* and Alfonso and colleagues classified this test as Quantitative Knowledge *Gq* and *Gf*. Interestingly, following the publication of the WISC-IV and its *WISC-IV Technical and Interpretive Manual* (The Psychological Corporation, 2003), The Psychological Corporation classified all of the WISC-IV subtests according to CHC theory on its website. These classifications are located in Rapid Reference 1.4, which shows that the classifications offered by The Psychological Corporation are similar to those provided in Rapid Reference 1.3, with

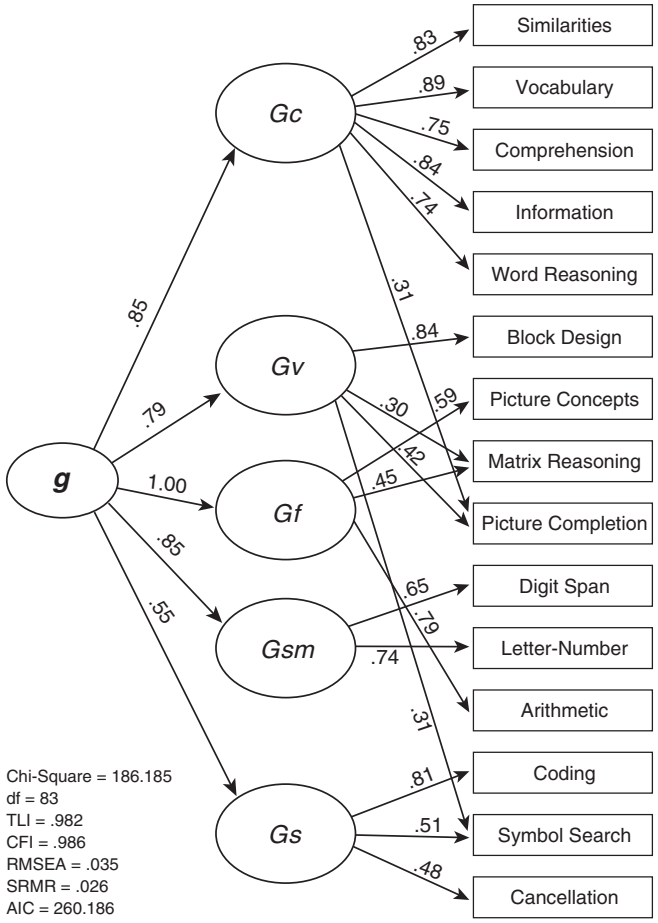

Figure 1.4 CHC Structure of the WISC-IV

Source: Keith et al. (2006). Printed with permission from authors.

Note: df = degrees of freedom; TLI = Tucker Lewis Index; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; AIC = Akaike Information Criterion.

only a few exceptions. That is, The Psychological Corporation classified Similarities and Word Reasoning as primarily measures of *Gf* and Arithmetic as primarily a measure of *Gq* and *Gsm*.

Although the factor analyses conducted by The Psychological Corporation and Keith and colleagues (2006) differ, it is important to understand that there is

no one *right* method of factor analysis. Indeed, the factor analyses, particularly the exploratory factor analyses, summarized in the *WISC-IV Technical and Interpretive Manual* provide strong support for the WISC-IV four-factor structure; while the confirmatory factor analyses conducted by Keith and colleagues provide strong support for a five-factor structure. Noteworthy is the fact that the five-factor CHC model is more in line with contemporary psychometric theory and research than is the four-factor structure that was used to develop the four WISC-IV Indexes. Nevertheless, our interpretive system permits examiners to interpret the WISC-IV according to either four or five factors. The latter option is made possible by the inclusion of clinical clusters and supplementary norms in our interpretive system (Chapter 4, Step 7).

Briefly, based on the results of independent factor analyses, expert consensus content-validity findings, the CHC classifications of the WISC-IV subtests offered by The Psychological Corporation (see Rapid References 1.3 and 1.4), and our own clinical judgment, we developed eight new clinical clusters:

1. Fluid Reasoning (*Gf*)
2. Visual Processing (*Gv*)
3. Nonverbal Fluid Reasoning (*Gf*-nonverbal)
4. Verbal Fluid Reasoning (*Gf*-verbal)
5. Lexical Knowledge (*Gc*-VL)
6. General Information (*Gc*-KO)
7. Long-Term Memory (*Gc*-LTM)
8. Short-Term Memory (*Gsm*-MW)

These clinical clusters may be used in what we call *Planned Clinical Comparisons* to gain information about a child's cognitive capabilities beyond the four Indexes and FSIQ, as well as to generate hypotheses about cognitive performance to be verified through other data sources. Figure 1.5 provides a *selective testing table* that may be used by the examiner to identify the different combinations of WISC-IV subtests that compose the four Indexes, FSIQ, and new clinical clusters. Use of the clinical clusters in Planned Clinical Comparisons is discussed as an optional interpretive step in Chapter 4.

Relationship to Other Wechsler Scales

In addition to factor analysis and content-validity research, the validity of the WISC-IV is supported by correlations with scores on other comprehensive measures of cognitive ability in normal and special group samples (Wechsler, 2003a; Kaufman & Kaufman, 2004, Table 8.17; Launey, Caroll, & Van Horn, 2007).

Subtest	Full Scale IQ (FSIQ)	Verbal Comprehension Index (VCI)	Perceptual Reasoning Index (PRI)	Working Memory Index (WMI)	Processing Speed Index (PSI)	Fluid Reasoning (*Gf*) Cluster	Visual Processing (*Gv*) Cluster	Verbal Fluid Reasoning (Gf-verbal) Cluster	Nonverbal Fluid Reasoning (*Gf*-nonverbal) Cluster	Lexical Knowledge(*Gc*-VL) Cluster	General Information (*Gc*-K0) Cluster	Long-Term Memory (*Gc*-LTM) Cluster	Short-Term Memory (*Gsm*-WM) Cluster[a]
	New Clinical Clusters												
1. Block Design	•		•				•						
2. Similarities	•	•						•					
3. Digit Span	•			•									•
4. Picture Concepts	•		•			•			•				
5. Coding	•				•								
6. Vocabulary	•	•								•		•	
7. Letter-Number Sequencing	•			•									•
8. Matrix Reasoning	•		•			•			•				
9. Comprehension	•	•									•		
10. Symbol Search	•				•								
11. Picture Completion							•						
12. Cancellation													
13. Information											•	•	
14. Arithmetic						•							
15. Word Reasoning								•		•			

[a]The Short-Term Memory (*Gsm*-WM) Cluster is identical to the WISC-IV Working Memory Index.

Figure 1.5 Selective Testing Table

Correlations with Full Scale IQ

Rapid Reference 1.9 shows the correlations between the WISC-IV FSIQ and the WISC-III FSIQ (.89) as well as the FSIQs from other Wechsler scales that are composed of both verbal and nonverbal subtests (i.e., WPPSI-III, WAIS-III, WAIS-IV, and WASI). Not surprisingly, the WISC-IV FSIQ is highly correlated with the FSIQs of these other Wechsler scales. Data are also included for the new *Wechsler Nonverbal Scale of Ability* (WNV; Wechsler & Naglieri. 2006, Table 5.16) for the Full Scale scores yielded by the WNV four-subtest battery and two-subtest battery. These coefficients (.76 and .58, respectively) are lower than the values for the other Wechsler scales, but that is sensible because the WNV is the only Wechsler scale that excludes verbal tasks.

Rapid Reference 1.9

Correlation of Full Scale IQs: WISC-IV and Other Wechsler Scales

	WISC-IV
WISC-III (*N* = 233)	.89
WPPSI-III (*N* = 144)	.89
WAIS-III (*N* = 183)	.89
WAIS-IV (*N* = 157)	.90
WASI (*N* = 254)	.86
WNV (*N* = 102)	
Full Scale Score (four subtests)	.76
Full Scale Score (two subtests)	.58

Note: All values are corrected for the variability of the standardization sample. Coefficients for the WPPSI-III, WISC-III, WAIS-III, and WASI are from *WISC-IV Technical and Interpretive Manual* (The Psychological Corporation, 2003, Tables 5.8, 5.10, 5.12, and 5.14). Coefficients for the Wechsler Nonverbal Scale of Ability are from the WNV Technical and Interpretive Manual (Wechsler & Naglieri, 2006, Table 5.16). Coefficients for the WAIS-IV are from the *WAIS-IV Technical and Interpretive Manual* (Pearson/PsychCorp, 2008, Table 5.9).

The correlation of .91 between WISC-IV and WAIS-IV Full Scale IQs is large, and is consistent with the fact that both new Wechsler scales compute Full Scale IQ the same way—namely, based on the 10 subtests that compose the four indexes. Even more impressive than the .91 coefficient are the values of .89 between WISC-IV FSIQ and previous FSIQs. These substantial coefficients suggest a continuity of the construct measured by the Full Scale, which is notable because the WISC-IV Full Scale is dramatically different from its predecessors. It shares only five subtests in common with the Full Scales of earlier versions of the WISC and other Wechsler scales.

The substantial correlation between WISC-IV FSIQ and previous FSIQs suggests a continuity of the construct measured by the Full Scale. Nonetheless, it is notable that the WISC-IV Full Scale is dramatically different from its predecessors. It shares only five subtests in common with the Full Scales of earlier versions of the WISC.

The two WMI and two PSI subtests, all of which have relatively "low *g*–loadings" (.40s to .60s; Tables C.1 and C.2), constitute 40% of the Full Scale. Of these four working-memory and processing speed subtests, only Coding was on previous WISC Full Scales. Excluded from the WISC-IV FSIQ are subtests that have "high *g*-loadings," like Arithmetic and Information (mid-.70s to low .80s). This different FSIQ better represents the constructs that compose the WISC-IV; however, despite the .89 coefficients with WISC-III FSIQ, it is possible that research findings with previous WISCs do not completely generalize to the WISC-IV (Kaufman, Flanagan, Alfonso, & Mascolo, 2006). For example, FSIQ differences between Whites and African Americans who were matched on SES and other background variables were found to be smaller on the WISC-IV (8.8 points) than on WISC-III (11.0 points) (Prifitera & Saklofske, 1998; Prifitera, Saklofske, & Weiss, 2005). This is a positive finding, discussed later in this chapter. But, as Kaufman and colleagues (2006) note, "clinicians and researchers need to be aware that with the clear-cut improvements in the structure of the major scales that comprise the WISC-IV comes the side effect of bringing into question the generalizability to the WISC-IV of IQ-based research results—even those that are time tested over the past 60 years with the Wechsler-Bellevue II, WISC, WISC-R, and WISC-III" (p. 281).

Convergent–Discriminant Validity Coefficients

The WISC-IV also shows good to excellent convergent–discriminant validity evidence. Rapid Reference 1.10 presents coefficients for the WISC-IV VCI and PRI with verbal and nonverbal scales on other Wechsler batteries. These scales include five Wechsler tests that yield scores on *both* verbal and nonverbal subtests: WPPSI-III, WISC-III, WAIS-III, WAIS-IV, and WASI. For these Wechsler scales, VCI has an average correlation of .84 with other measures of verbal ability, compared to a mean of .60 with measures of perceptual abilities. Similarly, Rapid Reference 1.10 shows that the PRI has an average correlation of .76 with other measures of visual-perceptual ability, compared to a mean of .61 with measures of verbal abilities.

In addition to the more traditional verbal-nonverbal Wechsler batteries, Rapid Reference 1.10 also presents convergent–discriminant coefficients for the WNV, which would be predicted to correlate substantially higher with WISC-IV PRI than VCI. Again, these data support the validity of the WISC-IV Indexes, as the two WNV Full Scale scores (based only on nonverbal subtests) correlated higher with PRI than VCI. The magnitude of the coefficients with the two WISC-IV Indexes is lower than the values for the other Wechsler scales (about .40 with VCI

Convergent–Discriminant Validity of the WISC-IV Verbal Comprehension Index (VCI) and Perceptual Reasoning Index(PRI)

	WISC-IV	
	VCI	**PRI**
WPPSI-III (n = 182, ages 6–7)		
Verbal IQ	**.83**	.63
Performance IQ	.65	**.79**
General Language Composite (GLC)	**.68**	.53
WISC-III (n = 244, ages 6–16)		
Verbal Comprehension Index (VCI)	**.88**	.59
Perceptual Organization Index (POI)	.62	**.72**
Verbal IQ	**.87**	.64
Performance IQ	.61	**.74**
WAIS-III (n = 198, age 16)		
Verbal Comprehension Index (VCI)	**.86**	.64
Perceptual Organization Index (POI)	.57	**.76**
Verbal IQ	**.86**	.69
Performance IQ	.61	**.76**
WASI-4 subtests (n = 260, ages 6–16)		
Verbal IQ	**.85**	.61
Performance IQ	.60	**.78**
WAIS-IV (n = 157, age 16)		
Verbal Comprehension Index	**.88**	.54
Perceptual Reasoning Index	.52	**.77**
WNV (n = 102, ages 7–16)		
Full Scale Score (four subtests)	.47	**.66**
Full Scale Score (two subtests)	.31	**.57**

Source: Convergent and divergent values for the Wechsler Nonverbal Scale of Ability (WNV) are from the *WNV Technical and Interpretive Manual* (Wechsler & Naglieri, 2006, Table 5.16). Convergent values for the WPPSI-III, WISC-III, WAIS-III, and WASI are from the *WISC-IV Technical and Interpretive Manual* (The Psychological Corporation, 2003, Tables 5.8, 5.10, 5.12, and 5.14). The divergent values (VCI with visual-perceptual ability, PRI with verbal ability) were provided by The Psychological Corporation. Convergent values for the *WAIS-IV* are from the *WAIS-IV Technical and Interpretive Manual* (Pearson/PsychCorp, 2008, Table 5.9). Divergent values were provided by The Psychological Corporation. Analysis results from the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV).

Note: Correlations of WISC-IV VCI and PRI with other measures of Wechsler's Verbal and Visual-Perceptual ability (average-corrected correlations across two testing orders), respectively, are printed in bold. Coefficients in bold denote convergent validity of WISC-IV VCI and PRI. All values are corrected for the variability of the standardization sample.

and about .60 with PRI) because the content of the WNV is so different from the content of the other Wechsler scales. The reason that the WNV correlated lower with VCI than the other Wechsler scales did is obvious, given the exclusion of verbal tasks from the WNV. The reasons for the lower correlations with PRI are less obvious, but undoubtedly relate to the fact that the WNV includes the novel nonverbal subtests of Recognition for ages 4 through 7 (memory for abstract figures) and Spatial Span for ages 8 though 21 (ability to reproduce the sequence of blocks tapped by the examiner). In addition, the WNV includes familiar Wechsler subtests that were not included in the WISC-IV—Object Assembly (ages 4–7) and Picture Arrangement (8–21). The net result is that the WNV correlates substantially enough with the WISC-IV and a variety of other cognitive batteries (Wechsler & Naglieri, 2006, Chapter 5) to support its construct validity, but it also has its own degree of uniqueness. As Kaufman (2006) notes in the Foreword to the WNV, "its built-in brief form makes it a flexible instrument for a variety of testing purposes within the 4- to 21-year age range, and its clever administrative aids make it a user-friendly instrument whenever it is desirable or essential to test a person's general intelligence nonverbally" (p. iv).

Relationship to WIAT-II

The validity of the WISC-IV was investigated further through an examination of its relationship to academic achievement. Rapid Reference 1.11 includes the

Rapid Reference 1.11

WISC-IV Indexes and Full Scale IQ: Correlations with WIAT-II Achievement Composites

WIAT-II Composite	VCI	PRI	WMI	PSI	FSIQ
Reading	.74	.63	.66	.50	.78
Math	.68	.67	.64	.53	.78
Written Language	.67	.61	.64	.55	.76
Oral Language	.75	.63	.57	.49	.75
Total Achievement	.80	.71	.71	.58	.87

Note: All values are corrected for the variability of the standardization sample. Coefficients are from *WISC-IV Technical and Interpretive Manual* (The Psychological Corporation, 2003, Table 5.15). Sample sizes range from 538 to 548.

correlations between the WISC-IV Indexes and FSIQ with the WIAT-II Achievement Composites. This Rapid Reference shows that the correlations between the FSIQ and WIAT-II Composites ranged from .75 (Oral Language) to .78 (Reading and Math), indicating that the WISC-IV FSIQ explains 56 to 60% of the variance in these achievement domains. Regardless of whether or not there is significant variability in factor scores, the FSIQ is a robust predictor of academic achievement in normal and clinical samples (Watkins, Glutting, & Lee, 2007; Glutting, Watkins, Konold, & McDermott, 2006; Weiss et al., 2006). The correlation between the FSIQ and WIAT-II Total Achievement Score is .87 (76% of variance explained), which is about as high as the correlation between the WISC-IV FSIQ and the FSIQs of other Wechsler scales (i.e., .89; see Rapid Reference 1.9). These correlations are among the highest ever reported between global IQ and achievement. According to Kenny (1979), "even highly developed causal models do not explain behavior very well. A good rule of thumb is that one is fooling oneself if more than 50% of the variance is predicted" (p. 9). It is likely that either overlapping content or standard deviations greater than 15 or some combination thereof led to spuriously high correlations.

Rapid Reference 1.12 summarizes the WISC-IV subtests that are the best and worst predictors of WIAT-II Achievement Composites. In general, Arithmetic, Vocabulary, and Information are the best predictors of the WIAT-II Composites; and Picture Concepts along with Coding and Cancellation (i.e., the Processing Speed subtests) are the worst predictors of these same composites.

In addition to the validity evidence summarized previously, the *WISC-IV Technical and Interpretive Manual* provides a number of special group studies to investigate the diagnostic utility of the instrument. These studies are discussed in detail in Chapter 6. Overall, the WISC-IV is a reliable and valid measure of a select number of cognitive abilities (viz., Verbal Comprehension [*Gc*], Perceptual Reasoning [*Gf*, *Gv*], Working Memory [*Gsm*], and Processing Speed [*Gs*]).

Ehnic Differences on the WISC-IV

Historically, Whites have scored about one standard deviation higher than African Americans on Wechsler's scales, with Full Scale IQs differing by 15.9 points on the WISC-R (Kaufman & Doppelt, 1976) and 14.9 points on the WISC-III (Prifitera, Weiss, & Saklofske, 1998, Table 1.1). For the WISC-IV, Full Scale IQ differences are reduced to ¾ SD, or 11.5 points; when SES and other background variables are controlled, that difference reduces to 8.8 points (Prifitera, Weiss, Saklofske, & Rolfhus, 2005, Tables 1.3 & 1.4). As shown in Rapid Reference 1.13, this overall difference in Full Scale IQ is merely an average of wide variations in African American–White

Rapid Reference 1.12

WISC-IV Subtests: The Best and Worst Predictors of WIAT-II Achievement Composites

Reading	Math	Written Language	Oral Language	Total Achievement
BEST				
Vocabulary (.72)	Arithmetic (.74)	Arithmetic (.67)	Vocabulary (.73)	Vocabulary (.76)
Information (.68)	Information (.67)	Vocabulary (.64)	Information (.69)	Information (.75)
Arithmetic (.68)	Vocabulary (.64)	Information (.62)	Similarities (.67)	Arithmetic (.75)
WORST				
Picture Concepts (.42)	Picture Concepts (.42)	Picture Concepts (.41)	Picture Concepts (.41)	Picture Concepts (.47)
Coding (.40)	Coding (.42)	Picture Completion (.40)	Coding (.38)	Coding (.45)
Cancellation (.14)	Cancellation (.11)	Cancellation (.14)	Cancellation (.15)	Cancellation (.15)

Note: Correlations of WISC-IV scaled scores with WIAT-II achievement composite standard scores are repeated in parentheses. All values are corrected for the variability of the standardization sample. Coefficients are from *WISC-IV Technical and Interpretive Manual* (The Psychological Corporation, 2003, Table 5.15). Sample sizes range from 531 to 548, except for the Arithmetic subtest (n = 301).

Rapid Reference 1.13

Difference in Mean Standard Scores for Whites versus African Americans on the WISC-IV Full Scale IQ and Indexes, Controlling for SES and Other Background Variables

	Difference in Mean Scores (White Minus African American)	
IQ or Index	**Ages 6–11**	**Ages 12–16**
Full Scale IQ	6.0	11.8
Verbal Comprehension (VCI)	5.6	12.2
Perceptual Reasoning (PRI)	6.8	10.5
Working Memory (WMI)	1.9	5.9
Processing Speed (PSI)	3.5	5.6

Source: Table 1.6 in Prifitera, A., Weiss, L. G., Saklofske, D. H., & Rolfhus, E. (2005). The WISC-IV in the clinical assessment context. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV: Clinical use and interpretation.* San Diego, CA: Elsevier Science.

differences based on the specific Index score and age group. With numerous variables controlled (parental education, gender, region, and number of parents living in the household), differences are smallest for the WMI and PSI (3-4 points) and for pre-adolescents than adolescents (6 versus 12 points).

Ethnic differences for Whites versus Hispanics averaged about 10 Full Scale IQ points (2/3 SD), in favor of Whites, on earlier versions of Wechsler's children's scales—11.2 points on the WISC-R (Mercer, 1979), 9.4 points on the WISC-III (Prifitera et al., 1998, Table 1.1)—and that same difference (9.9 points) characterizes the WISC-IV (Prifitera et al., 2005, Table 1.3). Traditionally, the difference in favor of Whites has been much larger on the Verbal than Performance Scale, as would be expected in view of the cultural and, especially, linguistic aspects of the Verbal subtests. Verbal IQ differences were 12 to 14 points on the WISC-R and WISC-III, compared to Performance IQ differences of 5 to 6 points (Mercer, 1979; Prifitera et al., 1998, Table 1.1). On the WISC-IV, that predictable verbal-nonverbal distinction was reflected in the VCI (11.4 points) versus the PRI (7.1 points) (Prifitera et al., 2005, Table 1.3). When controlling for SES and other pertinent variables, all WISC-IV differences are greatly reduced (FSIQ = 4.8; VCI = 6.0; PRI = 2.6) (Prifitera et al., 2005, Table 1.5).

Rapid Reference 1.14

Difference in Mean Standard Scores for Whites versus Hispanics on the WISC-IV Full Scale IQ and Indexes, Controlling for SES and Other Background Variables

	Difference in Mean Scores (White Minus Hispanic)	
IQ or Index	**Ages 6–11**	**Ages 12–16**
Full Scale IQ	1.3	8.0
Verbal Comprehension (VCI)	3.7	8.5
Perceptual Reasoning (PRI)	–0.2	3.9
Working Memory (WMI)	2.4	5.5
Processing Speed (PSI)	–1.5	3.4

Source: Table 1.7 in Prifitera, A., Weiss, L. G., Saklofske, D. H., & Rolfhus, E. (2005). The WISC-IV in the clinical assessment context. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV: Clinical use and interpretation.* San Diego, CA: Elsevier Science.

Furthermore, these differences are merely verbal versus nonverbal; some differences favor Hispanics and age plays a factor as well, as it did for African American–White differences. Rapid Reference 1.14 presents FSIQ and Index differences, by age, for Hispanics versus Whites on the WISC-IV for groups matched on parental education, gender, region, and number of parents living in the household. At ages 6 to 11, the two ethnic groups differed by 1 point on FSIQ and performed about equally well on all scales; differences ranged from a 3.7 advantage for Whites on VCI to a 1.5 discrepancy in favor of Hispanics on PSI. At ages 12 to 16, results were quite different. Whites scored higher on the FSIQ (8 points) and on all scales, with the differences highest on VCI (8.5 points) and lowest on PSI and PRI (3–4 points).

Analogous age-related findings characterized White–Hispanic differences on the Kaufman Brief Intelligence Test (K-BIT; Kaufman & Kaufman, 1990) for global, verbal (Vocabulary), and nonverbal (Matrices) scores adjusted for SES (Kaufman & Wang, 1992). Global IQ differences on the K-BIT favored Hispanics by 1 point at ages 4 to 7 years, whereas Whites scored higher at ages 8 to 12 (2.6 points) and 13 to 19 (6.0 points). Similarly, Vocabulary differences increased with increasing age group (from 3.5 points at ages 4–12 to 6.5 points at 13–19) and the

Matrices difference went from a 5-point advantage for Hispanics at ages 4 to 7 to a 4.4-point edge for Whites at ages 13 to 19.

None of these ethnic differences for Whites relative to African Americans and Hispanics have intuitive explanations; even the age-related findings defy simple understanding. The most insightful discussion that we have read on these complex issues is a chapter written by Weiss and colleagues (2006). They reviewed the often-ignored body of literature on the roles of mental health status, physical health status, education, income, home environment, cognitive stimulation, and individual differences on intellectual development, and they discuss how these variables have a differential impact on different ethnic groups. Weiss and colleagues (2006) emphasize that "children grow up with differing levels of opportunity for cognitive growth and development" (p. 18), that "race/ethnicity are likely to be proxy variables for a set of active mechanisms that have only been partially identified" (p. 32), and that "cognitive growth is malleable, within limits, based on environmental opportunities for cognitive development" (p. 51).

Weiss and colleagues (2006) also conducted an innovative set of multiple-regression analyses using WISC-IV data to demonstrate the mediating effect of SES variables on IQ differences often attributed to race/ethnicity. They showed, for example, that parent education alone accounted for 18.8% of the variance in FSIQ between African American and White samples, much higher than the 4.7% for race alone. Further, controlling for parent education and household income reduced the 4.7% to 1.6%. In their analyses of Hispanic–White differences, the percent of variance due to parental education was 17.5 and the ethnic status percent was 1.4; controlling for SES explained nearly all of the variance attributed to ethnic status. We recommend reading this exceptional chapter in its entirety to fully grasp the role of contextual factors in shaping the IQs earned by individuals from diverse ethnic groups and to be able to give 2-point responses to any questions you may be asked about SES, test bias, or ethnic differences on intelligence tests.

Other Quantitative and Qualitative Characteristics of the WISC-IV

Appendix D on the CD-ROM provides a quick reference to key quantitative and qualitative features of the WISC-IV subtests that may aid in interpretation. Several quantitative characteristics are *evaluated* in Table D.1 according to commonly accepted criteria, including internal consistency and test–retest reliabilities, *g*-loadings, subtest floors and ceilings, and item gradients. Table D.1 also includes important qualitative characteristics of the WISC-IV subtests. Specifically, each subtest is classified according to degree of cultural loading and linguistic demand.

Also, a list of the most probable factors that influence subtest performance is provided for each subtest. Table D.2 of this appendix provides definitions of the quantitative and qualitative characteristics included in Table D.1 along with an explanation of the criteria used to (a) evaluate the quantitative characteristics and (b) classify the WISC-IV subtests according to select qualitative characteristics. Finally, Table D.2 provides a brief description of the interpretive relevance of each characteristic included in Table D.1. The information included in Appendix D may be used to assist in the generation of hypotheses about a child's unique profile of cognitive capabilities.

CONCLUSION

The contributions to the science of intellectual assessment made by David Wechsler through his intelligence scales are many and substantial, if not landmark. Although he is not recognized as an important theoretician, this neither detracts from his accomplishments nor diminishes his innovations in applied psychometrics. Wechsler was a well-known clinician and, as such, he intentionally placed significant importance on developing tasks that had practical, clinical value, and not merely theoretical value. Thus, the driving force behind the development of the Wechsler scales was no doubt based more on practical considerations rather than theoretical ones. Zachary (1990) stated, "When David Wechsler published the original Wechsler-Bellevue scales in 1939, he said relatively little about the theoretical underpinnings of his new instrument; rather, he followed a pragmatic approach. He selected a set of tasks that were easy to administer and score." (p. 276). Detterman (1985) also attributed much of the popularity of the Wechsler family of tests to their "ease of administration fostered by an organization of subtests that are brief and have long clinical histories" (p. 1715). For better or worse, Wechsler's primary motivation for constructing his tests was to create an efficient, easy-to-use tool for clinical purposes; operationalizing them according to a specific theory of intelligence was not of paramount importance.

Despite these accomplishments and accolades, under the critical eye of subsequent advancements in the field, the failure of the Wechsler scales to keep abreast of contemporary intelligence research cannot be ignored. It is clear that meaningful use and interpretation of the Wechsler scales requires the adoption of a fourth-wave approach in which contemporary theory, research, and measurement principles are integrated.

We believe that clinical judgment and experience alone are insufficient stanchions upon which defensible interpretations can be built. Application of contemporary theory and research to intelligence test use and interpretation is needed.

The interpretive approach offered in this book is an efficient, theoretically and statistically defensible method for assessing and interpreting the array of cognitive abilities underlying the WISC-IV. The subsequent chapters of this book demonstrate how the principles and procedures of both Kaufman's and Flanagan's interpretive methods have been integrated to advance the science of measuring and interpreting cognitive abilities using the WISC-IV.

COMPREHENSIVE REFERENCES ON THE WISC-IV

The *WISC-IV Technical and Interpretive Manual* (The Psychological Corporation, 2003) provides important information about the development of the test and includes descriptions of the subtests and scales, as well as detailed information on standardization, reliability, and validity.

Also see the following resources:

- Prifitera, A., Saklofske, D. H., & Weiss, L. G. (Eds.). (2008). *WISC-IV clinical use and interpretation, Second Edition*. San Diego, CA: Elsevier Science.
- Sattler, J. M. (2008). *Assessment of children: Cognitive foundations* (5th ed.). San Diego, CA: Jerome M. Sattler.
- Weiss, L. G., Saklofske, D. H., Prifitera, A., & Holdnack, J. A. (Eds.). (2006). *WISC-IV advanced clinical interpretation*. San Diego, CA: Elsevier Science.

TEST YOURSELF

1. **Picture Arrangement, Object Assembly, and Mazes were deleted from the WISC-IV battery for which one of the following reasons?**
   (a) because they are most valid for preschool children
   (b) to deemphasize the timed nature of the battery
   (c) because surveys regarding WISC-IV development revealed that children did not like these tests
   (d) because these tests were deemed unfair to language-impaired children
2. **The Block Design subtest is primarily a measure of which of the following CHC abilities?**
   (a) Visual Processing (*Gv*)
   (b) Fluid Reasoning (*Gf*)
   (c) Working Memory (*Gsm*-MW)
   (d) Processing Speed (*Gs*)

3. **The average reliability of the WISC-IV core battery subtests can be best described as**
   (a) high.
   (b) low.
   (c) medium.
   (d) unacceptable.
4. **Which of the following WISC-IV indexes is the best predictor of written-language achievement?**
   (a) VCI
   (b) PRI
   (c) WMI
   (d) PSI
5. **The WISC-IV represents the most substantial revision of the Wechsler scales to date.**
   True or False?
6. **Cohen's significant contributions that largely defined the third wave of test interpretation included which of the following?**
   (a) empirical support for the FSIQ based on analysis of shared variance between subtests
   (b) development of the three-factor solution for interpretation of the Wechsler scales
   (c) revelation of limited subtest specificity, questioning individual subtest interpretaion
   (d) all of the above
7. **Kaufman's and Flanagan's intraindividual (ipsative) analysis method has improved upon traditional ipsative methods in several ways. One major difference between their approach and traditional approaches is that they recommend using composites or clusters, rather than subtests, in intraindividual analysis.**
   True or False?
8. **When SES is controlled for, the *smallest* differences between Whites and African Americans are found in *processing speed* and**
   (a) verbal comprehension.
   (b) global intelligence.
   (c) working memory.
   (d) perceptual reasoning.

Answers:

1. b; 2. a; 3. c; 4. a; 5. True; 6. d; 7. True; 8. c