

One

INTRODUCTION AND OVERVIEW

INTRODUCTION

The field of assessment, particularly intellectual assessment, has grown tremendously over the past couple of decades. New tests of cognitive abilities are being developed, and older tests of intelligence are being revised to meet the needs of the professionals utilizing them. There are several good sources for reviewing major measures of cognitive ability (e.g., Flanagan & Harrison, 2012; Naglieri & Goldstein, 2009; Sattler, 2008); however, the new and revised measures multiply rapidly, and it is often difficult to keep track of new instruments, let alone know how to administer, score, and interpret them. One of the goals of this book is to provide an easy reference source for those who wish to learn essentials of the *Wechsler Adult Intelligence Scale—Fourth Edition* (WAIS-IV) in a direct, no-nonsense, systematic manner.

Essentials of WAIS-IV Assessment was developed with an easy-to-read format in mind. The topics covered in the book emphasize administration, scoring, interpretation, and application of the WAIS-IV. Each chapter includes several “Rapid Reference,” “Caution,” and “Don’t Forget” boxes that highlight important points for easy reference. At the end of each chapter, questions are provided to help you solidify what you have read. The information provided in this book will help you to understand, in depth, the latest of the measures in the Wechsler family and will help you become a competent WAIS-IV examiner and clinician.

HISTORY AND DEVELOPMENT

The first assessment instrument developed by David Wechsler came on the scene in 1939. However, the history of intelligence testing began several decades before that, in the late 19th century, and is largely an account of the measurement of

2 ESSENTIALS OF WAIS®-IV ASSESSMENT

the intelligence of children or retarded adults. Sir Francis Galton (1869, 1883) studied adults and was interested in giftedness when he developed what is often considered the first comprehensive individual test of intelligence, composed of sensorimotor tasks (Kaufman, 2000b). But despite Galton's role as the father of the testing movement (Shouksmith, 1970), he did not succeed in constructing a true intelligence test. His measures of simple reaction time, strength of squeeze, or keenness of sight proved to assess sensory and motor abilities, skills that relate poorly to mental ability and that are far removed from the type of tasks that constitute contemporary intelligence tests.

BINET-SIMON SCALES

Alfred Binet and his colleagues (Binet & Henri, 1895; Binet & Simon, 1905, 1908) developed the tasks that survive to the present day in most tests of intelligence for children and adults. Binet (1890a, 1890b) mainly studied children; beginning with systematic developmental observations of his two young daughters, Madeleine and Alice, he concluded that simple tasks such as those used by Galton did not discriminate between children and adults. In 1904, the minister of public instruction in Paris appointed Binet to a committee to find a way to distinguish normal from retarded children. Fifteen years of qualitative and quantitative investigation of individual differences in children—along with considerable theorizing about mental organization and the development of a specific set of complex, high-level tests to investigate these differences—preceded the “sudden” emergence of the landmark 1905 Binet-Simon intelligence scale (Murphy, 1968).

The 1908 scale was the first to include age levels, spanning the range from 3 to 13. This important modification stemmed from Binet and Simon's unexpected discovery that their 1905 scale was useful for much more than classifying a child at one of the three levels of retardation: moron, imbecile, idiot (Matarazzo, 1972). Assessment of older adolescents and adults, however, was not built into the Binet-Simon system until the 1911 revision. That scale was extended to age 15 and included five ungraded adult tests (Kite, 1916). This extension was not conducted with the rigor that characterized the construction of tests for children, and the primary applications of the scale were for use with school-age children (Binet, 1911).

Measuring the intelligence of adults, except those known to be mentally retarded, was almost an afterthought. But Binet recognized the increased applicability of the Binet-Simon tests for various child assessment purposes just before his untimely death in 1911, when he “began to foresee numerous uses for his

method in child development, in education, in medicine, and in longitudinal studies predicting different occupational histories for children of different intellectual potential” (Matarazzo, 1972, p. 42).

TERMAN’S STANFORD-BINET

Lewis Terman was one of several people in the United States who translated and adapted the Binet-Simon scale for use in the United States, publishing a “tentative” revision (Terman & Childs, 1912) four years before releasing his painstakingly developed and carefully standardized Stanford Revision and Extension of the Binet-Simon Intelligence Scale (Terman, 1916). This landmark test, soon known simply as the Stanford-Binet, squashed competing tests developed earlier by Goddard, Kuhlmann, Wallin, and Yerkes. Terman’s success was undoubtedly due in part to heeding the advice of practitioners whose demand “for more and more accurate diagnoses . . . raised the whole question of the accurate placing of tests in the scale and the accurate evaluation of the responses made by the child” (Pintner & Paterson, 1925, p. 11).

Terman (1916) saw intelligence tests as useful primarily for the detection of mental deficiency or superiority in children and for the identification of “feble-mindedness” in adults. He cited numerous studies of delinquent adolescents and adult criminals, all of which pointed to the high percentage of mentally deficient juvenile delinquents, prisoners, or prostitutes, and concluded that “there is no investigator who denies the fearful role played by mental deficiency in the production of vice, crime, and delinquency” (p. 9). Terman also saw the potential for using intelligence tests with adults for determining “vocational fitness,” but, again, he emphasized employing “a psychologist . . . to weed out the unfit” or to “determine the minimum ‘intelligence quotient’ necessary for success in each leading occupation” (p. 17).

Perhaps because of this emphasis on the assessment of children or concern with the lower end of the intelligence distribution, Terman (1916) did not use a rigorous methodology for constructing his adult-level tasks. Tests below the 14-year level were administered to a fairly representative sample of about 1,000 children and early adolescents. To extend the scale above that level, data were obtained from 30 businessmen, 50 high school students, 150 adolescent delinquents, and 150 migrating unemployed men. Based on a frequency distribution of the mental ages of a mere 62 adults (the 30 businessmen and 32 of the high school students above age 16), Terman partitioned the graph into the Mental Age (MA) categories: 13 to 15 (inferior adults), 15 to 17 (average adults), and above 17 (superior adults).

WORLD WAR I TESTS

The field of adult assessment grew rapidly with the onset of World War I, particularly after U.S. entry into the war in 1917 (Anastasi & Urbina, 1997; Vane & Motta, 1984). Psychologists saw with increasing clarity the applications of intelligence tests for selecting officers and placing enlisted men in different types of service, apart from their generation-old use for identifying the mentally unfit. Under the leadership of Robert Yerkes and the American Psychological Association, the most innovative psychologists of the day helped translate Binet's tests into a group format. Arthur Otis, Terman's student, was instrumental in leading the creative team that developed the Army Alpha, essentially a group-administered Stanford-Binet, and the Army Beta, a novel group test composed of nonverbal tasks.

Yerkes (1917) opposed Binet's age-scale approach and favored a point-scale methodology, one that advocates selection of tests of specified, important functions rather than a set of tasks that fluctuates greatly with age level and developmental stage. The Army group tests reflect a blend of Yerkes's point-scale approach and Binet's notions of the kind of skills that should be measured when assessing mental ability. The Army Alpha included the Binet-like tests of Directions or Commands, Practical Judgment, Arithmetical Problems, Synonym-Antonym, Dissarranged Sentences, Analogies, and Information. Even the Army Beta had subtests resembling Stanford-Binet tasks: Maze, Cube Analysis, Pictorial Completion, and Geometrical Construction. The Beta also included novel measures, such as Digit Symbol, Number Checking, and X-O Series (Yoakum & Yerkes, 1920). Never before or since have tests been normed and validated on samples so large; 1,726,966 men were tested (Vane & Motta, 1984).

Another intelligence scale was developed during the war, one that became an alternative for those who could not be tested validly by either the Alpha or Beta. This was the Army Performance Scale Examination, composed of tasks that would become the tools of the trade for clinical psychologists, school psychologists, and neuropsychologists into the 21st century: Picture Completion, Picture Arrangement, Digit Symbol, and Manikin and Feature Profile (Object Assembly). Except for Block Design (developed by Kohs in 1923), Army Performance Scale Examination was added to the Army battery "to prove conclusively that a man was weakminded and not merely indifferent or malingering" (Yoakum & Yerkes, 1920, p. 10).

WECHSLER'S CREATIVITY

In the mid-1930s, David Wechsler became a prominent player in the field of assessment by blending his strong clinical skills and statistical training (he studied

under Charles Spearman and Karl Pearson in England) with his extensive experience in testing, gained as a World War I examiner. He assembled a test battery that comprised subtests developed primarily by Binet and World War I psychologists. His Verbal Scale was essentially a Yerkes point-scale adaptation of Stanford-Binet tasks; his Performance Scale, like other similar nonverbal batteries of the 1920s and 1930s (Cornell & Coxe, 1934; Pintner & Paterson, 1925), was a near replica of the tasks and items making up the individually administered Army Performance Scale Examination.

In essence, Wechsler took advantage of tasks developed by others for nonclinical purposes to develop a clinical test battery. He paired verbal tests that were fine-tuned to discriminate among children of different ages with nonverbal tests that were created for adult males who had flunked both the Alpha and Beta exams—nonverbal tests that were intended to distinguish between the unmotivated and the hopelessly deficient. Like Terman, Wechsler had the same access to the available tests as did other psychologists; like Terman and Binet before him, Wechsler succeeded because he was a visionary, a man able to anticipate the needs of practitioners in the field.

While others hoped intelligence tests would be psychometric tools used to subdivide retarded individuals into whatever number of categories was currently in vogue, Wechsler saw the tests as dynamic clinical instruments. While others looked concretely at intelligence tests as predictors of school success or guides to occupational choice, Wechsler looked abstractly at the tests as a mirror to the hidden personality. With the Great War over, many psychologists returned to a focus on IQ testing as a means of childhood assessment; Wechsler (1939), however, developed the first form of the Wechsler-Bellevue Intelligence Scale exclusively for adolescents and adults.

Most psychologists saw little need for nonverbal tests when assessing English-speaking individuals other than illiterates. How could it be worth 2 or 3 minutes to administer a single puzzle or block-design item when 10 or 15 verbal items could be given in the same time? Some test developers (e.g., Cornell & Coxe, 1934) felt that Performance scales might be useful for normal, English-speaking people to provide “more varied situations than are provided by verbal tests” (p. 9) and to “test the hypothesis that there is a group factor underlying general concrete ability, which is of importance in the concept of general intelligence” (p. 10).

Wechsler was less inclined to wait a generation for data to accumulate. He followed his clinical instincts and not only advocated for the administration of a standard battery of nonverbal tests to everyone but also placed the Performance Scale on an equal footing with the more respected Verbal Scale. Both scales

6 ESSENTIALS OF WAIS®-IV ASSESSMENT

would constitute a complete Wechsler-Bellevue battery, and each would contribute equally to the overall intelligence score.

Wechsler also had the courage to challenge the Stanford-Binet monopoly, a boldness not unlike Binet's when the French scientist created his own forum (the journal *L'Année Psychologique*) to challenge the preferred but simplistic Galton sensorimotor approach to intelligence (Kaufman, 2000b). Wechsler met the same type of resistance as Binet, who had had to wait until the French Ministry of Public Instruction "published" his Binet-Simon Scale. When Wechsler's initial efforts to find a publisher for his two-pronged intelligence test failed, he had no cabinet minister to turn to, so he took matters into his own hands. With a small team of colleagues, he standardized Form I of the Wechsler-Bellevue by himself. Realizing that stratification on socioeconomic background was more crucial than obtaining regional representation, he managed to secure a well-stratified sample from Brooklyn, New York (Kaufman, 2009).

The Psychological Corporation agreed to publish Wechsler's battery once it had been standardized, and the rest is history. Although an alternative form of the Wechsler-Bellevue Intelligence Scale (Wechsler, 1946) was no more successful than Terman and Merrill's (1937) ill-fated Form M, a subsequent downward extension of Form II of the Wechsler-Bellevue (to cover the age range 5 to 15 instead of 10 to 59) produced the wildly successful Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949). Although the Wechsler scales did not initially surpass the Stanford-Binet in popularity, instead serving an apprenticeship to the master in the 1940s and 1950s, the WISC and the subsequent revision of the Wechsler-Bellevue, Form I (WAIS; Wechsler, 1955), triumphed in the 1960s. According to Kaufman: "With the increasing stress on the psychoeducational assessment of learning disabilities in the 1960s, and on neuropsychological evaluation in the 1970s, the Verbal-Performance (V-P) IQ discrepancies and subtest profiles yielded by Wechsler's scales were waiting and ready to overtake the one-score Binet" (Kaufman, 1983, p. 107).

Irony runs throughout the history of testing. Galton developed statistics to study relationships between variables—statistics that proved to be forerunners of the coefficient of correlation, later perfected by his friend Pearson (DuBois, 1970). The ultimate downfall of Galton's system of testing can be traced directly to coefficients of correlation, which were too low in some crucial (but, ironically, poorly designed) studies of the relationships among intellectual variables (Sharp, 1898–99; Wissler, 1901). Similarly, Terman succeeded with the Stanford-Binet while the Goddard-Binet (Goddard, 1911), the Herring-Binet

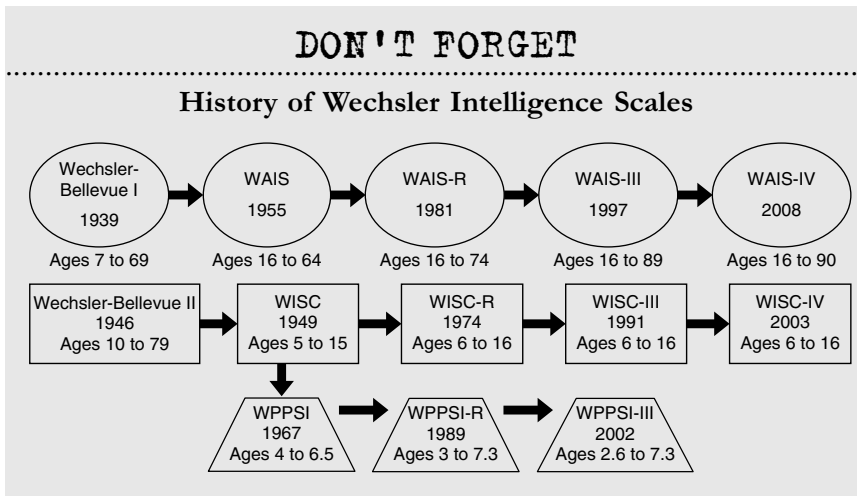
(Herring, 1922), and other Binet-Simon adaptations failed because Terman was sensitive to practitioners' needs. He patiently withheld a final version of his Stanford revision until he was certain that each task was placed appropriately at an age level consistent with the typical functioning of representative samples of U.S. children.

Terman continued his careful test development and standardization techniques with the first revised version of the Stanford-Binet (Terman & Merrill, 1937), but four years after his death in 1956, his legacy was devalued when the next revision of the Stanford-Binet merged Forms L and M *without a standardization* of the newly formed battery (Terman & Merrill, 1960). The following version saw a restandardization of the instrument but without a revision of the placement of tasks at each age level (Terman & Merrill, 1973). Unfortunately for the Binet, the abilities of children and adolescents had changed fairly dramatically in the course of a generation, so the 5-year level of tasks (for example) was now passed by the average 4-year-old.

Terman's methods had been ignored by his successors. The ironic outcome was that Wechsler's approach to assessment triumphed, at least in part because the editions of the Stanford-Binet in the 1960s and 1970s were beset by the same type of flaws as those of Terman's competitors in the 1910s. The fourth edition of the Stanford-Binet (Thorndike, Hagen, & Sattler, 1986) attempted to correct these problems and even adopted Wechsler's multisubtest, multiscale format; the fifth edition (Roid, 2003) is theory-based and of exceptional psychometric quality. However, these improvements in the Binet were too little and too late to reclaim the throne it had shared for decades with Wechsler's scales.

WAIS-IV AND ITS PREDECESSORS

The first in the Wechsler series of tests was the Wechsler-Bellevue Intelligence Scale (Wechsler, 1939), so named because Wechsler was the chief psychologist at Bellevue Hospital in New York City (a position he held from 1932 to 1967). That first test, followed in 1946 by Form II of the Wechsler-Bellevue, had as a key innovation the use of deviation IQs (standard scores), which were psychometrically superior to the mental age divided by chronological age (MA/CA) formula that Terman had used to compute IQ. The Don't Forget box that follows shows the history of Wechsler's scales. The WAIS-IV is the great-great-grandchild of the original 1939 Wechsler-Bellevue Form I; it is also a cousin of the WISC-IV, which traces its lineage to Form II of the Wechsler-Bellevue.



As prodigious as Wechsler's contribution was to the assessment of children and adolescents, his impact on adult assessment might have been profound. As (Kaufman, 2010a) stated:

For the first Stanford-Binet, Terman's (1916) adult sample was small, haphazard, and unrepresentative. . . . The Average Adult level was derived from the mental ages for 62 adults. . . . Just as incredibly, Terman and Merrill (1937) tested no one above age 18 years for the standardization sample of Forms L and M of the Stanford-Binet. . . . For the revised Stanford-Binet, "A mental age of fifteen years represents the norm for all subjects who are sixteen years of age or older." (p. 30)

Dr Wechsler was not deterred by the difficulties in identifying representative samples of adults when he developed the Wechsler-Bellevue in the 1930s for ages 7 to 70 years. . . . For all practical purposes, *Dr Wechsler developed the first real test of intelligence for adults in 1939*, even though the Binet had been used to assess the mental ability of the adult population for a generation. (pp. xiv–xv)

The development of Wechsler's tests was originally based on practical and clinical perspectives rather than on theory per se. (The origin of each of the WAIS-IV subtests is shown in Rapid Reference 1.1.) Wechsler's view of IQ tests was that they were a way to peer into an individual's personality. Years after the development of the original Wechsler scales, extensive theoretical speculations have been made about the nature and meaning of these tests and their scores, and the newest WAIS-IV subtests were developed with specific

Rapid Reference 1.1

Origin of WAIS-IV Subtests

Verbal Comprehension Subtest

Similarities	Stanford-Binet
Vocabulary	Stanford-Binet
Information	Army Alpha
Comprehension	Stanford-Binet/Army Alpha

Working Memory Subtest

Digit Span	Stanford-Binet
Arithmetic	Stanford-Binet/Army Alpha
Letter-Number Sequencing	Gold, Carpenter, Randolph, Goldberg, & Weinberger (1997)

Perceptual Reasoning Subtest

Block Design	Kohs (1923)
Matrix Reasoning	Raven's Progressive Matrices (1938)
Visual Puzzles	Paper Form Board tasks trace back to the late 1920s (Roszkowski, 2001)
Figure Weights	Novel task developed by Paul E. Williams, PsyD (2005; pers. comm.)
Picture Completion	Army Beta/Army Performance Scale Examination

Processing Speed Subtest

Symbol Search	Shiffrin & Schneider (1977) and S. Sternberg (1966)
Coding	Army Beta/Army Performance Scale Examination
Cancellation	Diller et al. (1974); Moran & Mefford (1959); Talland & Schwab (1964)

theory in mind. However, the original Wechsler tasks were developed without regard to theory. Nonetheless, his influence continues to reverberate. He was one of the founders of the field of clinical psychology (Wasserman, 2012); “his tests and clinical approach have changed the lives of an infinite number of children, adolescents, and adults referred for evaluation for nearly a century” (Kaufman, in press). As Matarazzo (1981) aptly stated, “Probably the work of no other psychologists, including Freud or Pavlov, has so *directly* impinged upon the lives of so many people” (p. 1542).

WECHSLER-BELLEVUE SUBTESTS THAT SURVIVE ON THE WAIS-IV

Wechsler selected tasks for the Wechsler-Bellevue from among the numerous tests available in the 1930s, many of which were developed to meet the assessment needs of World War I. Although Wechsler chose not to develop new subtests for his intelligence battery, his selection process incorporated a blend of clinical, practical, and empirical factors. His rationale for each of the nine well-known original Wechsler-Bellevue subtests that survive to the present day on the WAIS-IV is discussed in the sections that follow.¹ (Note: The WAIS-III contained three new subtests that were not part of the earlier Wechsler batteries: Letter-Number Sequencing, Symbol Search, and Matrix Reasoning. The WAIS-IV contains three additional new subtests: Visual Puzzles, Figure Weights, and Cancellation. Subtests that were not a part of the original Wechsler batteries are discussed in separate sections of this chapter and in later chapters.)

Similarities (Verbal Comprehension Index)

Wechsler (1958) noted that before the Wechsler-Bellevue (W-B), “similarities questions have been used very sparingly in the construction of previous scales . . . [despite being] one of the most reliable measures of intellectual ability” (p. 72). Wechsler felt that this omission was probably due to the belief that language and vocabulary were necessarily too crucial in determining successful performance. However,

while a certain degree of verbal comprehension is necessary for even minimal performance, sheer word knowledge need only be a minor factor. More important is the individual’s ability to perceive the common elements of the terms he or she is asked to compare and, at higher levels, his or her ability to bring them under a single concept. (Wechsler, 1958, p. 73)

A glance at the most difficult items on the W-B I, WAIS, WAIS-R, and WAIS-III Similarities subtests (fly-tree, praise-punishment) makes it evident that Wechsler was successful in his goal of increasing “the difficulty of test items without resorting to esoteric or unfamiliar words” (p. 73).

Wechsler (1958) saw several merits in the Similarities subtest: It is easy to administer, has an interest appeal for adults, has a high *g* loading, sheds light on the logical nature of the person’s thinking processes, and provides other qualitative information as well. Regarding the latter point, he stressed the

1. Wechsler’s (1958) original quotes have been modified to avoid sexist language but are otherwise verbatim.

obvious difference both as to maturity and as to level of thinking between the individual who says that a banana and an orange are alike because they both have a skin, and the individual who says that they are both fruit. . . . But it is remarkable how large a percentage of adults never get beyond the superficial type of response. (Wechsler, 1958, p. 73)

Consequently, Wechsler considered his 0–1–2 scoring system to be an important innovation to allow simple discrimination between high-level and low-level responses to the same item. He also found his multipoint system helpful in providing insight into the evenness of a person's intellectual development. Whereas some individuals earn almost all 1s, others earn a mixture of 0, 1, and 2 scores. "The former are likely to bespeak individuals of consistent ability, but of a type from which no high grade of intellectual work may be expected; the latter, while erratic, have many more possibilities" (p. 74).

Vocabulary (Verbal Comprehension Index)

Contrary to lay opinion, the size of a person's vocabulary is not only an index of schooling, but also an excellent measure of general intelligence. Its excellence as a test of intelligence may stem from the fact that the number of words a person knows is at once a measure of learning ability, fund of verbal information and of the general range of the person's ideas. (Wechsler, 1958, p. 84)

The Vocabulary subtest formed an essential component of Binet's scales and the WAIS, but, surprisingly, this task, which has become prototypical of Wechsler's definition of verbal intelligence, was not a regular W-B I subtest. In deference to the objection that the word *knowledge* "is necessarily influenced by . . . educational and cultural opportunities" (p. 84), Wechsler included Vocabulary only as an alternative test during the early stages of W-B I standardization. Consequently, the W-B I was at first a 10-subtest battery, and Vocabulary was excluded from analyses of W-B I standardization data, such as factor analyses and correlations between subtest score and total score. Based on Wechsler's (1944) reconsideration of the value of Vocabulary and concomitant urging of examiners to administer it routinely, Vocabulary soon became a regular W-B I component. When the W-B II was developed, 33 of the 42 W-B I words were included in that battery's Vocabulary subtest. Since many W-B I words were therefore included in the WISC when the W-B II was revised and restandardized to become the Wechsler children's scale in 1949, Wechsler (1955) decided to include an all-new Vocabulary subtest when the W-B I was converted to the WAIS.

This lack of overlap between the W-B I Vocabulary subtest and the task of the same name on the WAIS, WAIS-R, WAIS-III, and WAIS-IV is of some concern

12 ESSENTIALS OF WAIS®-IV ASSESSMENT

regarding the continuity of measurement from the W-B I to its successors. Wechsler (1958) noted:

The WAIS list contains a larger percentage of action words (verbs). The only thing that can be said so far about this difference is that while responses given to verbs are easier to score, those elicited by substantives are frequently more significant diagnostically. (pp. 84–85)

This difference in diagnostic significance is potentially important because Wechsler found Vocabulary so valuable, in part because of its qualitative aspects: “The type of word on which a subject passes or fails is always of some significance” (p. 85), yielding information about reasoning ability, degree of abstraction, cultural milieu, educational background, coherence of thought processes, and the like.

Nonetheless, Wechsler was careful to ensure that the various qualitative aspects of Vocabulary performance had a minimal impact on quantitative score:

What counts is the number of words that a person knows. Any recognized meaning is acceptable, and there is no penalty for inelegance of language. So long as the subjects show that they know what a word means, they are credited with a passing score. (1958, p. 85)

Information (Verbal Comprehension Index)

Wechsler (1958) included a subtest designed to tap a person’s range of general information, despite “the obvious objection that the amount of knowledge which a person possesses depends in no small degree upon his or her education and cultural opportunities” (p. 65). Wechsler had noted the surprising finding that the fact-oriented information test in the Army Alpha group examination had among the highest correlations with various estimates of intelligence:

It correlated . . . much better with the total score than did the Arithmetical Reasoning, the test of Disarranged Sentences, and even the Analogies Test, all of which had generally been considered much better tests of intelligence. . . . The fact is, all objections considered, the range of a person’s knowledge is generally a very good indication of his or her intellectual capacity. (1958, p. 65)

Wechsler was also struck by a variety of psychometric properties of the Army Alpha Information Test compared to other tasks (excellent distribution curve, small percentage of zero scores, lack of pile-up of maximum scores), and the long history of similar factual information tests being “the stock in trade of mental examinations, and . . . widely used by psychiatrists in estimating the intellectual level of patients” (p. 65).

Always the astute clinician, Wechsler (1958) was aware that the choice of items determined the value of the Information subtest as an effective measure of intelligence. Items must not be chosen whimsically or arbitrarily but must be developed with several important principles in mind, the most essential being that, generally, “the items should call for the sort of knowledge that average individuals with average opportunity may be able to acquire for themselves” (p. 65). Wechsler usually tried to avoid specialized and academic knowledge, historical dates, and names of famous individuals, “but there are many exceptions to the rule, and in the long run each item must be tried out separately” (p. 66). Thus, he preferred an item such as “What is the height of the average American woman?” to ones like “What is iambic tetrameter?” or “In what year was George Washington born?” but occasionally items of the latter type appeared in his Information subtest. Wechsler was especially impressed with the exceptional psychometric properties of the Army Alpha Information Test “in view of the fact that the individual items on [it] left much to be desired” (p. 65).

Although Wechsler (1958) agreed with the criticism that factual information tests depended heavily on educational and cultural opportunities, he felt that the problem “need not necessarily be a fatal or even a serious one” (p. 65). Similarly, he recognized that certain items would vary in difficulty in different locales or when administered to people of different nationalities: “Thus, ‘What is the capital of Italy?’ is passed almost universally by persons of Italian origin irrespective of their intellectual ability” (p. 66). Yet he was extremely fond of information, considering it “one of the most satisfactory in the battery” (p. 67).

Comprehension (Verbal Comprehension Index)

Measures of general comprehension were plentiful in tests used before the W-B I, appearing in the original Binet scale and its revisions and in such group examinations as the Army Alpha and the National Intelligence Test. However, the test in multiple-choice format, though still valuable, does not approach the contribution of the task when individuals have to compose their own responses:

[O]ne of the most gratifying things about the general comprehension test, when given orally, is the rich clinical data which it furnishes about the subject. It is frequently of value in diagnosing psychopathic personalities, sometimes suggests the presence of schizophrenic trends (as revealed by perverse and bizarre responses) and almost always tells us something about the subject’s social and cultural background. (Wechsler, 1958, p. 67)

In selecting questions for the W-B I Comprehension subtest, Wechsler (1958) borrowed some material from the Army Alpha and the Army Memoirs

(Yoakum & Yerkes, 1920) and included a few questions that were also on the old Stanford-Binet, “probably because they were borrowed from the same source” (p. 68). He was not bothered by overlap because of what he perceived to be a very small practice effect for Comprehension: “It is curious how frequently subjects persist in their original responses, even after other replies are suggested to them” (p. 68).

The WAIS Comprehension subtest was modified from its predecessor by adding two very easy items to prevent a pile-up of zero scores and by adding three proverb items “because of their reported effectiveness in eliciting paralogical and concretistic thinking” (Wechsler, 1958, p. 68). Wechsler found that the proverbs did not contribute to the subtest exactly what he had hoped; they were useful for mentally disturbed individuals,

but “poor” answers were also common in normal subjects . . . [and] even superior subjects found the proverbs difficult. A possible reason for this is that proverbs generally express ideas so concisely that any attempt to explain them further is more likely to subtract than add to their clarity. (p. 68)

Despite the shortcomings of proverbs items, particularly the fact that they seem to measure skills that differ from prototypical general comprehension items (Kaufman, 1985), Wechsler (1981) retained the three proverbs items in the WAIS-R Comprehension subtest. Because these three items are relatively difficult (they are among the last five in the sequence), they are instrumental in distinguishing among the most superior adults regarding the abilities measured by WAIS-R Comprehension. Only two of the proverb items were retained on the WAIS-III, but the WAIS-IV includes four such items.

According to Wechsler (1958), Comprehension was termed a test of common sense on the Army Alpha, and successful performance

seemingly depends on the possession of a certain amount of practical information and a general ability to evaluate past experience. The questions included are of a sort that average adults may have had occasion to answer for themselves at some time, or heard discussed in one form or another. They are for the most part stereotypes with a broad common base. (pp. 68–69)

Wechsler was also careful to include no questions with unusual words “so that individuals of even limited education generally have little difficulty in understanding their content” (p. 69). Comprehension scores are, however, dependent on the ability to express one’s thoughts verbally.

Digit Span (Working Memory Index)

Memory Span for Digits (renamed Digit Span) combines in a single subtest two skills that subsequent research has shown to be distinct in many ways (Costa, 1975; Jensen & Figueroa, 1975): repetition of digits in the same order as they are spoken by the examiner and repetition of digits in the reverse order. Wechsler (1958) combined these two tasks for pragmatic reasons, but not theoretical ones: Each task alone had too limited a range of possible raw scores, and treating each set of items as a separate subtest would have given short-term memory too much weight in determining a person's IQ—1/6 instead of 1/11.

Wechsler was especially concerned about overweighing memory because Digit Span proved to be a relatively weak measure of general intelligence (*g*). He gave serious consideration to dropping the task altogether but decided to retain it for two reasons:

1. Digit Span is particularly useful at the lower ranges of intelligence; adults who cannot recall five digits forward and three backward are mentally retarded or emotionally disturbed “in 9 cases out of 10” (Wechsler, 1958, p. 71), except in cases of neurological impairment.
2. Poor performance on Digit Span is of unusual diagnostic significance, according to Wechsler, particularly for suspected brain dysfunction or concern about mental deterioration across the life span.

Digit Span also has several other advantages that may account for Wechsler's (1958) assertion that “perhaps no test has been so widely used in scales of intelligence as that of Memory Span for Digits” (p. 70): It is simple to administer and score, it measures a rather specific ability, and it is clinically valuable because of its unusual susceptibility to anxiety, inattention, distractibility, and lack of concentration. Wechsler noted that repetition of digits backward is especially impaired in individuals who have difficulty sustaining concentrated effort during problem solving. The test has been popularly “used for a long time by psychiatrists as a test of retentiveness and by psychologists in all sorts of psychological studies” (p. 70); because Wechsler retained Digit Span as a regularly administered subtest on the WAIS-R but treated it as supplementary on the WISC-R, it is evident that he saw its measurement as a more vital aspect of adult assessment than of child assessment.

The WAIS-IV provided an important innovation by adding a third section to the subtest—Digit Span Sequencing. For that section, examinees need to recall the numbers in ascending order, which (like Digits Backward) provides an excellent measure of working memory. “This change increases the role of mental manipulation and results in greater demands on working memory, relative to

previous versions of Digit Span” (Drozdick, Wahlstrom, Zhu, & Weiss, 2012, p. 198).

Arithmetic (Working Memory Index)

Wechsler (1958) included a test of arithmetical reasoning in an adult intelligence battery because such tests correlate highly with general intelligence; are easily created and standardized; are deemed by most adults as “worthy of a grownup”; have been “used as a rough and ready measure of intelligence” before the advent of psychometrics; and have “long been recognized as a sign of mental alertness” (p. 69). Such tests are flawed by the impact on test scores of attention span, temporary emotional reactions, and of educational and occupational attainment. As Wechsler notes: “Clerks, engineers and businessmen usually do well on arithmetic tests, while housewives, day laborers, and illiterates are often penalized by them” (p. 69). However, he believed that the advantages of an arithmetical reasoning test far outweighed the negative aspects. He pointed out that adults “may be embarrassed by their inability to do certain problems, but they almost never look upon the questions as unfair or inconsequential” (p. 69).

He took much care in developing the specific set of items for the W-B I and the WAIS and believed that his particular approach to constructing the Arithmetic subtest was instrumental in the task’s appeal to adults. Wechsler constructed items dealing with everyday, practical situations such that the solutions generally require computational skills taught in grade school or acquired “in the course of day-to-day transactions” (p. 70), and the responses avoid “verbalization or reading difficulties” (p. 69). Whereas the WISC-R and W-B I involve the reading of a few problems by the subject, all items on the WAIS, WAIS-R, WAIS-III, and WAIS-IV are read aloud by the examiner.

Bonus points for quick, perfect performance are not given to children on the WISC-R, but Wechsler considered the ability to respond rapidly to relatively difficult arithmetic problems to be a pertinent aspect of adult intelligence; bonus points are given to two items on the W-B I Arithmetic subtest, to four items on the WAIS task, to five items on WAIS-R Arithmetic, and to two items on WAIS-III Arithmetic. No bonus points are awarded on WAIS-IV Arithmetic, but only 30 seconds are allowed for each item. On the WAIS-IV, Arithmetic was modified, “to more purely reflect working memory, as opposed to verbal comprehension skills or mathematical knowledge. For example, difficult items require several successive simple mathematical steps that have to be represented in working memory instead of the complex calculations included in earlier versions” (Drozdick et al., 2012, p. 198).

Block Design (Perceptual Reasoning Index)

Kohs (1923) developed the Block Design test, which used blocks and designs that were red, white, blue, and yellow. His test was included in numerous other tests of intelligence and neuropsychological functioning before Wechsler adapted it for the W-B I. Wechsler (1958) shortened the test substantially; used designs having only two colors (although the W-B I blocks included all four colors, unlike the red and white WAIS and WAIS-III blocks); and altered the patterns that the examinee had to copy. Block Design has been shown to correlate well with various criterion measures, to be a good measure of *g*, and to be quite amenable to qualitative analysis (Wechsler, 1958). It intrigued Wechsler that those who do very well on this subtest are not necessarily the ones who treat the pattern as a gestalt; more often they are individuals who are able to break up the pattern into its component parts.

Wechsler (1958) believed that observation of individuals while they solve the problems, such as their following the entire pattern versus breaking it into small parts, provided qualitative, clinical information about their problem-solving approach, attitude, and emotional reaction that is potentially more valuable than the obtained scores. “One can often distinguish the hasty and impulsive individual from the deliberate and careful type, a subject who gives up easily or becomes disgusted, from the one who persists and keeps on working even after his time is up” (p. 80). He also felt that the Block Design subtest is most important diagnostically, particularly for persons with dementia or other types of neurological impairment. From Goldstein’s (1948) perspective, those with brain damage perform poorly on Block Design because of loss of the “abstract approach,” although Wechsler (1958) preferred to think that most “low scores on Block Design are due to difficulty in visual-motor organization” (p. 80).

Picture Completion (Perceptual Reasoning Index)

This subtest was commonly included in group-administered tests such as the Army Beta. A variant of this task, known as Healy Picture Completion II, which involves placing a missing piece into an uncompleted picture, was given individually in various performance scales, including the Army Performance Scale Examination; however, individual administration of Picture Completion, though conducted with the Binet scale for an identical task named Mutilated Pictures, was less common. Wechsler (1958) was unimpressed with the group-administered versions of Picture Completion because the subject had to draw in (instead of name or point to) the missing part, too few items were used, unsatisfactory items were included, and items were chosen haphazardly (a typical

set of items incorporated many that were much too easy and others that were unusually difficult).

Wechsler (1958) nonetheless believed that the test's "popularity is fully deserved" (p. 77); he tried to select an appropriate set of items while recognizing the difficulty of that task. "If one chooses familiar subjects, the test becomes much too easy; if one turns to unfamiliar ones, the test ceases to be a good measure of intelligence because one unavoidably calls upon specialized knowledge" (p. 77). He thought that the W-B I set of items was generally successful, although he had to increase the subtest length by 40% when developing WAIS Picture Completion to avoid a fairly restricted range of obtained scores. Although Wechsler was critical of the group-administered Picture Completion tasks, it is still noteworthy that four of the W-B I and WAIS items were taken directly from the Army Beta test, and an additional four items were clear adaptations of Beta items (using the same pictures, with a different part missing, or the same concept).

The subtest has several psychometric assets, according to Wechsler (1958), including brief administration time, minimal practice effect even after short intervals, and good ability to assess intelligence for low-functioning individuals. Two of these claims are true, but the inconsequential practice effect is refuted by data in the *WAIS-III Manual* (Psychological Corporation, 1997) and *WAIS-IV Technical and Interpretive Manual* (Psychological Corporation, 2008), which show test-retest gains for Picture Completion to average about 2 scaled-score points over intervals of a few weeks. Limitations of the task are that subjects must be familiar with the object in order to have a fair opportunity to detect what is missing and the susceptibility of specific items to sex differences. Wechsler (1958) notes that women did better in finding the missing eyebrow in the girl's profile and that men did better in detecting the missing thread on the electric light bulb. Similarly, on the WISC-R, about two-thirds of the boys but only about one-third of the girls across the entire 6–16 age range were able to find the missing "slit" in the screw; in contrast, many more girls than boys detected the sock missing from the girl who is running.

Because a person must first have the basic perceptual and conceptual abilities to recognize and be familiar with the object pictured in each item, Wechsler (1958) saw Picture Completion as measuring "the ability of the individual to differentiate essential from non-essential details" and "to appreciate that the missing part is in some way essential either to the form or to the function of the object or picture." But because of the total dependence of the assessment of this skill on the person's easy familiarity with the content of the item, "unfamiliar, specialized and esoteric subject matter must therefore be sedulously avoided when pictures are chosen for this test" (p. 78).

Coding (Processing Speed Index)

“The Digit Symbol [Coding on WAIS-IV] or Substitution Test is one of the oldest and best established of all psychological tests. It is to be found in a large variety of intelligence scales, and its wide popularity is fully merited” (Wechsler, 1958, p. 81). The W-B I Digit Symbol subtest was taken from the Army Beta, the only change being the reduction in response time from 2 minutes to 1 1/2 minutes to avoid a pile-up of perfect scores. For the WAIS, the number of symbols to be copied was increased by about one-third, although the response time remained unchanged.

Wechsler’s (1958) main concern regarding the use of Digit Symbol for assessing adult intelligence involved its potential dependency on visual acuity, motor coordination, and speed. He discounted the first two variables, except for people with specific visual or motor disabilities, but gave much consideration to the impact of speed on test performance. He was well aware that Digit Symbol performance drops dramatically with increasing age and is especially deficient for older individuals, who

do not write or handle objects as fast as younger persons, and what is perhaps equally important, they are not as easily motivated to do so. The problem, however, from the point of view of global functioning, is not merely whether the older persons are slower, but whether or not they are also “slowed up.” (p. 81)

Because correlations between Digit Symbol performance and total score remain high (or at least consistent) from age 16 through old age, Wechsler concluded that older people deserve the penalty for speed, “since resulting reduction in test performance is on the whole proportional to the subject’s overall capacity at the time he is tested” (p. 81). Although neurotic individuals also have been shown to perform relatively poorly on Digit Symbol, Wechsler attributed that decrement to difficulty in concentrating and applying persistent effort, that is, “a lessened mental efficiency rather than an impairment of intellectual ability” (p. 82).

Compared to earlier Digit Symbol or Substitution tests, Wechsler saw particular advantages to the task he borrowed from the Army Beta and included on his scales: It includes sample items to ensure that examinees understand the task, and it requires copying the unfamiliar symbols, not the numbers, lessening “the advantage which individuals having facility with numbers would otherwise have” (1958, p. 82).

Optional procedures were added to the WAIS-III Digit Symbol—Coding subtest, which were developed to help examiners assess what skills (or lack

thereof) may be impacting examinees' performance on the subtest. These optional procedures involve recalling shapes from memory (Pairing and Free Recall) and perceptual and graphomotor speed (Digit Symbol—Copy). However, these optional procedures were removed on WAIS-IV Coding.

WECHSLER'S LEGACY

When put into historical perspective, Wechsler made some mighty contributions to the clinical and psychometric assessment of intelligence. His insistence that every person be assessed on both Verbal and Performance scales went against the conventional wisdom of his time. Yet discrepancies between Verbal and Performance IQs (and ultimately among the four indexes that replaced the two IQs) would prove to have critical value for understanding brain functioning and theoretical distinctions between fluid and crystallized intelligence. Furthermore, Wechsler's stress on the clinical value of intelligence tests would alter the face of intellectual assessment forever, replacing the psychometric, statistical emphasis that accompanied the use and interpretation of the Stanford-Binet. Finally, Wechsler's inclusion of a multiscore subtest profile (as well as three IQs instead of one) met the needs of the emerging field of learning disabilities assessment in the 1960s, to such an extent that his scales replaced the Stanford-Binet as king of IQ during that decade. It has maintained that niche ever since for children, adolescents, and adults (Alfonso, LaRocca, Oakland, & Spanakos, 2000; Archer, Buffington-Vollum, Stredny, & Handel, 2006; Archer & Newsom, 2000; Camara, Nathan, & Puente, 2000; Rabin, Barr, & Burton, 2005).

The popularity of the adult Wechsler tests, starting with the WAIS and continuing with the WAIS-R, WAIS-III, and WAIS-IV, is remarkable and pervasive. Wechsler's adult scales are by far the first choice for measuring intelligence among clinical neuropsychologists (Rabin et al., 2005), psychologists who conduct forensic assessments (Archer et al., 2006), clinical psychologists (Camara et al., 2000), psychologists who conduct evaluations in state correctional facilities (Gallagher, Somwaru, & Ben-Porath, 1999), psychology professors who train doctoral-level students (Belter & Piotrowski, 2001), and, indeed, psychologists who conduct assessments with adults for any other reason (Groth-Marnat, 2009; Sattler & Ryan, 2009; Weiss et al., 2010). Harrison, Kaufman, Hickman, and Kaufman (1988) reported data from a survey of 402 clinical psychologists that showed 97% of these professionals utilized the WAIS or WAIS-R when administering an adult measure of intelligence. Even if the 97% figure is no longer exactly precise, the WAIS-IV has clearly continued the Wechsler tradition as being by far the most popular test of adult intelligence.

PURPOSES OF ASSESSING ADULTS AND ADOLESCENTS

As mentioned previously, adults were historically assessed because of a need to place men into the appropriate level of the military service or to determine a person's mental deficiency. Today, reasons for assessing adolescents and adults commonly include measuring cognitive potential or neurological dysfunction, obtaining clinical information, making educational or vocational placement decisions, and developing interventions for educational or vocational settings. Harrison et al. (1988) found that practitioners who assess adults most often report using intelligence tests to measure cognitive potential and to obtain clinically relevant information. About 77% of practitioners reported using intelligence tests for obtaining information about neurological functioning, and fewer than 50% reported using intelligence tests for making educational or vocational placements or interventions (Harrison et al., 1988). Camara and colleagues (2000) also reported that a large proportion of the assessment services of clinical psychologists and neuropsychologists are in the areas of intellectual/achievement assessment (20–34%) and neuropsychological assessment (13–26%).

FOUNDATIONS OF THE WAIS-IV: THEORY AND RESEARCH

Wechsler defined intelligence as “the capacity to act purposefully, to think rationally, and to deal effectively with his [or her] environment” (1944, p. 3). His concept of intelligence was that of a global entity that could also be categorized by the sum of many specific abilities. The most recent revision of Wechsler's adult intelligence scale, the WAIS-IV, has enhanced measures of more discrete domains of cognitive functioning, such as working memory and processing speed (Psychological Corporation, 2008), while continuing to provide a measure of global intelligence. Unlike the earliest Wechsler tests, the WAIS-IV also was developed with specific theoretical foundations in mind. In fact, revisions were made purposely to reflect the latest knowledge from literature in the areas of intelligence theory, adult cognitive development, and cognitive neuroscience. The theoretical constructs of fluid reasoning, working memory, and processing speed were of particular importance during the development of the WAIS-IV, just as they were in the development of the WISC-IV. Rapid Reference 1.2 defines these three theoretical constructs.

Wechsler's adult tests, from the Wechsler-Bellevue (1939) to the WAIS (1955) to the WAIS-R (1981), took the same basic form, with 6 subtests constituting the Verbal Scale, 5 making up the Performance Scale, and all 11 yielding the global

Rapid Reference 1.2

Updated WAIS-IV Theoretical Foundations

Theoretical Construct	Fluid Reasoning	Working Memory	Processing Speed
Definition	Ability to process or manipulate abstractions, rules, generalizations, and logical relationships	Ability to actively maintain information in conscious awareness, perform some operation or manipulation with it, and produce a result	Ability to process information rapidly (which is dynamically related to one's ability to perform higher-order cognitive tasks)
References for the Construct	Carroll (1997); Cattell (1943, 1963); Cattell & Horn (1978); Sternberg (1995)	Beuhner, Krumm, Ziegler, & Pluecken (2006); Unsworth & Engle (2007)	Fry & Hale (1996); Kail (2000); Kail & Hall (1994); Kail & Salthouse (1994)

entity of intelligence characterized by the Full Scale IQ. The WAIS-III departed slightly from the original form by offering four separate indexes (i.e., Verbal Comprehension Index, Perceptual Organization Index, Working Memory Index, and Processing Speed Index), in addition to the Verbal, Performance, and Full Scale IQs. The WAIS-IV, like the WISC-IV, departed dramatically from the longtime Wechsler tradition by eliminating the Verbal and Performance IQs and, hence, the ever-popular V-P IQ discrepancy. The four indexes were retained in the WAIS-IV, alongside the Full Scale IQ, providing a more modern and conceptually clearer scale structure. The WAIS-IV and WISC-IV now offer the same four indexes: Verbal Comprehension (VCI), Perceptual Reasoning (PRI), Working Memory (WMI), and Processing Speed (PSI). (To achieve this synchrony, the WAIS-IV and WISC-IV Perceptual Organization Index was renamed the Perceptual Reasoning Index, and WISC-IV Freedom from Distractibility Index became the Working Memory Index.)

The focus on the four indexes in the WAIS-IV psychometric profile is a plus when it comes to understanding how to interpret individual profiles, from both a theoretical and a clinical perspective. However, this shift in focus also affects WAIS-IV Full Scale IQ (FSIQ), which is now computed from the sum of the 10 subtests that compose the four scales (3 VCI, 3 PRI, 2 WMI, and 2 PSI).

 Rapid Reference 1.3

**Comparison of the Subtest Composition of the WAIS-III
and WAIS-IV Full Scales**

WAIS-III Full Scale Subtests	WAIS-IV Full Scale Subtests
Verbal	
Vocabulary	Vocabulary (VCI)
Similarities	Similarities (VCI)
Information	Information (VCI)
Comprehension	
Arithmetic	Arithmetic (WMI)
Digit Span	Digit Span (WMI)
Performance	
Block Design	Block Design (PRI)
Matrix Reasoning	Matrix Reasoning (PRI)
	Visual Puzzles (PRI)
Picture Completion	
Picture Arrangement	
Digit Symbol—Coding	Coding (PSI)
	Symbol Search (PSI)

Traditionally, the WAIS FSIQ has been composed of 11 subtests, 6 Verbal and 5 Performance. The end result of these changes is a WAIS-IV FSIQ that differs substantially from WAIS-III FSIQ, as shown in Rapid Reference 1.3. Of the 11 WAIS-III Full Scale subtests, only 8 are retained on the WAIS-IV Full Scale. Although this shift is not as dramatic as the change from the WISC-III to the WISC-IV Full Scale (which share only 5 of 10 subtests), it is nonetheless notable.

Although two global scores were eliminated from the WAIS-IV (Verbal and Performance IQs), one new global score was added, the optional General Ability Index (GAI). The GAI is derived from the sum of scaled scores on the three Verbal Comprehension and three Perceptual Reasoning subtests, thereby eliminating the WMI and PSI from consideration and forming a global composite composed solely of the verbal and perceptual constructs. This new global score aids examiners in interpreting test profiles and is included in our step-by-step interpretive system (see Chapter 5), just as the WISC-IV GAI is incorporated into its interpretive system (Flanagan & Kaufman, 2009).

DON'T FORGET

New WAIS-IV Four-Factor Structure

Verbal Indexes

1. Verbal Comprehension
3. Working Memory

Nonverbal Indexes

2. Perceptual Reasoning
4. Processing Speed

Note: The Perceptual Reasoning Index (PRI) was called the Perceptual Organization Index (POI) on the WAIS-III. This name change (on the WISC-IV as well) reflects “the increased emphasis of fluid reasoning on the WAIS-IV” (Drozdick et al., 2012, p. 198).

Description of WAIS-IV

Several issues prompted the revision of the WAIS-IV; the Manual clearly details these issues and what changes were made (Psychological Corporation, 2008, pp. 7–23). Rapid Reference 1.4 lists key features that were adapted for the Fourth Edition.

WAIS-III examiners will recognize many of the core Wechsler subtests in the WAIS-IV, but there have been several notable changes with the addition of new

Rapid Reference 1.4

WAIS-IV Key Revisions

- Updated theoretical foundations
- Updated norms
- Increased developmental appropriateness
- Increased user-friendliness
- Enhanced clinical utility
- Decreased reliance on timed performance
- Enhancement of fluid reasoning measurement by adding Figure Weights and Visual Puzzles subtests
- Enhancement of working memory measurement by adding Digit Span Sequencing as the third component of Digit Span
- Strengthening the framework based on factor analysis, including state-of-the-art confirmatory factor analysis
- Statistical linkage to other measures of cognitive functioning and achievement, most notably the Wechsler Memory Scale—Fourth Edition (WMS-IV)
- Extensive testing of reliability and validity

subtests and modifications to the overall structure. (Rapid Reference 1.5 lists a description of all WAIS-IV subtests.) There are three new subtests:

1. **Visual Puzzles** (added to the Perceptual Reasoning Index; it is a visual-perceptual variation of the old Object Assembly subtest, a task that required both visual-perceptual abilities and visual-motor coordination)
2. **Figure Weights** (added to the Performance Reasoning Index as a supplemental subtest)
3. **Cancellation** (added to the Processing Speed Index as a supplemental subtest)

Rapid Reference 1.5

WAIS-IV Subtest Abbreviations and Descriptions

Subtest	Abbreviation	Description
Verbal Comprehension Subtest		
Similarities	SI	The examinee is presented with two words that represent common objects or concepts and describes how they are similar.
Vocabulary	VC	For picture items, the examinee names the object presented visually. For verbal items, the examinee defines words that are presented visually and orally.
Information	IN	The examinee answers questions that address a broad range of general knowledge topics.
Comprehension	CO	The examinee answers questions based on his or her understanding of general principles and social situations.
Perceptual Reasoning Subtest		
Block Design	BD	Working within a specified time limit, the examinee views a model and a picture or a picture only and uses red-and-white blocks to recreate the design.
Matrix Reasoning	MR	The examinee views an incomplete matrix or series and selects the response option that completes the matrix or series.
Visual Puzzles ^a	VP	Working within a specified time limit, the examinee views a completed puzzle and selects three response options that, when combined, reconstruct the puzzle.

(continued)

26 ESSENTIALS OF WAIS®-IV ASSESSMENT

Subtest	Abbreviation	Description
Figure Weights ^a	FW	Working within a specified time limit, the examinee views a scale with missing weight(s) and selects the response option that keeps the scale balanced.
Picture Completion ^b	PC	Working within a specified time limit, the examinee views a picture with an important part missing and identifies the missing part.
Working Memory Subtest		
Digit Span	DS	For Digit Span Forward, the examinee is read a sequence of numbers and recalls the numbers in the same order. For Digit Span Backward, the examinee is read a sequence of numbers and recalls the numbers in reverse order. For Digit Span Sequencing, the examinee is read a sequence of numbers and recalls the numbers in ascending order.
Arithmetic	AR	Working within a specified time limit, the examinee mentally solves a series of arithmetic problems.
Letter-Number Sequencing	LN	The examinee is read a sequence of numbers and letters and recalls the numbers in ascending order and the letters in alphabetical order.
Processing Speed Subtest		
Symbol Search	SS	Working within a specified time limit, the examinee scans a search group and indicates whether one of the symbols in the target group matches.
Coding	CD	Using a key, the examinee copies symbols that are paired with numbers within a specified time limit.
Cancellation ^a	CA	Working within a specified time limit, the examinee scans a structured arrangement of shapes and marks target shapes.

^a New WAIS-IV subtest.
^b The WAIS-IV Record Form and Manual uses PCm as the abbreviation for Picture Completion, but in this book, we use PC.

How these new subtests were created gives interesting insight into the process of test development and revision. Professionals on the Research Development (RD) Team for the WAIS-IV shared how Figure Weights and Visual Puzzles were developed for the WAIS-IV (Cancellation was developed

first for the WISC-IV). Dr. Susan Raiford (personal communication, November 25, 2008) revealed:

Visual Puzzles was inspired by Object Assembly as an abstract nonmotor task that was similar. Jim Holdnack, one of the WMS-IV RDs, submitted the item type for consideration in April of 2005, and it was originally named “Puzzle Pieces.” . . . As the subtest evolved we were aware of the similarities to the old Paper Form Board tests through reviews of Carroll’s work and of existing measures (Quasha & Likert) published many years ago by Psychcorp. We found as we worked with the item type that difficulty could be controlled with complexity of cut and with internal cues (colors or lines), which is why the internal cues are there on the easier items and the complexity of piece cut gets greater as the items progress.

Dr. Holdnack (personal communication, November 25, 2008) continued:

The subtest was inspired from the Object Assembly subtest and the Visual Puzzles and Geometric Puzzles on NEPSY-II, although, the make-up of this test varies considerably from those subtests. Mostly, I was shooting for the items to have elements of mental construction and rotation while limiting other confounding factors such as verbalization, processing speed, and fine-motor integration.

Paul Williams, a research director at the Psychological Corporation, submitted the original Figure Weights item in 2005 (Raiford, personal communication). Dr. Williams explained (personal communication, December 1, 2008):

[T]he hard part was coming up with a way to create a relationship between the objects. I couldn’t use symbols such as $=$ $+-$ because this would require prior knowledge. So the thought came to me that another way to symbolize $>$ and $<$ is by weight; which led to the idea of using a balance to create a rule or relationship between the figures. With this information a series of rules can be presented which has to be reasoned out by the examinee to balance the final scale. Susie then took it from there and did an amazing job building the items and doing the science necessary to develop the idea into a functional subtest.

Dr. Raiford (personal communication) continued:

Paul told me at the time that he intended it to be a new item type for Matrix Reasoning, but we thought we could make a whole subtest out of it, and wanted to because it seemed to be measuring quantitative reasoning, which

28 ESSENTIALS OF WAIS®-IV ASSESSMENT

we weren't measuring nonverbally yet. I switched the item type to a scale from the seesaws . . . because it seemed more intuitive. I also found we could get all the difficulty we needed with just two scales establishing relationships and a third scale with an empty tray.

In addition to these three new subtests, other modifications to the WAIS-III include the removal of two of Wechsler's original group of subtests from the revised test: Picture Arrangement and Object Assembly. The rationale for deleting these subtests was to lessen the motor demands of the test and to deemphasize time bonus points. When Object Assembly was originally developed, Wechsler (1958) "wanted at least one test which required putting things together into a familiar configuration" (pp. 82–83). He included Object Assembly, but only "after much hesitation" (p. 82), because of its known liabilities: relatively low reliability and predictive value, large practice effects, and low correlations with other subtests. In the development of Picture Arrangement, Wechsler selected items for his test based on "interest of content, probable appeal to subjects, ease of scoring and discriminating value" (p. 75). Yet he was never satisfied with the result, noting that "the final selection leaves much to be desired."

He spent much time and statistical analysis trying to discern which alternative responses deserved credit and even called in a team of four judges, yet the final system for assigning credit for alternative arrangements "turned out to be more or less arbitrary" (p. 76). Although bonus points were included on earlier editions of the WAIS Picture Arrangement, Wechsler (1981) reversed this trend for the WAIS-R and deemphasized speed greatly by not allowing bonus points for any of the Picture Arrangement items. Thus, Wechsler's concerns about these two subtests are consistent with The Psychological Corporation's decision to eliminate them from the WAIS-IV (and from the WISC-IV). Nonetheless, had he been alive, Wechsler undoubtedly never would have agreed to eliminate these original subtests from any version of the WAIS or WISC. He would, however, have gained solace from the fact that both Object Assembly and Picture Arrangement are included in the Wechsler Nonverbal Scale of Ability (WNV; Wechsler & Naglieri, 2006).

Further deletions from the WAIS-III to the WAIS-IV included removal of the optional procedures: Digit Symbol—Incidental Learning and Digit Symbol—Copy. However, process scores were added to the WAIS-IV Block Design, Digit Span, and Letter-Number Sequencing subtests that allow examiners to analyze errors and qualitatively interpret test performance. For example, Block Design No Time Bonus is a process score that reflects a person's performance without additional time bonus for rapid completion of items. The Digit Span task offers three process scores that reflect an examinee's performance on the separate tasks of

repeating digits forward, backward, and then sequencing digits. The addition of the Digit Span Sequencing task is consistent with the test publisher's theoretical emphasis on working memory. An additional process score is offered for another Working Memory subtest, which involves the calculation of the longest Letter-Number sequence recalled. A comparison of Digit Span Sequencing and Letter-Number Sequencing will provide an auditory analog of a comparison of Trail Making A and B. Rapid Reference 1.6 describes the subtests' process analyses.

Rapid Reference 1.6

Subtests with Process Analysis

Subtest	Abbreviation	Process Score	Use
Block Design			
Block Design No Time Bonus	BDN	Score reflects performance on BD without additional time bonus for rapid completion.	Useful when physical limitations, problem-solving strategies, or personality characteristics affect performance on timed tasks.
Digit Span			
Digit Span Forward	DSF	Raw scores reflect the total number of DSF trials correctly completed before discontinuing.	May help to explain variable performance on Digit Span Tasks. DSF requires immediate auditory recall, whereas DSB and DSS place demands on working memory and attention.
Digit Span Backward	DSB	Raw scores reflect the total number of DSB trials correctly completed before discontinuing.	
Digit Span Sequencing	DSS	Raw scores reflect the total number of DSS trials correctly completed before discontinuing.	

(continued)

30 ESSENTIALS OF WAIS®-IV ASSESSMENT

Subtest	Abbreviation	Process Score	Use
Longest Digit Span Forward	LDSF	Raw scores reflect the number of forward digits recalled on the last trial scored 1 point.	May help to explain variable performance on DS tasks. Some examinees may arrive at their DS total raw score by inconsistently earning 1s and 0s across trials, whereas other examinees may show a pattern of consistently earning 1s until they discontinue the task.
Longest Digit Span Backward	LDSB	Raw scores reflect the number of backward digits recalled on the last trial scored 1 point.	
Longest Digit Span Sequencing	LDSS	Raw scores reflect the number of digits correctly sequenced on the last trial scored 1 point.	
Letter-Number Sequencing			
Longest Letter-Number Sequence	LLNS	Raw scores reflect the number of letters and numbers correctly sequenced on the last trial scored 1 point.	May help to explain variable performance on LN tasks. Some examinees may arrive at their LN total raw score by inconsistently earning 1s and 0s across trials, whereas other examinees may show a pattern of consistently earning 1s until they discontinue the task.

Validity of the WAIS-IV Model

With the addition of the 3 new subtests and removal of 2 subtests, the complete WAIS-IV comprises 15 subtests, although only 10 are core subtests needed to compute the 4 indexes and FSIQ. Like the WISC-IV structure, the WAIS-IV structure focuses users on the middle tier of scores—the Factor Indexes (see Figure 1.1). FSIQ and the indexes have a mean of 100 and a standard deviation of 15. Subtest scaled scores have a mean of 10 and standard deviation of 3.

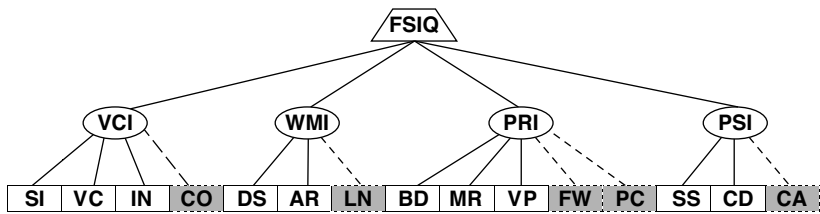


Figure 1.1. WAIS-IV Structure: Three-Tier Hierarchy

Note: Shaded subtests that are bordered with dashed lines and connected to indexes with dashed lines are supplemental and contribute to the calculation of the Index score only if they have substituted for one of the core subtests.

Of the five supplemental subtests, three are normed only for ages 16 to 69: Letter-Number Sequencing (WMI), Figure Weights (PRI), and Cancellation (PSI). Comprehension (VCI) and Picture Completion (PRI) are normed for the complete 16- to 90-year range. Supplemental subtests are not included in calculation of any of the Index scores.

The *WAIS-IV Technical and Interpretive Manual* (Psychological Corporation, 2008) reports the details of several confirmatory factor analysis (CFA) studies that support the underlying four-factor structure of the WAIS-IV. For all ages, there is strong construct validity support for the four Indexes. However, at both ages 16–69 and ages 70–90, a model that allows Arithmetic to load on both the Working Memory Factor and the Verbal Comprehension Factor fits the data best. For ages 16–69, the Arithmetic subtest had a Factor loading of .75 on the Working Memory Factor and a small loading of .08 on the Verbal Comprehension Factor. For ages 70–90, the Arithmetic subtest had a loading of .48 on the Working Memory Factor and .33 on the Verbal Comprehension Factor. The Figure Weights subtest also had a split factor loading for ages 16–69, with factor loadings of .37 and .43 on the Working Memory Factor and Perceptual Reasoning Factor, respectively. The *WAIS-IV Technical and Interpretive Manual* (Psychological Corporation, 2008) also reported a different four-factor model, in which it allowed a correlated error for Digit Span and Letter-Number Sequencing and a cross-loading for Arithmetic on a *Gc* factor; these changes helped the WAIS-IV scoring model considerably.

Subsequent CFAs by Ward, Bergman, and Hebert (2012) and by Canivez and Watkins (2010a, 2010b) offered empirical support for the WAIS-IV four-factor model, but these research teams differed in the best way to interpret the data. Ward et al. (2012) modified details of the structural model based on cognitive

theory (e.g., they emphasized the role of working memory on Figure Weights) and, ultimately, supported the notion that methodologists must take into consideration the “conceptual coherence and theoretical adherence in addition to statistical fit” (p. 1). Canivez and Watkins identified the four designated WAIS-IV factors, but noted that these four first-order factors accounted for relatively small amounts of variance compared to the second-order *g* factor; they argued “that the WAIS-IV provides strong measurement of general intelligence, and clinical interpretation should be primarily at that level” (2010b, p. 827).

It was in the spirit of alternative approaches to interpretation—both empirical and theoretical—that we offered a five-factor theory-based supplementary interpretive system in the first edition of this book. We based that alternative interpretive system on the results of CFAs of the WAIS-IV conducted by Tim Keith (personal communication, January 30, 2009). He analyzed the averaged matrix for ages 16–90 shown in the *WAIS-IV Manual* (Psychological Corporation, 2008, p. 62) and used the technique of higher-order CFA. Keith’s analyses compared various models, including the Four-Factor WAIS-IV model and a Five-Factor model that is in line with the Cattell-Horn-Carroll (CHC) theory. This CHC model included Matrix Reasoning and Figure Weights on the Fluid Reasoning (*Gf*) Factor, along with Arithmetic. The Visual Processing (*Gv*) Factor included Block Design, Visual Puzzles, and Picture Completion. The Crystallized Knowledge (*Gc*) Factor included Similarities, Vocabulary, Comprehension, and Information. Short-Term Memory (*Gsm*) included Digit Span and Letter-Number Sequencing, and Processing Speed (*Gs*) included Coding, Symbol Search, and Cancellation. Keith reported that a CHC model with separate *Gf* and *Gv* Factors fits the data especially well. Arithmetic, though included on the WMI, is associated with the *Gf* factor in Keith’s analysis. The loadings are shown in Figure 1.2. Note that *Gf* is indistinguishable from the general factor (*g*). Also note that Figure Weights shows a high loading (.77) on a *Gf* Factor.

Benson, Hulac, and Kranzler (2010) conducted CFAs of the WAIS-IV that supported Keith’s five CHC factors. Benson et al. (2010) concluded from their analyses: “Results suggest that a CHC model provides a better explanation of test performance than does the WAIS-IV scoring structure” (p. 124). The authors also addressed the issue of whether CHC factor structure was invariant across ages 16–17 through 65–69. Their analyses “provide modest support for the conclusion that the WAIS-IV measures the same constructs across age groups. . . . [S]ome age differences exist, although these differences may not be large enough to be practically meaningful” (Benson et al., 2010, p. 129). Ward et al. (2012) and Weiss, Keith, Zhu, and Chen (in press) also compared five-factor CHC-based solutions to conventional four-factor WAIS-IV solutions. Ward et al. (2012) commented:

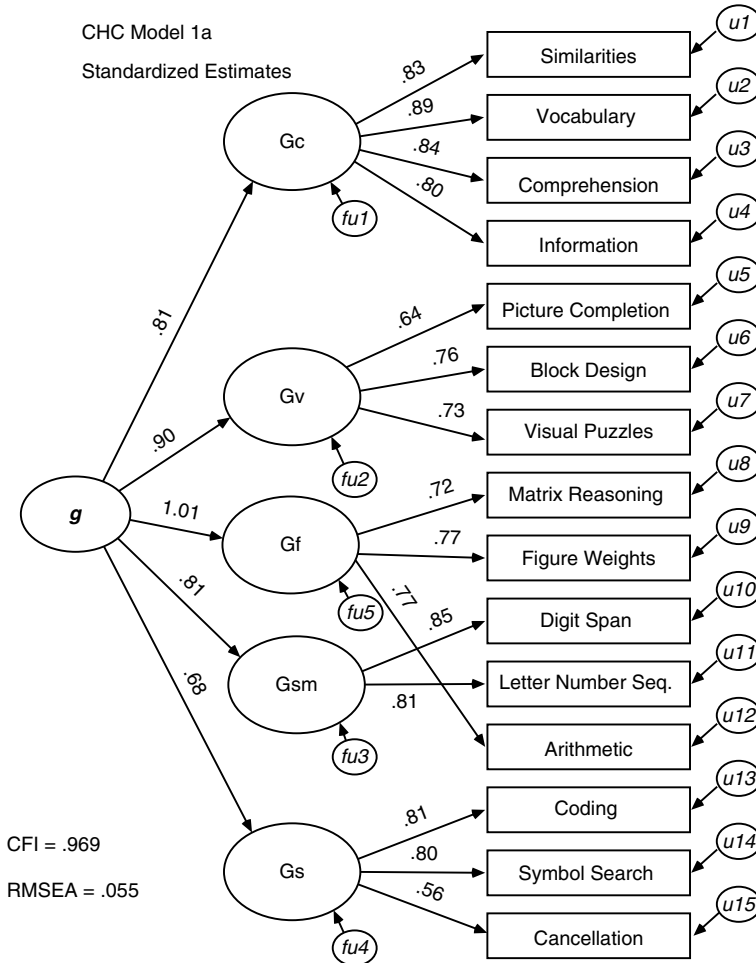


Figure 1.2. Keith's CHC Model for the WAIS-IV for Ages 16–69 Years

Note that this analysis is based on data for the total WAIS-IV sample.

Source: T. Keith, personal communication, January 30, 2009.

“The Benson et al. (2010) research is important because it is based on current taxonomic theory [but] . . . their statistical and theoretical appraisal of the model is relatively uncritical” (p. 331). Ward et al. (2012) concluded from their sophisticated set of analyses that both four-factor and five-factor models of the WAIS-IV have merit, but they favor a four-factor model because it is more consistent with WAIS-IV Index structure, with the test's theory and constructs,

with “current cognitive theory (e.g., Kane et al., 2004)” (p. 339), and with “WAIS-IV factor-analytic antecedents” (p. 339).

Weiss et al. (in press) also found empirical support for both four- and five-factor solutions, noting that the five-factor model provided a better fit. Most importantly, Weiss et al. (in press) found that the CFAs for both four and five factors were invariant for normal ($N = 1,800$) and clinical ($N = 411$) samples of adults. These authors interpret their results as providing strong validity evidence for both models and see the four-factor and CHC solutions as being complementary for normal adults as well as clinical patients. Their inclination is to begin interpretation with the four Indexes, but to consider switching to five factors if there are discrepant scores within one or more Indexes.

We have retained Keith’s original analysis in Figure 1.2, based on ages 16–90, rather than limiting the analysis to ages 16–69 or reporting data from recently published studies. We continued to use Keith’s data because, in the first edition of this book, we frequently referred to the CHC factors as the “Keith factors,” and we will continue to do so in this edition when referring to the CHC interpretive system for ages 16–69. Moreover, the comparable WAIS-IV analysis for ages 16–69 (Benson et al., 2010, Figure 3) is nearly identical to Keith’s analysis regarding the structure and factor loadings.

Benson et al. (2010) support a five-factor CHC solution; Ward et al. (2012) prefer four-factor models; Weiss et al. (in press) consider both models viable and complementary; and Canivez and Watkins (2010a, 2010b) advocate a one-factor (g) interpretation. We disagree emphatically with the clinical value of the g approach (see Chapters 4 and 5), but support *both* the four-factor and five-factor interpretive models. We believe that examiners should select whichever solution best fits their professional orientation or—on a case-by-case basis—whichever approach best fits the subtest profile for an individual referred for evaluation. For example, Keith (personal communication) concluded from his CFA analyses that “a CHC-based interpretation of the WAIS-IV is, at minimum, worth considering. I would certainly consider that interpretation if there were inconsistencies among the Perceptual Reasoning tasks, or between Arithmetic versus the Working Memory tasks.”

In the previous edition of this book, we offered the four-factor solution for all ages, 16–90 years; however, we only offered the five-factor CHC solution for individuals ages 16–69 because two of the component subtests (Figure Weights and Letter-Number Sequencing) are not administered to ages 70–90 years. This limitation is no longer true for the second edition of *Essentials of WAIS-IV Assessment*. Recent CFA research produced a new CHC five-factor model for adults ages 70–90 years (Näileksela, Reynolds, & Kaufman, 2012).

Niileksala et al.'s approach rests on two main decisions: (a) to consider Arithmetic as primarily a measure of Gf , which has empirical support from Keith's study (see Figure 1.2) and from other CFA investigations of the WAIS-IV (Benson et al., 2010; Ward et al., 2012); and (b) to treat the three components of Digit Span (Forward, Backward, Sequencing) as separate variables in the factor analyses. The latter decision is supported by data presented in Chapters 4 and 5 of the *WAIS-IV Technical and Interpretive Manual* (Psychological Corporation, 2008): Each section of Digit Span has adequate internal consistency reliability across the age range (means of .81–.83) and for 13 clinical samples (.86–.89) (Tables 4.1 and 4.2 of WAIS-IV manual); each section has adequate stability (means of .71–.77 for four age groups, comparable to the value of .74 for Matrix Reasoning; Table 4.5 of manual); and the three parts intercorrelated .42–.53 for all ages (Table 5.1 of manual), a moderate degree of relationship that supports entering them as separate variables in CFAs (Niileksala et al., 2012). Figure 1.3 (Niileksala, personal communication, October 21, 2011) shows the new CHC structure for ages 70–90 years. As was found for the Keith CHC structure, Gf is indistinguishable from g , even though the “new” Gf factor is composed only of Matrix Reasoning and Arithmetic, and a strong Gsm factor emerged based on substantial loadings by all three sections of Digit Span. Notably, the g loading for the Gsm factor in the analysis for ages 70–90 (.84) is similar to the g loading of .81 for Keith's Gsm factor. The other three factors— Gc , Gv , and Gs —are virtually identical for ages 16–69 and 70–90, and the “fit” statistics (CFI and RMSEA) are excellent for both models (shown in Figures 1.2 and 1.3).

Although Niileksala et al. (2012) had as their main research goal the extension of the CHC model to ages 70–90, they also investigated the new model for the entire 16–90 age range and verified that this model was invariant across age groups; the factor structure, factor loadings, and “fit” statistics for ages 16–69 mirrored the values shown in Figure 1.3 for the elderly sample. These analyses provide empirical support for an alternative CHC model for ages 16–69 in those instances where examiners did not administer Figure Weights and Letter-Number Sequencing (both supplemental). The Keith CHC factors require those two subtests, but the Niileksala model does not.

The technique of CFA, championed most notably by Tim Keith and Matt Reynolds, has dramatically changed the way clinicians interpret all current measures of cognitive abilities for children and adults; for a thorough, insightful explanation of how sophisticated psychometrics has enhanced examiners' understanding of the intricacies of the patterns of strengths and weaknesses in test profiles on widely used intelligence tests, see the recent chapter by Keith and Reynolds (2012).

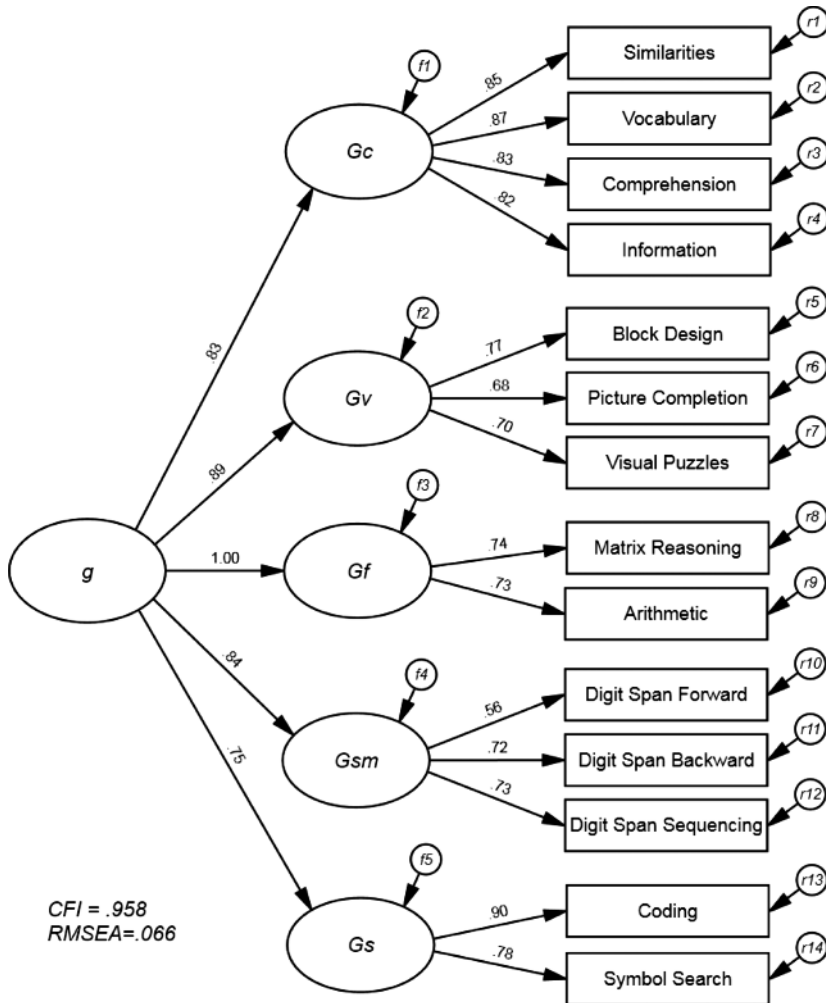


Figure 1.3. New CHC Model for the WAIS-IV for Ages 70–90 Years

Source: C. Nieleksala, personal communication, October 21, 2011 (based on analyses conducted by Nieleksala et al., 2012)

WAIS-IV's Relationship with the WMS-IV

The WAIS-IV was codeveloped and conormed with the Wechsler Memory Scale—Fourth Edition (WMS-IV; Wechsler, 2009b), and the two tests are commonly administered together as part of clinical, forensic, and neuropsychological evaluations,

especially for older individuals (Groth-Marnat, 2009). The integration of these two Wechsler batteries as clinical and psychometric tools has been discussed insightfully by Lisa Drozdick, Jim Holdnack, and colleagues in *Essentials of WMS-IV Assessment* (Drozdick, Holdnack, & Hilsabeck, 2011) and elsewhere (Drozdick et al., 2012; Holdnack & Drozdick, 2010; Holdnack, Zhou, Larrabee, Millis, & Salthouse, 2011; Wechsler, 2009a). Consult the complete set of articles in a special issue of *Assessment* devoted to advancing WAIS-IV and WMS-IV clinical interpretation (Frazier, 2011).

Holdnack et al. (2011) explored several different CFA models to best explain the relationship between the constructs measured by the two test batteries for ages 16–69 ($N = 900$). Variables included the 10 primary “Full Scale” WAIS-IV subtests and six WMS-IV subtests. The six WMS-IV subtests were Logical Memory (recall for a short story), Verbal Paired Associates (recall for related and unrelated word pairs), Designs (recall of spatial locations and visual details), Visual Reproduction (recall of geometric designs), Spatial Addition (ability to manipulate visual-spatial information in working memory), and Symbol Span (ability to manipulate designs in working memory). Only the delayed portions of Logical Memory, Verbal Paired Associates, Designs, and Visual Reproduction were entered into the CFA, because it is unlikely that CFA “can differentiate immediate from delayed memory functioning as this often leads to model specification errors consequent to method variance” (Holdnack et al., 2011, p. 180). They examined 13 measurement models and concluded that two were especially good and equally effective at demonstrating the interrelatedness of the WAIS-IV and WMS-IV.

The first model identified five second-order factors, along with a hierarchical first-order g factor; the second model extracted seven factors, but no g dimension (Holdnack et al., 2011). Because the empirical and conceptual support for both models is equally strong, examiners may choose to interpret whichever model is most consistent with their personal, professional, and theoretical orientation. As discussed in the following paragraphs, we favor the seven-factor solution.

The five factors in the hierarchical model were named Verbal Comprehension, Perceptual Reasoning, Processing Speed, Working Memory, and Memory. The first four correspond to the WAIS-IV Indexes; the fifth is composed of the four WMS-IV delayed recall subtests, tasks that measure primarily Long-Term Retrieval (Glr) and Visual Processing (Gv) from CHC theory (D. Flanagan, personal communication, October 21, 2011). The Working Memory factor was the only one to incorporate subtests from both batteries—Arithmetic, Digit Span, Spatial Addition, and Symbol Span.

The loadings on the first-order g factor ranged from .72 for Processing Speed to .94 for Working Memory (WM). This finding is consistent with the claim that:

WM comprises the functions of focusing attention, conscious rehearsal, and transformation and mental manipulation of information, while *g* reflects the component variance that is common to all tests of ability. The centrality of WM in individual differences in information processing leads to some cognitive theorists to equate it with *g*. (Colom, Rebollo, Palacios, Juan-Espinosa, & Kyllonen, 2004, p. 277)

The alternative joint CFA recommended by Holdnack et al. (2011) produced seven factors, four composed only of WAIS-IV subtests (Verbal Comprehension, Perceptual Reasoning, Processing Speed, and Auditory Working Memory) and three composed solely of WMS-IV subtests (Visual Working Memory, Auditory Memory, and Visual Memory). From CHC theory: (a) Auditory Memory measures *Gl* (specifically, the narrow abilities of meaningful memory and associative memory); (b) Visual Memory measures *Gv* (Visual Memory) with a touch of *Gsm* (Working Memory); and (c) Visual Working Memory is an amalgam of *Gv* (Visual Memory) and *Gsm* (Working Memory and Memory Span) (D. Flanagan, personal communication, October 21, 2011).

From the vantage point of CHC theory, the first model has the great disadvantage of merging *Gv* and *Gl* into a single factor, whereas the seven-factor structure does a better job of distinguishing *Gv* from *Gl*, especially when contrasting the Visual Memory and Auditory Memory factors. Furthermore, the seven-factor model also accords well with the cognitive neuroscience research and theory that underlies the WAIS-IV, including the distinction between Auditory Working Memory and Visual Working Memory, and it demonstrates the total separation of the constructs measured by the WAIS-IV and WMS-IV, providing strong empirical support for the common clinical and neuropsychological practice of administering both test batteries as part of a comprehensive assessment. We endorse that practice and recommend that clinicians integrate the scores yielded by the two instruments by (a) internalizing the joint seven-factor structure of the WAIS-IV and WMS-IV (Holdnack et al., 2011); (b) performing some or all of the two dozen or so comparisons between the two instruments (Drozdzick et al., 2011, pp. 171–172); and (c) studying the growing body of literature on joint interpretation of profiles on the two Wechsler tests (Drozdzick et al., 2011; Frazier, 2011; Holdnack & Drozdzick, 2010).

The case report of Jim W., a 64-year-old man referred for possible dementia (see Chapter 10), demonstrates the integration of data from the WAIS-IV and WMS-IV and provides a good illustration of the dynamic degree to which these two instruments complement each other in the diagnostic process.

WAIS-IV's Relationship with the WAIS-III

The relationship between the WAIS-IV and its predecessor, the WAIS-III, was examined in a sample of 240 adults aged 16 to 88 (Psychological Corporation, 2008). Each test was administered in a counterbalanced order with a 1- to 23-week interval (mean = 5 weeks) between the testings. The overall correlation coefficients showed that the Full Scale IQs for the WAIS-III and WAIS-IV were the most highly related ($r = .94$) of the global scales, followed by the Verbal Comprehension Indexes ($r = .91$), Working Memory Indexes ($r = .87$), Processing Speed Indexes ($r = .86$), and the Perceptual Organization/Reasoning Indexes ($r = .84$). Thus, despite the substantial changes from the WAIS-III to the WAIS-IV in the composition of the Full Scale (see Rapid Reference 1.3), the extremely high coefficient of .94 indicates that the construct measured by Wechsler's Full Scale has not changed at all.

As shown in Table 1.1, the average WAIS-IV Full Scale IQ was 2.9 points lower than the WAIS-III Full Scale IQ, which is the same difference the WAIS-III FSIQ was from the WAIS-R FSIQ. The difference between the two instruments on both the Working Memory Index and the Processing Speed Index is negligible (0.7 points for both), but it is more substantial for the Verbal Comprehension Index (4.3 points)

Table 1.1. Changes in Scores From the WAIS-III to the WAIS-IV

Scale	WAIS-III		WAIS-IV		WAIS-III—WAIS-IV Standard Score	WAIS-III— WAIS-IV
	Mean ^a	SD	Mean ^a	SD	Difference	Correlation ^b
VCI	104.4	15.5	100.1	14.9	4.3	0.91
PI or POI	103.7	15.3	100.3	15.5	3.4	0.84
WMI	100.0	14.5	99.3	13.7	0.7	0.78
PSI	100.8	17.2	100.1	14.9	0.7	0.86
FSIQ	102.9	15.0	100.0	15.2	2.9	0.94

^a The values in the Mean columns are the average of the means of the two administration orders.

^b The weighted average was obtained with Fisher's z transformation.

Note: Sample sizes ranged from 238 to 240. Correlations were computed separately for each order of administration in a counterbalanced design and corrected for the variability of the WAIS-III standardization sample (Guilford & Fruchter, 1978).

Source: Data are adapted from Table 5.5 of the *WAIS-IV Technical and Interpretive Manual* (Wechsler, 2008).

and the Perceptual Organization/Reasoning Index (3.4 points). These differences are entirely consistent with the well-known Flynn Effect (Flynn, 1987, 2007, 2009a, 2009b; Flynn & Weiss, 2007) and indicate that a person's standard scores on an old test, with outdated norms (e.g., the WAIS-III), will tend to be spuriously high. The WAIS-IV will yield scores that are a little lower than the WAIS-III, especially on the FSIQ, VCI, and PRI, but these lower scores present a more accurate estimate of the person's intellectual abilities because they are derived from contemporary standards (i.e., the most recent norms groups).

The Flynn Effect and Capital Punishment

Overall, the Flynn effect (FE) has shown that, on average, American children and adults have increased their scores on intelligence tests at the rate of 3 points per decade between the 1930s and 1990s, with gains of 5 to 8 points per decade occurring for other developed nations, such as France, The Netherlands, and Japan (Flynn, 2007, 2009b; Kaufman & Lichtenberger, 2006). The mean FSIQ difference in the WAIS-III/WAIS-IV study, which translates to a gain of 2.7 IQ points per decade (Zhou, Gregoire, & Zhu, 2010), confirms the maintenance of the FE in the United States into the first decade of the 21st century. Post-2000 data from Norway and Denmark suggest that the FE has stopped occurring in those countries and that there may even be a reverse FE (i.e., decline in IQ) taking place, especially in Denmark (Sundet, Barlaug, & Torjussen, 2004; Teasdale & Owen, 2005, 2008). However, those studies are limited by the fact that they are based solely on data from 18- to 19-year-old males; generalizations are made to entire *countries* even though the studies did not include females, children, or adults age 20 or above (Kaufman, 2010c).

Within the United States, the debate has escalated well beyond research laboratories and clinical practice (e.g., Should examiners adjust a child's IQ for the FE when diagnosing children for special education?). Now the debate has entered the arena of litigation, with a wide array of ongoing capital punishment cases in progress throughout the nation. Kaufman and Weiss (2010a) frame the current issues in their introduction to a special issue of the *Journal of Psychoeducational Assessment* devoted to the FE:

Researchers differ on why the FE occurs. Some claim that genetics is the key variable (Rodgers & Wanstrom, 2007), though most stress environmental factors such as nutrition (Colom, Lluís-Font, & Andres-Pueyo, 2005), education (Teasdale & Owen, 2005), or improvement in public health (Steen, 2009). . . . [Regardless of causality], the dispute about the

scientific validity of the FE has entered the U.S. courtrooms in a big way as a burgeoning array of law cases asks whether the FE should be considered when sentencing low-functioning criminals convicted of a capital crime. Ever since the Supreme Court's decision in *Atkins v. Virginia* (2002), which stipulated that a criminal who is mentally retarded cannot be executed, whether or not to adjust IQs for the FE (i.e., by subtracting 3 points for each decade that the norms are out of date) has literally been a matter of life or death for some individuals. If a convicted criminal in a capital punishment case earned a global IQ of 73 on a test with 20-year-old norms, should that IQ be adjusted by 6 points to account for their datedness? Is the best estimate of the person's mental functioning 73 or 67? Of course, standard errors of measurement and adaptive behavior enter the equation as well, but the questions that arise from the FE are intriguing. (pp. 379–380)

The courts remain split on whether a criminal's IQ should be adjusted for the FE, with verdicts differing from state to state and from courtroom to courtroom within states. Those who argue for the FE adjustment claim that an IQ obtained on out-of-date norms is spuriously high; the downward adjustment often places a person's IQ in the range associated with Mental Retardation (Intellectual Disability), and the criminal is exempt from execution. Those who are unconvinced of the scientific validity of the FE would *not* adjust a person's IQ of, say, 75, making that criminal eligible for the death penalty. Numerous professionals have served as expert witnesses in these "Adkins" cases, such as Jack Fletcher, Alan Kaufman, Kevin McGrew, Cecil Reynolds, and James Flynn himself. Some argue for the prosecution (e.g., Hagan, Drogin, & Guilmette, 2010), others for the defense (e.g., Fletcher, Steubing, & Hughes, 2010; Reynolds, Niland, Wright, & Rosenn, 2010).

The arguments persist and are not yet resolved. The editors of the special journal issue devoted to the FE disagree about the adjustment. Weiss (2010) believes that, "the ethical position of an expert witness providing testimony is not to argue either for or against FE adjustments but to inform the court about the extant research on the topic" (p. 491). In contrast, Kaufman (2010b) states:

I respect the diversity of opinion on the topic of capital punishment, and the statistical complexity that surrounds the FE, but I am firmly in the camp with Reynolds et al. (2010), Fletcher et al. (2010), and Flynn (2006, 2007, 2009b) that IQs obtained on outdated norms should be adjusted for the FE in capital punishment cases. (p. 500)

The arguments on both sides are complex and too detailed for this book; interested readers are referred to the diverse articles in the special issue on the FE

(Kaufman & Weiss, 2010b), which is not restricted to the issue of capital punishment but covers a wide variety of topics, including causality, clinical applications, and ethical considerations (e.g., Ceci & Kanaya, 2010; Flynn, 2010; Kaufman, 2010a; McGrew, 2010; Reynolds et al., 2010; Sternberg, 2010).

In the meantime, the FE continues to be a hot topic for ongoing research investigations, and readers are encouraged to seek out state-of-the-art studies as well. For example, Zhou and colleagues (Zhou et al., 2010; Zhou, Zhu, & Weiss, 2010) conducted a particularly innovative set of analyses using data from various versions of Wechsler's scales and other tests as well. They found that: (a) males had a larger FE than females; (b) individuals with IQs of 110 and above had a notably smaller FE than those with IQs below 110; (c) adults ages 55–90 had a substantially larger FE than virtually every age group between 3–4 and 40–54 years; (d) infants and toddlers showed a *reverse* FE on the Bayley Scales, as mean cognitive scores *dropped* 6.3 points per decade from the Bayley-II to the Bayley-III standardizations.

But the main thrust of the FE research is not found in the laboratory. As Cecil Reynolds (personal communication, March 13, 2010) noted: “Whether to apply the Flynn Correction is a dire matter with implications we seldom encounter in psychology.” Reynolds et al. (2010) commented on the societal responsibilities that fall, at least partly, into the hands of psychologists subsequent to *Atkins v. Virginia*: “The importance of understanding and assessing mental retardation in criminal defendants has become critical, indeed a true matter of life and death, in capital felony cases. . . . No one’s life should depend on when an IQ test was normed” (pp. 477, 480).

STANDARDIZATION AND PSYCHOMETRIC PROPERTIES OF THE WAIS-IV

The standardization sample for the WAIS-IV ($N = 2,200$) was selected according to 2005 U.S. Census data and was stratified according to age, sex, race/ethnicity, geographic region, and education level. Thirteen age groups were created from a large sample of adolescents and adults, with 100 to 200 subjects in each group between ages 16–17 and 85–90.

Reliability

The average split-half reliability for the FSIQ across the 13 age groups was strong, ranging from .97 to .98 (see Rapid Reference 1.7 for split-half and test-retest reliability for all scales and subtests) (The Psychological Corporation, 2008). The Factor Indexes had average reliability coefficients ranging from .90 for

Rapid Reference 1.7

Average WAIS-IV Reliability

Subtest/Composite Score	Split-Half Reliability	Test-Retest Reliability
Block Design	.87	.80
Similarities	.87	.87
Digit Span	.93	.83
Matrix Reasoning	.90	.74
Vocabulary	.94	.89
Arithmetic	.88	.83
Symbol Search	.81	.81
Visual Puzzles	.89	.74
Information	.93	.90
Coding	.86	.86
Letter-Number Sequencing	.88	.80
Figure Weights	.90	.77
Comprehension	.87	.86
Cancellation	.78	.78
Picture Completion	.84	.77
Verbal Comprehension Index	.96	.96
Perceptual Reasoning Index	.95	.87
Working Memory Index	.94	.88
Processing Speed Index	.90	.87
Full Scale IQ	.98	.96

Note: For Coding and Symbol Search, and the composite of these two (Processing Speed), only test-retest coefficients are reported because of the timed nature of the subtests.

Source: Data are from Tables 4.1 and 4.5 of the *WAIS-IV Technical and Interpretive Manual* (Psychological Corporation, 2008).

Processing Speed to .96 for Verbal Comprehension. Individual subtest reliabilities ranged from an average of .94 on Vocabulary to .78 on Cancellation; median values were .89 for the 10 core subtests and .87 for the 5 supplemental subtests. A subset of the standardization sample (298 adults) provided test-retest data, with an average of three weeks between testings. The results of the test-retest study showed similar reliability coefficients for the four age-group subsamples (16–29, 30–54, 55–69, and 70–90 years). Average stability coefficients across all ages were

.96 for the Full Scale IQ and Verbal Comprehension Index, .88 for the Working Memory Index, and .87 for both the Perceptual Reasoning and Processing Speed Index. The highest stability coefficient for the core subtests was .90 for Information, and the lowest was .74 for Matrix Reasoning and Visual Puzzles. Of the supplemental subtests, Comprehension had the highest stability coefficients, ranging from .86 for Comprehension to .77 for Figure Weights and Picture Completion.

Salthouse and Saklofske (2010, Table 8.4) reported reliability and stability of the WAIS-IV subtests for two broad age groups: 16–64 and 65–90. Most coefficients were within .02 of each other for the two groups. However, the following differences of .03 or greater were noted: (a) split-half reliability was *higher* for ages 65–90 than 16–64 on Information, Similarities, and Symbol Search, but was *lower* for Block Design; and (b) stability was *higher* for ages 65–90 than 16–64 on Information, Digit Span, and Picture Completion, but was lower on Matrix Reasoning and Visual Puzzles. The only truly notable differences were for the stability of Digit Span and Visual Puzzles. Digit Span had a stability coefficient of .84 for the elderly sample versus .74 for the younger group. The stability for Visual Puzzles was .72 for ages 16–64 versus an *unacceptably low value of .57 for the older sample*.

Loadings on the General Factor

General intelligence or general mental ability (Spearman, 1927) is denoted by *g*. The measurement of *g* may be done by several methods. Preliminary findings from Keith's WAIS-IV higher-order CFA (personal communications, January 30 and March 14, 2009), based on the average correlation matrix for ages 16 to 90 (Psychological Corporation, 2008, p. 62), provided the *g*-loadings reported here. These *g*-loadings are the Factor loadings for each WAIS-IV subtest on the second-order general Factor that was obtained from the CFA. Factor loadings of .70 or greater are usually considered “good” measures of *g*; loadings of .50 to .69 are deemed “fair” *g* loadings; and loadings below .50 are considered poor. Rapid Reference 1.8 contains data on how well each subtest loads on the *g* factor.

Contrary to previous Wechsler scales on which measures of verbal comprehension and expression tended to yield the highest *g* loadings, the best measures of *g* on the WAIS-IV were Arithmetic and two Perceptual Reasoning tasks. Among the Verbal Comprehension subtests, only Vocabulary emerged as a good measure of *g*. The traditionally good measures, such as Comprehension, Information, and Similarities, were only fair measures, loading in the mid- to high .60s. Not surprisingly, the Processing Speed subtests were the weakest measures of *g*, but only Cancellation, with a dismal loading of .38, qualifies as a poor measure of *g*.

Rapid Reference 1.8

WAIS-IV Subtests as Measures of General Ability (*g*)

	<i>g</i> loading	Strength as a measure of <i>g</i>
Arithmetic	.78	Good
Figure Weights	.77	Good
Matrix Reasoning	.73	Good
Vocabulary	.72	Good
Digit Span	.69	Fair
Block Design	.68	Fair
Comprehension	.68	Fair
Similarities	.68	Fair
Visual Puzzles	.66	Fair
Letter-Number Sequencing	.66	Fair
Information	.65	Fair
Picture Completion	.57	Fair
Coding	.55	Fair
Symbol Search	.54	Fair
Cancellation	.38	Poor

Source: T. Keith (personal communication, January 30, 2009).

Salthouse and Saklofske (2010, Figure 8.2) reported *g* loadings for the four WAIS-IV Indexes separately for two broad age groups: 16–64 and 65–90. These loadings are shown here:

Age Group	VCI	PRI	WMI	PSI
16–64	.80	.89	.82	.71
65–90	.81	.89	.95	.79

The PRI was the best measure of *g* for ages 16–64, but WMI was the best measure of *g* for elderly individuals. The PSI was the lowest for both age groups but was, nonetheless, higher for the older than the younger sample. However, Salthouse and Saklofske (2010) did not believe that the age-related fluctuations were meaningful; they concluded that the coefficients were “very similar, which suggests nearly

equivalent composition of the higher-order cognitive ability factor at different ages” (p. 233). As was mentioned in the discussion of the joint CFA of WAIS-IV and WMS-IV, the strong *g* loadings for Working Memory are consistent with a large body of cognitive research on the topic (Colom et al., 2004).

The concept of general intelligence is one whose usefulness has been debated in the intelligence literature. Interestingly, Horn (1989) and Carroll (1993) were at the opposite poles of this debate, although their theories were merged to form CHC theory. Horn was a devout anti-*g* theorist, whereas Carroll had great respect for *g* and considered general ability to be Stratum III of his theory of intelligence. Because of their disagreements about the *g* construct, CHC theory focuses on Broad Abilities (Stratum II) and Narrow Abilities (Stratum I) and rarely addresses the role of *g* (McGrew, 2005).

From our perspective, *g* pertains to a practical, clinical construct that corresponds to FSIQ and, therefore, provides an overview of each person’s diverse abilities. There is also evidence that the *g* that underlies tests of intellectual ability is closely related to the *g* that underlies tests of academic achievement (correlations of .77–.94 for ages 4–5 to 16–19) when intelligence and achievement are assessed by the Kaufman and Woodcock test batteries (Kaufman, Reynolds, Liu, Kaufman, & McGrew, 2012).

But we do not interpret it as a theoretical construct. Other theorists have argued otherwise (Carroll, 1993; Jensen, 1998; Spearman, 1904); even Wechsler² (1974) was a strong believer in *g*, maintaining that “[i]ntelligence is the overall capacity of individuals to understand and cope with the world around them” (p. 5). We believe that a subtest with a strong *g* loading should not be interpreted as one that is *the* representation of an individual’s overall level of cognitive ability. Rather, as discussed in Chapters 4 and 5 on interpretation, a cognitive test assesses diverse cognitive abilities, all of which need to be understood. The person’s pattern of strengths and weaknesses on the four Indexes or five Factors is far more important to interpret than FSIQ. The *g* loadings do represent how well psychometrically the subtests hang together as a whole but do not reflect a theoretical construct that underlies human intellect. The *g* loadings do offer aids to clinical interpretation by providing expectancies. For example, Arithmetic’s high *g* loading and strong loading on the fluid reasoning Factor in Keith’s CFA lead us to expect that a person will score about as well on the Arithmetic subtest as he or she scored on FSIQ and PRI. If, for example, the person scored much lower on Arithmetic than on FSIQ and PRI, that is contrary to expectations, and we would seek an explanation, such as distractibility, anxiety, poor working memory, or poor ability

2. Wechsler’s (1974) quote has been modified to avoid sexist language but is otherwise verbatim.

to manipulate numbers. By contrast, an extremely high or low score on Cancellation is anticipated and would not cause us to think twice about it.

ETHNIC DIFFERENCES IN IQ

Differences Between Whites and African Americans

The difference of about one standard deviation in the IQs earned by Whites and African Americans, identified for numerous samples with a wide variety of tests (Hauser, 1998; Lichtenberger, Broadbooks, & Kaufman, 2000; Reynolds & Lowe, 2009), is similar to the overall findings on the WAIS-III (Heaton, Taylor, & Manly, 2003; Kaufman & Lichtenberger, 2006) and the WAIS-IV. For the WAIS-IV standardization sample ages 16–90 years, unadjusted mean Full Scale IQs of 103.2 and 88.7 were obtained for Whites ($n = 1,540$) and African Americans ($n = 260$), respectively, yielding a difference of 14.5 points. Largest Index differences occurred in fluid and visual-spatial abilities ($PIR = 14.5$), with smaller but notable 10- to 12-point differences emerging for verbal abilities ($VCI = 11.8$), working memory ($WMI = 10.6$), and processing speed ($PSI = 10.0$) (Weiss et al., 2010, Table 4.3).

However, these global differences do not come close to telling the whole story. Weiss et al. (2010) conducted an array of analyses, often ingenious and always state-of-the-art in terms of psychometrics, in an attempt to better understand ethnic difference on the WAIS-IV within a societal context (see Rapid Reference 1.9 for the key findings). In that respect, the WAIS-IV research conducted by Weiss and colleagues mirrors the important series of WISC-IV investigations that Weiss et al.

Rapid Reference 1.9

Key Results in WAIS-IV Analyses on Differences Between Whites and African Americans (Weiss et al., 2010)

- Full Scale IQ differences varied as a direct function of birth cohort. White–African American differences were 19 points for those born between 1917 and 1942; 17 points for birth cohorts 1943–1962; 13 points for 1963–1987; and 10 points for 1988–1991 (Weiss et al., 2010, p. 123). Thus, race differences were almost twice as large for ages 65–90 years as for ages 16–19. Age trends were also noted on the WISC-IV, where race differences averaged 12 points for adolescents and 6 points for children (Prifitera, Saklofske, Weiss, & Rolfus, 2005). Nisbett (2009) reported that race differences in IQ for 12-year-olds had dropped by nearly 40% over the past three decades.

(continued)

- The IQ difference of about 15 points for adults ages 20–90 was reduced to 11 points when controlling for a variety of mediating variables—education, occupation, income, region, and gender (Weiss et al., 2010, p. 125). These findings conform to previous results with a variety of intelligence tests (Reynolds & Lowe, 2009), including the WAIS-III (Heaton et al., 2003; Kaufman & Lichtenberger, 2006).
- For adults ages 20–90, the variable of race accounted for 15% of the variance in multiple regression analysis, whereas educational attainment accounted for almost twice as much variance (29%). When occupation, income, region, and gender are added to the mix, that value increases to 35% (Weiss et al., 2010, p. 124).

(2006) reported in a groundbreaking chapter; indeed, Flanagan and Kaufman (2009) recommended reading that 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” (p. 49).

Differences Between Whites and Hispanics

For the WAIS-IV standardization sample ages 16–90 years, Weiss et al. (2010) also provided differences in scores earned by Whites and Hispanics. Unadjusted mean Full Scale IQs of 103.2 and 91.6 were obtained for Whites ($n = 1,540$) and Hispanics ($n = 289$), respectively, yielding a difference of 11.6 points. Smallest Index differences occurred in processing speed ($PSI = 6.1$) and fluid/visual-spatial abilities ($PRI = 8.8$) with the largest differences occurring on language tasks: verbal abilities ($VCI = 11.5$) and working memory ($WMI = 10.9$) (Weiss et al., 2010, Table 4.3). These Index differences mirror almost exactly data obtained on the WAIS-III (Heaton et al., 2003; Kaufman & Lichtenberger, 2006).

As with White–African American differences, Weiss et al. (2010) observed the mediating effects of age, education, and other background variables in their analyses of White-Hispanic discrepancies on the WAIS-IV, as shown in Rapid Reference 1.10.

Rapid Reference 1.10

Key Results in WAIS-IV Analyses on Differences Between Whites and Hispanics (Weiss et al., 2010)

- Similar to the findings for Whites vs. African Americans, Full Scale IQ differences for Whites vs. Hispanics varied as an almost direct function of birth cohort. Differences were 18 points for those born between 1917 and 1942;

14 points for birth cohorts 1943–1977; 7 points for 1978–1987; and 9 points for 1988–1991 (Weiss et al., 2010, p. 123). Thus, once again, ethnic differences were almost twice as large for ages 65–90 years as for ages 16–19. These results conform to WISC-IV findings, where ethnic differences averaged 8 points for adolescents and 1 point for children (Prifitera et al., 2005).

- The IQ difference of about 12 points for adults at ages 20–90 was reduced to 6.5 points when controlling for a variety of mediating variables—education, occupation, income, region, and gender (Weiss et al., 2010, p. 125), consistent with previous analyses of adjusted White–Hispanic differences on the WAIS-III and other cognitive tests (Heaton et al., 2003; Kaufman & Lichtenberger, 2006).
- For adults ages 20–90, the variable of ethnicity (White vs. Hispanic) accounted for 11% of the variance in multiple regression analysis, whereas educational attainment accounted for about three times as much variance (31%). When occupation, income, region, and gender are added to the mix, that value increases to 37% (Weiss et al., 2010, p. 124).

OVERVIEW OF ETHNIC DIFFERENCES ON THE WAIS-IV

The results of the analyses by Weiss and his colleagues on the WISC-IV (Weiss et al., 2006) and WAIS-IV (Weiss et al., 2010) make it abundantly clear that socioeconomic status and an array of other background, behavioral, and personal variables impact a person's IQ and profile of test scores far more than the variable of ethnicity alone, and that these variables mediate the role played by ethnicity in affecting a person's IQ. Weiss and colleagues urge the interpretation of ethnic differences within a multifaceted societal context. We encourage clinicians and researchers to read these two powerful chapters on Wechsler's scales to understand the interactive and complex roles played by ethnicity, education, diverse indicators of socioeconomic status, personal beliefs, and developmental variables in shaping a person's cognitive test profile. Consult also Suzuki, Short, and Lee (2011) and *IQ Testing 101* (Kaufman, 2009) for a thorough discussion of the genetic and environmental factors that interactively affect IQ test performance.

Ultimately, we agree with Weiss et al. (2010), who state that,

racial/ethnic differences are likely to be proxies for a multitude of other variables that we are just beginning to identify and study. . . . We suggest that future researchers go beyond these easily collected proxy variables (i.e., race/ethnicity, and SES) and directly study the factors that are related to the development and maintenance of cognitive abilities both within and across culturally and linguistically diverse groups. (pp. 135–136)

COMPREHENSIVE REFERENCES ON TEST

The *WAIS-IV Administrative and Scoring Manual* (Wechsler, 2008) and the *WAIS-IV Technical and Interpretive Manual* (Psychological Corporation, 2008) currently provide the most detailed information about the WAIS-IV. These manuals review the development of the test, descriptions of each of the subtests and scales, standardization, reliability, and validity. In addition, a particularly valuable WAIS-IV reference is a book edited by professionals who were intimately involved with the development and standardization of the test: *WAIS-IV—Clinical Use and Interpretation* (Weiss, Saklofske, Coalson, & Raiford, 2010b). *Assessing Adolescent and Adult Intelligence, Third Edition* (Kaufman & Lichtenberger, 2006) provides an excellent review of the research on the WAIS, WAIS-R, and WAIS-III, much of which is still pertinent for the WAIS-IV. Rapid Reference 1.11 provides basic information on the WAIS-IV and its publisher.

Our second edition of *Essentials of WAIS-IV Assessment*, the edited book by Weiss et al. (2010b), Sattler and Ryan's (2009) *Assessment with the WAIS-IV* (2009), a recent chapter by Drozdick et al. (2012), and a special section of the June 2011 issue of

Rapid Reference 1.11

Wechsler Adult Intelligence Scale—Fourth Edition

Author: David Wechsler

Publication Date: 2008

What the Test Measures: verbal comprehension, perceptual reasoning, working memory, processing speed, and general intelligence

Age Range 16–90 years

Administration Time: 10 core subtests to obtain 4 indexes = 65–90 minutes;
15 core and supplemental subtests = 85–114 minutes

Qualification of Examiners: Graduate- or professional-level training in psychological assessment

Publisher: Pearson

19500 Bulverde Road

San Antonio, TX 78259

Customer Service: (800) 211-8378

<http://pearsonassess.com>

Price: WAIS-IV Basic Kit Includes *Administration and Scoring Manual*, *Technical Manual*, 2 Stimulus Books, 25 Record Forms, 25 Response Booklet 1, 25 Response Booklet 2, Symbol Search Scoring Key, Coding Scoring Key, Cancellation Scoring Templates in a box. ISBN: 015-8980-808. \$1,120.00 (in box); \$1,190.00 (in hard- or soft-sided case).

Assessment (Frazier, 2011) provide the most authoritative sources for administering, scoring, interpreting, and applying WAIS-IV test profiles.



TEST YOURSELF



1. **Many of the tasks that David Wechsler used in his WAIS, WAIS-R, WAIS-III, and WAIS-IV were adapted from what sources?**
2. **Updating the WAIS-IV's theoretical foundations was achieved by considering the following theoretical constructs EXCEPT**
 - a. Fluid reasoning
 - b. Working memory
 - c. Processing speed
 - d. Phonological processing
3. **What was the major structural change implemented from the WAIS-III to the WAIS-IV?**
4. **Which of the following WAIS-IV subtests is a CORE subtest that is used to compute FSIQ?**
 - a. Visual Puzzles
 - b. Letter-Number Sequencing
 - c. Picture Completion
 - d. Comprehension
 - e. Figure Weights
5. **Which subtest is NOT new to the WAIS-IV?**
 - a. Visual Puzzles
 - b. Figure Weights
 - c. Cancellation
 - d. Symbol Search
6. **Which WAIS-IV subtest does NOT offer Process scores?**
 - a. Digit Span
 - b. Visual Puzzles
 - c. Block Design
 - d. Letter-Number Sequencing
7. **The results of Keith's confirmatory factor analysis that supported a Five-Factor CHC model for ages 16–69 showed three WAIS-IV subtests to load highly on the fluid reasoning (Gf) factor. These subtests are Figure Weights, Matrix Reasoning, and**
 - a. Block Design
 - b. Picture Completion

(continued)

- c. Letter-Number Sequencing
- d. Similarities
- e. Arithmetic

8. The new Five-Factor CHC model of the WAIS-IV for ages 70-90

- a. Has strong empirical support
- b. Treats the three sections of Digit Span as separate variables
- c. Measures the same theoretical constructs from CHC theory as the Keith model
- d. Is an alternate model for ages 16–69 when examiners do not administer supplemental subtests
- e. All of the above

9. Which index includes the subtests with the lowest loadings on the general (g) factor?

- a. Verbal Comprehension
- b. Perceptual Reasoning
- c. Working Memory
- d. Processing Speed

10. What term refers to the phenomenon that IQ test norms in the United States get out of date at the rate of about 3 points per decade?

- a. Flynn Effect
- b. CFA
- c. CHC
- d. Horn Effect
- e. g theory

11. TRUE OR FALSE? Analyses of ethnic differences on the WAIS-IV have shown that ethnicity accounts for more variance in IQ than socioeconomic status.

Answers: 1. Army Alpha, Army Beta, Army Performance Scale Examination, and Stanford-Binet; 2. d; 3. Removal of the VIQ and PIQ; 4. a; 5. d; 6. b; 7. e; 8. e; 9. d; 10. a; 11. false.