CHAPTER 14

The Woodcock-Johnson Battery—Third Edition (WJ III)

KEVIN S. MCGREW
University of Minnesota

RICHARD WOODCOCK
Measurement/Learning/Consultants

LAURIE FORD
University of South Carolina

HISTORY AND EVOLUTION OF THE WJ III

Original Woodcock-Johnson (WJ)

The Woodcock-Johnson—Third Edition (WJ III; Woodcock, McGrew, & Mather, 2001) is the most recent revision of a battery of tests first published in 1977 as the Woodcock-Johnson Psycho-Educational Battery (WJ; Woodcock & Johnson, 1977). The WJ was the first comprehensive co-normed battery of tests of cognitive abilities, achievement, and interests. The WJ was normed on 4,732 subjects from ages three through 80+ and provided both age- and grade-based norm scores.

The authors express their appreciation for the constructive received feedback from Dr. Fred Schrank and Dr. Dawn Flanagan. We also wish to express our thanks to Jeff Evans and Julie Evans for their assistance in the preparation of the tables and the editing of the text.

The WJ Tests of Cognitive Ability (COG) included 12 individual tests designed to represent a sampling of cognitive abilities ordered from lower- to higher-level cognitive processing (Woodcock, 1978). The 12 COG tests were combined into a differentially weighted estimate of *g* or general intelligence (Broad Cognitive Ability cluster), the first differentially weighted *g*-score provided with an individually administered battery of cognitive tests. The WJ COG included four cognitive ability factors (Reasoning, Verbal Ability, Memory, Perceptual Speed) that were empirically derived from a series of exploratory factor and cluster analyses. As such, the WJ COG factors were not based on any particular structure-of-intellect model of intelligence available at that time.

Four Scholastic Aptitude (SAPT) clusters were also included in the WJ. The SAPT clusters were the cornerstone of one component of the WJ's featured *pragmatic decision-making model*, a model that utilized a decision-based test design strategy (McGrew, 1986; Woodcock, 1984a, 1984b). Briefly, the WJ was designed to provide the data necessary for practitioners needing to make important psychoeducational decisions (McGrew, 1986). Each SAPT cluster was composed of the best differentially weighted combination of four cognitive tests that predicted a respective achievement domain. The SAPT clusters were pivotal in the quantification of aptitude/achievement discrepancies via the aptitude/achievement Relative Performance Indexes (RPIs). Within-cognitive and within-achievement discrepancies comprised the intra-cognitive and intra-achievement components of the pragmatic decision-making model, and were operationalized via comparison of test/cluster confidence bands.

The WJ Tests of Achievement (ACH) were comprised of 10 tests and resultant clusters organized around the curricular areas reading, mathematics, written language, and knowledge. Broad cluster scores were provided in each of these achievement domains. The Tests of Interest Level provided scores for the measurement of Scholastic (Reading, Mathematics, and Written Language) and Nonscholastic (Physical and Social) Interests. A Spanish version (Woodcock, 1982) included 10 of the cognitive tests and 7 of the achievement tests. In 1984 the battery was linked to the Scales of Independent Behavior (Bruininks, Woodcock, Hill, & Weatherman, 1985) as part of the WJ/SIB Assessment System.

Woodcock-Johnson—Revised (WJ-R)

In 1989 a revised and re-standardized WJ-R battery was published (Woodcock & Johnson, 1989a, 1996b). The primary goal of the WJ-R was to expand the diagnostic capabilities of the test and to complement the pragmatic decision-making model with a validated structure-of-intellect model (viz., Cattell-Horn *Gf–Gc* theory) (McGrew, 1994; McGrew et al., 1991). The WJ Tests of Interest were dropped and the WJ-R was divided into two main batteries: Tests of Cognitive Ability (WJ-R COG) and Tests of Achievement (WJ-R ACH), each with standard and supplemental battery. The complete norm sample was comprised of 6,359 subjects from age two through 90+. For the first time, separate college/university norms were provided.

The WJ-COG was enhanced by the addition of new tests and clusters grounded in the Cattell-Horn *Gf–Gc* theory of cognitive abilities. The WJ-R COG was the first of the major intelligence test batteries to utilize a *multiple intelligences* approach with several (seven) empirically supported cognitive ability factors. The WJ-R ACH was extended from 10 to 14 tests with several new tests of reading, written language, and mathematics added. Broad achievement clusters were supplemented with subdomain clusters (e.g., Basic Reading Skills and Reading Comprehension) in reading, math, and written language. Parallel, alternative forms (Forms A & B) were also introduced to the Tests of Achievement. A complete Spanish version of both the WJ-R COG and WJ-R ACH were made available in 1996 (Woodcock & Munoz-Sandoval, 1996a, 1996b).

A number of the WJ interpretation features were refined and other new features added to the WJ-R. In response to concerns about the complexity of hand-scoring the WJ, the differential weighting of the SAPT and the BCA clusters were replaced with equally weighted clusters (McGrew, 1994). A major improvement was the development of actual norm-based scores for evaluating the intra-cognitive and intra-achievement discrepancies. Finally, an extensive technical manual (McGrew, Werder, & Woodcock, 1991) provided detailed information that supported the construct validity of the *Gf–Gc* structure-of-intellect model for the battery.

The current WJ III has benefited from 23 years of development and ongoing revision. Detailed historical information regarding the development and evolution of the WJ to the WJ-R, as well as a synthesis of related research literature and independent reviews, can be found in a number of sources (Hessler, 1982, 1993; McGrew, 1986, 1994; McGrew et al., 1991). The presentation of an overview of the latest version of the WJ battery (WJ III), with a particular emphasis on its characteristics and use with adolescents and adults, is the purpose of this chapter.

THEORETICAL FOUNDATIONS OF THE WJ III

The validity of psychological tests hinges on the degree to which empirical evidence and theory support the use and interpretation of the test scores. Over time, a number of prominent measurement experts (Benson, 1998; Cronbach, 1971; Cronbach & Meehl, 1955; Loevinger, 1957; Messick, 1989; Nunnally, 1978) have, in one way or another, outlined a three- or four-source validity framework, a framework that now serves as the foundation for most of the validity standards in the joint APA, AERA, NCME *Standards on Educational and Psychological Testing* (1999). As per the joint *Test Standards*, substantive, internal, and external validity evidence is the cornerstone of a strong program of construct validity.

The *substantive* stage of construct validity defines the *theoretical* and *measurement* (empirical) domains of the theoretical constructs. In the current context, the substantive question to be answered is "*How should intelligence be defined and operationally measured by the WJ III?*" According to Benson (1998), a strong psychological theory maximizes substantive (content) validity vis-à-vis the specification of a well-bounded construct domain, which, in turn, guides the development of measures in the empirical domain. The Cattell-Horn-Carroll (CHC) Theory of Cognitive Abilities served this function for the WJ III. The theory and its use as the WJ III test-design blueprint is described here.

The Cattell-Horn-Carroll (CHC) Theory of Cognitive Abilities

CHC theory evolved from the psychometric tradition of defining intelligence (Flanagan et al., 2000; McGrew & Flanagan, 1998). The psychometric approach is the oldest and most well-established approach to describing the structure of intelligence, dating back to Galton's attempt, in the late 1800s, to measure intelligence with psychophysical measures. Correlation and factor-analytic methods are employed typically to analyze scores from psychological tests in an attempt to objectively identify the primary dimensions that form the structure of individual differences in cognitive ability.

CHC theory is a *strong* psychological theory, as it represents one of the best examples of cumulative science in applied psychology. The original roots of CHC theory can be traced to Spearman's (1904, 1927) presentation of the general or *g*-factor theory of intelligence, a development some consider the formal birth of the psychometric research tradition. The fundamental premise of Spearman's theory is that a single *g* or general intelligence ability accounts for the performance of individuals on most tasks of cognitive ability. Spearman's work was followed by decades of correlation and factor-analytic investigations by researchers who were committed to "slicing and dicing" the construct of intelligence into a taxonomy of cognitive abilities.

One of these researchers, Raymond Cattell, diverged from Spearman's single *g*-factor model to propose the existence of two general types of intelligence: fluid intelligence (*Gf*) and crystallized intelligence (*Gc*) (Cattell, 1941). *Fluid Intelligence* (*Gf*) is influenced by both biological and neurological factors and incidental learning through interaction with the environment and includes inductive and deductive reasoning as hallmark *Gf* indicators (Taylor, 1994). In contrast, *Crystallized Intelligence* (*Gc*) is comprised of abilities that reflect the influences of acculturation (viz., verbal–conceptual knowledge; Taylor, 1994; Gustafsson, 1994). Over the next 30 to 40 years, a variety of individual scholars (Carroll & Horn, 1981; Ekstrom, French, & Harmon, 1979; Horn, 1965, 1968, 1985, 1988, 1991; Thurstone, 1935, 1938; Thurstone & Thurstone, 1941) applied factor-analytic methods to a diverse array of ability measures in a wide array of samples. These efforts produced a number of empirically-based human cognitive ability taxonomies, the most notable being the *Well Replicated Common Factors (WERCOF)* (Ekstrom et al., 1979) and *Thurstone's Primary Mental Abilities (PMA)* model (1938). Collectively, these structure-seeking research activities indicated that human intellectual ability is best represented by a multitude of abilities that vary by degree of breadth or generality.

Cattell-Horn *Gf–Gc* Model

As reflected in Figure 14.1, by the early 1980s, John Horn, a student of Cattell's, articulated the relatively "complete" *Gf–Gc* model of intelligence that included eight broad abilities, which, in turn, subsumed the WERCOF and PMA abilities. Horn, like Cattell, continued to dismiss the notion of *g* and posited the broad abilities of fluid intelligence (*Gf*), crystallized intelligence (*Gc*), visual processing (*Gv*), auditory processing (*Ga*), short-term acquisition and retrieval (*SAR*, later referred to as short-term memory or *Gsm*), tertiary storage and retrieval (*TSR*, later referred to as long-term storage and retrieval or *Glr*), processing speed (*Gs*), and correct decision speed (*CDS*) (Horn, 1991).

In a series of publications between 1981 and 1991, Horn presented evidence for a broad quantitative knowledge ability (*Gq*) and sketched the faint outlines of a broad "English language usage" ability. The latter language factor was formally defined and described by Woodcock (1994) as a broad reading and writing ability (*Grw*). By 1994, contemporary Cattell-Horn *Gf–Gc* theory included 9 to 10 broad abilities. Although the Cattell-Horn *Gf–Gc* model was derived heavily from structural or factor-analytic research, support for the model comes from a number of divergent sources. Horn and Noll (1997) summarize the developmental, heritability, neurocognitive, and external/outcome validity evidence that also support the structure of the model.

Carroll's Three-Stratum Model

Aside from the specification of the 9 to 10 broad ability Cattell-Horn model, the most important influence in applied intelligence testing during the past decade has been Carroll's (1993) meta-analysis integration of the extant psychometric factor-analytic research (see Figure 14.1). Briefly, Carroll retrieved, sampled, and then factor-analyzed (via exploratory "let the data speak for themselves" methods) the reported correlation coefficients or raw data from 461 post-1925 data sets, including four sets drawn from the 1977 WJ norming sample. Carroll (1993, 1997) articulated a hierarchical *three-stratum theory*. Sixty-nine specific, or *narrow*, abilities were identified and classified as *Stratum I* abilities. The narrow abilities were subsumed under *broad* categories of cognitive ability (*Stratum II*), which he labeled Fluid Intelligence (*Gf*), Crystallized Intelligence (*Gc*), General Memory and Learning (*Gy*), Broad Visual Perception (*Gv*), Broad Auditory Perception (*Gu*), Broad Retrieval Ability (*Gr*), Broad Cognitive Speediness (*Gs*), and Processing Speed (*Gt*). At the apex of his model (*Stratum III*), Carroll identified a higher-order factor above the broad factors, which he interpreted as General Intelligence, or *g*.

Notwithstanding the differences between the Cattell-Horn and Carroll models, the two models are very similar. Carroll (1993) reached the same conclusion when, after reviewing all the major theories of intelligence, he described the *Gf–Gc* model as the best available model of the structure of human intellect:

> *The Cattell-Horn model, as summarized by Horn (1985, 1988), is a true hierarchical model covering all major domains of intellectual functioning. Numerous*

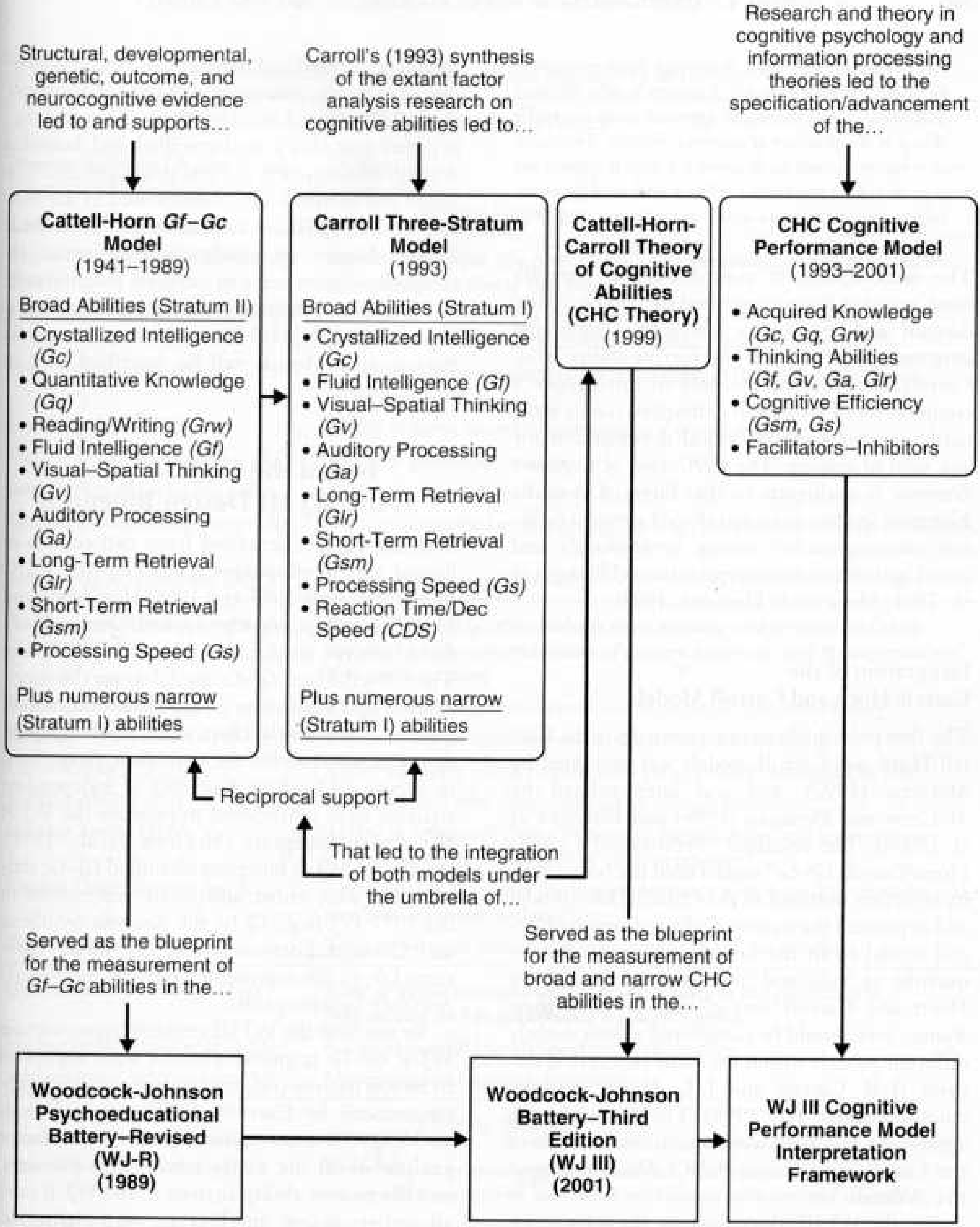

FIGURE 14.1

The evolution of Cattell-Horn-Carroll (CHC) theory and the Woodcock-Johnson III *details remain to be filled in through further research, but among available models it appears to offer the most well-founded and reasonable approach to an acceptable theory of the structure of cognitive abilities. The major reservation I would make about it is that it appears not to provide for a third-order* g *factor to account for correlations among the broad second-order factors.* (p. 62)

The most significant contribution of Carroll's work was that it circumscribed the *Gf–Gc* ability domain and imposed a systematic taxonomic structure, particularly at the narrow ability level. Carroll has provided the field of intelligence a common set of terms and definitions (i.e., a standard nomenclature), a critical development for any field of science. The *CHC Table of Cognitive Elements* is analogous to the Table of Periodic Elements in chemistry and should serve to facilitate communication among professionals and guard against test misinterpretations (Flanagan et al., 2001; McGrew & Flanagan, 1998).

Integration of the Cattell-Horn and Carroll Models

The first published attempt to integrate the Cattell-Horn and Carroll models was presented by McGrew (1997), and was later refined by McGrew and Flanagan (1998) and Flanagan et al. (2000). The resultant "Synthesized Cattell-Horn/Carroll *Gf–Gc*" model used the broad ability structure outlined in the Cattell-Horn model and organized the narrow abilities from the Carroll model to fit the broader structure. Subsequently, as indicated in Figure 14.1, in 1999 Horn and Carroll agreed that both of their frameworks should be considered as two slightly different models within the same theoretical domain (J. B. Carroll and J. L. Horn, personal communication, July 1999). The result was the agreement that both were operational models of the *Cattell-Horn-Carroll (CHC) Theory of Cognitive Abilities.*

For the WJ III, the answer to the substantive validity question (How should intelligence be defined and operationally measured?) was: "It should be defined and measured as per CHC theory." Therefore, consistent with the *Test Standards*, the WJ III is based on a strong psychological theory that provides a well-specified and bounded domain of constructs for test development. The broad and narrow CHC abilities used in the revision of the WJ III are summarized in Table 14.1. Finally, Figure 14.1 indicates that recent attempts have been made to integrate psychometric CHC theory with information-processing theory. Woodcock's (1993, 1997) Cognitive and Academic Performance Models will be described later in this chapter.

Use of the CHC Theory as the WJ III Design Blueprint

The WJ III has benefited from two rounds of formal test development guided by the CHC theory. Between 1985 and 1986, Woodcock and McGrew independently applied confirmatory factor-analytic methods to the 1977 WJ battery. The Cattell-Horn *Gf–Gc* model drove the specification and evaluation of the structural models tested. Concurrently, Horn and Carroll independently factor-analyzed the same data. As outlined in Figure 14.1, these four sets of independent analyses were synthesized to produce the WJ-R test design blueprint (McGrew et al., 1991). Briefly, the WJ-R blueprint identified *Gf–Gc* abilities that were either adequately represented in the 1977 WJ (e.g., *Gf* by the Analysis–Synthesis and Concept Formation tests) or underrepresented (e.g., *Glr* represented only by the Visual-Auditory Learning test).

By the time the WJ III revision began, certain WJ-R *Gf–Gc* cognitive clusters were suggested to be too narrow (i.e., inadequate construct representation) by Carroll's (1993) work, Woodcock's (1990) cross-battery confirmatory factor analysis of all the major intelligence batteries, and the narrow ability analysis of the WJ-R (and all other major intelligence test batteries) (McGrew, 1997; McGrew & Flanagan, 1998). For example, the WJ-R *Ga* cluster was com-

TABLE 14.1 Broad and narrow CHC abilities incorporated into the WJ III revision

Narrow Stratum I Name (Code)	CHC Broad Stratum II Ability Definition
	Acquired Knowledge
Verbal Comprehension Knowledge (Gc)	***Breadth and depth of a person's acquired knowledge of a culture and the effective application of this knowledge***
Language Development (LD)	General development, or the understanding of words, sentences, and paragraphs (not requiring reading) in spoken native language skills
Lexical Knowledge (VL)	Extent of vocabulary that can be understood in terms of correct word meanings
Listening Ability (LS)	Ability to listen to and comprehend oral communications
General (verbal) Information (KO)	Range of general knowledge
Information about Culture (K2)	Range of cultural knowledge (e.g., music, art)
General Science Information (K1)	Range of scientific knowledge (e.g., biology, physics, engineering, mechanics, electronics)
Geography Achievement (AS)	Range of geography knowledge
Reading and Written Language (Grw)	***Store of knowledge that includes basic reading and writing and skills required for the comprehension of written language and the expression of thought in writing***
Reading Decoding (RD)	Ability to recognize and decode words or pseudowords in reading
Reading Comprehension (RC)	Ability to comprehend connected discourse during reading
Cloze Ability (CZ)	Ability to supply words deleted from prose passages that must be read
Spelling Ability (SG)	Ability to spell (not clearly defined by existing research)
Writing Ability (WA)	Ability to write with clarity of thought, organization, and good sentence structure (not clearly defined by existing research)
English Usage Knowledge (EU)	Knowledge of writing in the English language with respect to capitalization, punctuation, usage, and spelling
Reading Speed (RS)	Time required to silently read a passage as quickly as possible
Quantitative Knowledge (Gq)	***Store of mathematical knowledge; the ability to use quantitative information and to manipulate numeric symbols***
Mathematical Knowledge (KM)	Range of general knowledge about mathematics
Mathematical Achievement (A3)	Measured mathematics achievement
	Thinking Abilities
Visual–Spatial Thinking (Gv)	***The ability to generate, perceive, analyze, synthesize, store, retrieve, manipulate, transform, and think with visual patterns and stimuli***
Visualization (Vz)	Ability to mentally manipulate objects or visual patterns and to "see" how they would appear under altered conditions

(Continues)

TABLE 14.1 (Continued)

Narrow Stratum I Name (Code)	CHC Broad Stratum II Ability Definition
	Thinking Abilities
Spatial Relations (SR)	Ability to rapidly perceive and manipulate visual patterns or to maintain orientation with respect to objects in space
Visual Memory (MV)	Ability to form and store a mental representation or image of a visual stimulus and then recognize or recall it later
Spatial Scanning (SS)	Ability to accurately and quickly survey a spatial field or pattern and identify a path through the visual field or pattern
Auditory Processing (Ga)	***Ability to perceive, analyze, and synthesize patterns among auditory stimuli and discriminate subtle nuances in patterns of sound and speech***
Phonetic Coding: Analysis (PC:A)	Ability to segment larger units of speech sounds into smaller units of speech sounds
Phonetic Coding: Synthesis (PC:S)	Ability to blend smaller units of speech together into larger units of speech
Speech Sound Discrimination (US)	Ability to detect differences in speech sounds under conditions of little distraction or distortion
Resistance to Auditory Stimulus Distortion (UR)	Ability to understand speech that has been distorted or masked in one or more ways
Long-Term Retrieval (Glr)	***Ability to store information in and fluently retrieve new or previously learned information from long-term memory***
Associative Memory (MA)	Ability to recall one part of a previously learned but unrelated pair of items when the other part is presented (i.e., paired-associative learning)
Meaningful Memory (MM)	Ability to recall a set of items in which there is a meaningful relation between items or the items create a meaningful story or connected discourse
Ideational Fluency (FI)	Ability to rapidly produce a series of ideas, words, or phrases related to a specific condition or object
Naming Facility (NA)	Ability to rapidly produce names for concepts
Fluid Reasoning (Gf)	***Mental operations involved when faced with novel tasks that cannot be performed automatically, including forming and recognizing concepts, drawing inferences, comprehending implications, problem solving, and extrapolating***
General Sequential Reasoning (RG)	Ability to start with stated rules, premises, or conditions and to engage in one or more steps to solve a problem
Induction (I)	Ability to discover the underlying characteristic (e.g., rule, concept, process, trend, class membership) that governs a problem or a set of materials
Quantitative Reasoning (RQ)	Ability to inductively and deductively reason with concepts involving mathematical relations and properties

TABLE 14.1 (Continued)

	CHC Broad Stratum II Ability
Narrow Stratum I Name (Code)	**Definition**
	Cognitive Efficiency
Processing Speed (Gs)	***Ability to fluently and automatically perform cognitive tasks, especially when under pressure to maintain focused attention and concentration***
Perceptual Speed (P)	Ability to rapidly search for and compare visual symbols presented side-by-side or separated in a visual field
Rate-of-Test-Taking (R9)	Ability to rapidly perform tests that are relatively easy or that require very simple decisions
Number Facility (N)	Ability to rapidly and accurately manipulate and deal with numbers, from elementary skills of counting and recognizing numbers to advanced skills of adding, subtracting, multiplying, and dividing numbers
Semantic Processing Speed (R4)	Speed of making a decision that requires some encoding and mental manipulation of stimulus content
Short-Term Memory (Gsm)	***Ability to apprehend and hold information in immediate awareness and to use it within a few seconds***
Memory Span (MS)	Ability to attend to and immediately recall temporally ordered elements in the correct order after a single presentation
Working Memory (MW)	Ability to hold information in mind for a short time while performing some operation on it; requires divided attention and the management of the limited capacity of short-term memory

NOTE: Narrow abilities not listed here can be found in Flanagan, McGrew, and Ortiz (2001).

From D. P. Flanagan, K. S. McGrew, and S. O. Ortiz, *The Wechsler intelligence scales and Gf-Gc theory: A contemporary approach to interpretation.* Copyright © 2000 by Allyn and Bacon. Adapted by permission.

prised of the Sound Blending and Incomplete Words tests. Both of these tests are indicators of a single narrow *Ga* ability, namely, Phonetic Coding (PC). Instead of representing a broad *Ga* domain, the WJ-R *Ga* cluster was in reality an index of a single narrow ability (PC) within *Ga*.

These analyses resulted in a revision that focused on selecting the tests that would improve the construct domain coverage of each WJ III CHC cognitive cluster. Special attention was paid to Carroll's hierarchical ability taxonomy to determine which narrow abilities should be measured. Each broad WJ III CHC cluster is now comprised of two qualitatively different narrow, or Stratum I, abilities. For example, the WJ III *Glr* cluster includes a measure of associative memory (Visual–Auditory Learning) and a measure of ideational fluency (Retrieval Fluency); the *Gv* cluster includes a measure of visualization (Spatial Relations) and a measure of visual memory (Picture Recognition). This blueprint for the WJ III had, as a primary goal, the assurance of adequate construction representation of the WJ III cognitive clusters and the minimization of construct-irrelevant variance in the tests. The process has resulted in a battery of individually administered tests that operationalizes the three-stratum CHC model of cognitive abilities.

DESCRIPTION AND ORGANIZATION OF THE WJ III TESTS AND CLUSTERS

Organization

The WJ III is a comprehensive collection of individually administered co-normed tests organized as two distinct test batteries. The Woodcock-Johnson Tests of Cognitive Abilities (WJ III COG) and the Woodcock-Johnson Tests of Achievement (WJ III ACH) are designed to measure a wide array of cognitive, oral language, and academic achievement abilities for individuals ages two through the geriatric population.

Each battery consists of two separate easel books organized into a Standard and Extended battery that can be used independently, together, or in conjunction with other tests. Tables 14.2 and 14.3 list the 20 Tests of Cognitive Abilities and the 22 Tests of Achievement. The cognitive tests are organized by both the broad CHC clusters and by three broader categories related to cognitive performance: verbal ability, thinking ability, and cognitive efficiency. The achievement tests are organized by curricular area (reading, mathematics, written language, and academic knowledge) and oral language and by clusters within these areas, with additional groupings for special purpose clusters. Examiners would rarely administer all tests but are, instead, encouraged to be *selective* in their testing. As examiners become more familiar and skilled with the tests in the battery, they can craft better referral-specific assessments by selecting different evaluation *tools* from their chest of tests. Different referral questions may require different tools for assessment.

TABLE 14.2 Organization of the WJ III Tests of Cognitive Abilities

Cognitive Performance Category/ CHC Ability Factor	Standard Battery Tests	Extended Battery Tests
Verbal Ability		
Comprehension–Knowledge (*Gc*)	Test 1: Verbal Comprehension	Test 11: General Information
Thinking Ability		
Long-Term Retrieval (*Glr*)	Test 2: Visual–Auditory Learning *Test 10: Visual–Auditory Learning-Delayed*	Test 12: Retrieval Fluency *Test 18: Rapid Picture Naming*
Visual–Spatial Thinking (*Gv*)	Test 3: Spatial Relations	Test 13: Picture Recognition *Test 19: Planning*
Auditory Processing (*Ga*)	Test 4: Sound Blending *Test 8: Incomplete Words*	Test 14: Auditory Attention
Fluid Reasoning (*Gf*)	Test 5: Concept Formation	Test 15: Analysis–Synthesis
Cognitive Efficiency		
Processing Speed (*Gs*)	Test 6: Visual Matching	Test 16: Decision Speed *Test 20: Pair Cancellation*
Short-Term Memory (*Gsm*)	Test 7: Numbers Reversed *Test 9: Auditory Working Memory*	Test 17: Memory for Words

NOTE: Italic font designates supplemental tests.

Copyright © 2001 by The Riverside Publishing Company. Adapted from the Woodcock-Johnson® III (WJ III™) by Richard W. Woodcock, Kevin S. McGrew and Nancy Mather, with permission of the publisher. All rights reserved.

TABLE 14.3 Organization of the WJ III Tests of Achievement

Curricular Area/CHC Ability Factor	*Standard Battery Tests*	*Extended Battery Tests*
Reading (*Grw*)		
Basic Reading Skills	Test 1: Letter–Word Identification	Test 13: Word Attack *Test 21: Sound Awareness*
Reading Fluency	Test 2: Reading Fluency	
Reading Comprehension	Test 9: Passage Comprehension	Test 17: Reading Vocabulary
Oral Language (*Gc*)		
Oral Expression	Test 3: Story Recall *Test 12: Story Recall-Delayed*	Test 14: Picture Vocabulary
Listening Comprehension	Test 4: Understanding Directions	Test 15: Oral Comprehension
Mathematics (*Gq*)		
Math Calculation Skills	Test 5: Calculation	
Math Fluency	Test 6: Math Fluency	
Math Reasoning	Test 10: Applied Problems	Test 18: Quantitative Concepts
Written Language (*Grw*)		
Basic Writing Skills	Test 7: Spelling	Test 16: Editing *Test 20: Spelling of Sounds* *Test 22: Punctuation & Capitalization*
Writing Fluency	Test 8: Writing Fluency	
Written Expression	Test 11: Writing Samples *H: Handwriting Legibility Scale*	
Knowledge (*Gc*)		Test 19: Academic Knowledge

NOTE: Italic font designates supplemental tests.

Copyright © 2001 by The Riverside Publishing Company. Adapted from the Woodcock-Johnson® III (WJ III™) by Richard W. Woodcock, Kevin S. McGrew and Nancy Mather, with permission of the publisher. All rights reserved.

The Cognitive Battery

The Tests of Cognitive Abilities (WJ III COG) consists of 20 tests: a Standard Battery (Tests 1–10) and an Extended Battery (Tests 11–20), with each test designed to measure a different aspect of cognitive abilities. The 20 tests in the WJ III COG provide a wide variety of clusters useful for clinical and diagnostic purposes. The WJ III COG provides measurement of seven broad CHC factors described previously in this chapter. While each of the 20 tests is an independent test, interpretation of the individual tests serves primarily to understand the broader clusters. The clusters provide the primary basis for test interpretation. Assessment time will vary based on the choice of clusters and the resultant number of tests selected for administration. Most tests in the cognitive battery take about 5 to 10 minutes to administer, with the seven tests needed to get a General Intellectual Ability-Standard score (GIA-Std) requiring approximately 45 minutes and the GIA-Extended (GIA-Ext) taking about 75 to 90 minutes.

The broad abilities from the CHC model are represented by clusters in the WJ III COG. Each test (1–20) was designed to measure a narrow ability within its broad ability. The tests in the WJ III COG were revised from the WJ-R or were newly designed to decrease sources of variance that are irrelevant to the narrow ability measured by the test and to insure adequate construct representation for each broad *Gf–Gc* cluster (McGrew & Woodcock, 2001). Table 14.4 describes the broad and narrow CHC abilities measured by the WJ III COG tests.

The WJ III General Intellectual Ability (GIA) score is a measure of *g* or general intelligence and is the best single predictor of a performance, *on average*, across a wide variety of academic and cognitive outcomes. The GIA score is a differentially weighted combination of tests of seven broad CHC abilities. The test weightings for the GIA-Std score are derived from the first principal components analysis of Tests 1 through 7 (one indicator of each CHC factor). The GIA-Ext test weights are similarly calculated using Tests 1 through 7 and Tests 11 through17 (two indicators of each CHC factor). While both have strong reliabilities throughout the life span, the broader mix of CHC narrow abilities assessed makes the GIA-Ext the single best measure of theoretical *g*. A Brief Intellectual Ability (BIA) cluster comprised of three tests (Verbal Comprehension, Concept Formation, and Visual Matching) is appropriate for screening purposes. Unlike the weighted GIA scores, the BIA is an arithmetic average of the three tests.

The *Predicted Achievement* scores are differentially weighted combinations of WJ III COG-Std Tests 1 through 7 and are used to predict achievement in reading, mathematics, written language, oral language, and academic knowledge. Each of the five Predicted Achievement scores is based on a differential weighting procedure that produces the best prediction for the achievement criterion in the near term. The seven tests are differentially weighted by age to allow for the best prediction in an academic area without including tests that overlap with the achievement criteria. The design and use of the GIA and Predicted Achievement options in ability/achievement comparisons are discussed later in this chapter.

The Achievement Battery

Like the WJ III COG, the WJ III ACH tests are organized into a Standard Battery (Tests 1–12) and an Extended Battery (Tests 13–22). A primary goal of the revision of the WJ-R was to enhance the diagnostic capabilities of the battery. Both breadth and depth have been added to the WJ III ACH, improving the utility of many of the tests from the WJ-R for use with young children. In addition, a number of new "special purpose clusters" were added in each curricular area.

The 22 tests and 20 clusters provided by the WJ III ACH allow for a comprehensive or selective diagnostic assessment. Seven of the clusters on the WJ III ACH are aligned with the seven areas of Learning Disability specified in the Individuals with Disabilities Education Act (IDEA, 1997). In addition, the WJ III includes an important eighth area of disability, expressive oral language. Of particular relevance to assessments of students at the secondary level are the fluency measures in reading, mathematics, and written language and the Academic Fluency cluster. The movement of the oral language tests from the WJ III COG also adds increased diagnostic utility to the WJ III ACH battery. Testing time will vary, depending on the age of the person evaluated, the reason for referral, and the number of tests administered. The 22 tests in the WJ III ACH are described in Table 14.5.

Broad cluster scores are available in all three academic areas (Reading, Mathematics, and Written Language), and an Oral Language—Standard score is also available by administering tests in the Standard Battery. Additional cluster scores in each academic area and oral language are available by administering additional tests in the Extended Battery. A general academic proficiency cluster score, Total Achievement, is available, based on nine tests in the standard battery.

Table 14.4 WJ III COG CHC ability construct and content coverage of the individual tests

Test	Primary Broad CHC Factor	Narrow CHC Ability	Task Requirement
Test 1: Verbal Comprehension	Comprehension–Knowledge (*Gc*)	Lexical Knowledge Language development	Naming pictures; knowledge of synonyms and antonyms; completing verbal analogies
Test 2: Visual–Auditory Learning	Long-Term Retrieval (*Glr*)	Associative memory	Learning and recalling pictographic representations of words
Test 3: Spatial Relations	Visual–Spatial Thinking (*Gv*)	Visualization Spatial relations	Identifying the subset of pieces needed to form a complete shape
Test 4: Sound Blending	Auditory Processing (*Ga*)	Phonetic coding: Synthesis	Synthesizing language sounds (phonemes)
Test 5: Concept Formation	Fluid Reasoning (*Glr*)	Induction	Identifying, categorizing, and determining rules for a set of objects
Test 6: Visual Matching	Processing Speed (*Gs*)	Perceptual Speed	Rapidly locating and circling identical numbers from sets of numbers
Test 7: Numbers Reversed	Short-Term Memory (*Gsm*)	Working memory	Holding a span of numbers in immediate awareness while reversing the sequence
Test 8: Incomplete Words	Auditory Processing (*Ga*)	Phonetic coding: analysis	Identifying words with missing phonemes
Test 9: Auditory Working Memory	Short-Term Memory (*Gsm*)	Working memory	Holding a mixed set of numbers and words in immediate awareness while reporting the sequence of words, then numbers
Test 10: Visual–Auditory Learning-Delayed	Long-Term Retrieval (*Glr*)	Associative Memory	Recalling and relearning pictographic representations of words from 30 minutes to 8 days later
Test 11: General Information	Comprehension–Knowledge (*Gc*)	General (verbal) information	Identifying where objects are found and what people typically do with an object
Test 12: Retrieval Fluency	Long-Term Retrieval (*Glr*)	Ideational Fluency	Naming as many examples as possible from given categories
Test 13: Picture Recognition	Visual–Spatial Thinking (*Gv*)	Visual memory	Identifying a subset of previously presented pictures within a field of distracting pictures
Test 14: Auditory Attention	Auditory Processing (*Ga*)	Speech-sound discrimination Resistance to auditory stimulus distortion	Identifying auditorily presented words amid increasingly intense background noise

(Continues)

TABLE 14.4 (Continued)

Test	Primary Broad CHC Factor	Narrow CHC Ability	Task Requirement
Test 15: Analysis–Synthesis	Fluid Reasoning (*Gfr*)	General sequential (deductive) reasoning	Analyzing puzzles (using symbolic formulations) to determine missing components
Test 16: Decision Speed	Processing Speed (*Gs*)	Semantic processing speed	Locating and circling two pictures most similar conceptually in a row
Test 17: Memory for Words	Short-Term Memory (*Gsm*)	Memory span	Repeating a list of unrelated words in correct sequence
Test 18: Rapid Picture Naming	Processing Speed (*Gs*)	Naming facility	Recognizing objects, then retrieving and articulating their names rapidly
Test 19: Planning	Visual–Spatial Thinking (*Gv*) Fluid Reasoning (*Gf*)	Spatial scanning General sequential reasoning	Tracing a maze pattern without removing the pencil from the paper or retracing any lines
Test 20: Pair Cancellation	Processing Speed (*Gs*)	Attention and concentration	Identifying and circling instances of a repeated pattern rapidly

Copyright © 2001 by The Riverside Publishing Company. Adapted from the Woodcock-Johnson® III (WJ III™) by Richard W. Woodcock, Kevin S. McGrew and Nancy Mather, with permission of the publisher. All rights reserved.

TABLE 14.5 *WJ III ACH CHC ability construct and content coverage of the individual tests*

Test	Primary Broad CHC Factor	Narrow CHC Ability	Test Requirement
Test 1: Letter–Word Identification	Reading & Writing (*Grw*)	Reading decoding	Identifying printed letters and words
Test 2: Reading Fluency	Reading & Writing (*Grw*) Processing Speed (*Gs*)	Reading speed	Reading printed statements rapidly and responding true or false (Yes or No)
Test 3: Story Recall	Comprehension–Knowledge (*Gc*) Long-term Retrieval (*Glr*)	Language development Listening ability Meaningful memory	Listening to and recalling details of stories
Test 4: Understanding Directions	Comprehension–Knowledge (*Gc*)	Listening ability Language development	Listening to a sequence of instructions and then following the directions
Test 5: Calculation	Quantitative Knowledge (*Gq*)	Math achievement	Performing various mathematical calculations
Test 6: Math Fluency	Quantitative Knowledge (*Gq*) Processing Speed (*Gs*)	Math achievement Number fluency	Adding, subtracting, and multiplying rapidly
Test 7: Spelling	Reading & Writing (*Grw*)	Spelling ability	Spelling orally presented words
Test 8: Writing Fluency	Reading & Writing (*Grw*) Processing Speed (*Gs*)	Writing speed	Formulating and writing simple sentences rapidly
Test 9: Passage Comprehension	Reading & Writing (*Grw*)	Reading Comprehension Verbal (printed) language comprehension Cloze ability	Identifying a missing key word that makes sense in the context of a written passage
Test 10: Applied Problems	Quantitative Knowledge (*Gq*)	Quantitative reasoning Math achievement Math knowledge	Performing math calculations in response to orally presented problems
Test 11: Writing Samples	Reading & Writing (*Grw*)	Writing ability	Writing meaningful sentences for a given purpose
Test 12: Story Recall-Delayed	Long-Term Retrieval (*Glr*)	Meaningful memory	Recalling previously presented story elements
Test 13: Word Attack	Reading & Writing (*Grw*)	Reading decoding Phonetic coding: Analysis & synthesis	Recalling phonically regular nonwords
Test 14: Picture Vocabulary	Comprehension–Knowledge (*Gc*)	Language development Lexical knowledge	Naming pictures

(Continues)

TABLE 14.5 (Continued)

Test	Primary Broad CHC Factor	Narrow CHC Ability	Test Requirement
Test 15: Oral Comprehension	Comprehension–Knowledge (*Gc*)	Listening ability	Identifying a missing key word that makes sense in an oral passage
Test 16: Editing	Reading & Writing (*Grw*)	Language development English usage	Identifying and correcting errors in written passages
Test 17: Reading Vocabulary	Reading & Writing (*Grw*) Comprehension–Knowledge (*Gc*)	Verbal (printed) language comprehension Lexical knowledge	Reading words and supplying appropriate meanings
Test 18: Quantitative Concepts	Quantitative Knowledge (*Gq*) Fluid Reasoning (*Gf*)	Math knowledge Quantitative reasoning	Identifying math terms and formulae Identifying number patterns
Test 19: Academic Knowledge	Comprehension–Knowledge (*Gc*)	General information Science information Cultural information Geography achievement	Responding to questions about science, social studies, and humanities
Test 20: Spelling of Sounds	Reading & Writing (*Grw*) Auditory Processing (*Ga*)	Spelling ability Phonetic coding: Analysis & synthesis	Letter combinations that are regular patterns in written English
Test 21: Sound Awareness	Auditory Processing (*Ga*)	Phonetic coding: Analysis & synthesis	Providing rhyming words, removing, substituting, and reversing parts of words to make new words
Test 22: Punctuation & Capitalization	Reading & Writing (*Grw*)	English usage	Applying punctuation and capitalization rules

Copyright © 2001 by The Riverside Publishing Company. Adapted from the Woodcock-Johnson® III (WJ III™) by Richard W. Woodcock, Kevin S. McGrew and Nancy Mather, with permission of the publisher. All rights reserved.

Interpretive Features of the WJ III

As a comprehensive co-normed battery, the WJ III provides a wide array of interpretive features, some not typically available in other test batteries. Four levels of interpretive information are available for the tests and clusters. While each of the four levels provides unique information about the person's test performance, the levels of information cannot be used interchangeably. Table 14.6 provides a summary of the levels of interpretation available on the WJ III. At *Level One*, qualitative aspects of the subject's performance are noted, including test session observations and error pattern analysis. *Level Two* provides information from the raw scores about the individual's stage or level of development. At *Level Three*, the quality of a person's performance on criterion tasks at different difficulty levels is indicated. Normative comparisons to peers in the standardization sample are available at *Level Four*.

Both age- and grade-based norms are available. Grade norms are available for students in Grades K through 12, two-year colleges, and four-year colleges, including graduate school. Age norms are available for individuals age two through 95+. While the age and grade equivalents are not affected by selection of age or grade-based norms, relative proficiency index, standard score, and percentile rank are affected by an examiner's choice of norms. While the decision of which norms should be used is left to the examiner, if age norms are used to score the WJ III ACH, then age norms also should be used to score the WJ III COG and vice versa.

TABLE 14.6 Level of interpretive information available on the WJ III

Level	Type of Information	Basis	Information & Scores
1	Qualitative (Criterion-Referenced)	Observation during testing and analysis of responses	Description of the subject's reaction to the test situation Performance on finely defined skills at the item content level
2	Level of Development (Norm-Referenced)	Sum of item scores Age or grade level in the norming sample at which the average is the same as the subject's score	Raw Score Rasch Ability Score (W Score) Age Equivalent (AE) Grade Equivalent (GE)
3	Proficiency (Criterion-Referenced)	Subject's distance on a Rasch scale from an age or grade reference point	Quality of performance on reference tasks Rasch Difference score Relative Proficiency Index (RPI) CALP Level Developmental or Instructional Zone
4	Relative Standing in a Group (Norm-Referenced)	Relative position (a transformation of a difference score, such as dividing by the standard deviation of the reference group)	Rank Order Standard Score (SS) (including T score, z score, NCE, Discrepancy SD DIFF) Percentile Rank (PR) (including Discrepancy PR)

Copyright © 2001 by The Riverside Publishing Company. Adapted from the Woodcock-Johnson® III (WJ III™) my Richard W. Woodcock, Kevin S. McGrew and Nancy Mather, with permission of the publisher. All rights reserved.

Types of Scores

A variety of scores are provided by the WJ III, including age and grade equivalents, relative proficiency indexes (RPI), cognitive–academic language proficiency (CALP) levels, percentile ranks, and standard scores. It is not possible to obtain any derived scores by hand except estimated age and grade equivalents. All scores are obtained through the use of the *WJ III Compuscore and Profiles Program* (Schrank & Woodcock, 2001). The program provides a number of special and unique features that would not be possible if scoring were done by hand. Several of the scores provided on the WJ III are discussed below.

W Score

All raw scores are converted to *W* scores, which are a special transformation of the Rasch ability scale (Rasch, 1960; Wright & Stone, 1979). The equal interval properties of the *W* scale make it a useful intermediate step in test interpretation and for measuring growth. The *W* scale for each test is centered on a value of 500 (the approximate average performance of a ten-year-old). Cluster scores are the average of the *W* scores for the individual tests in the cluster (Mather & Woodcock, 2001).

Age and Grade Equivalents

The WJ III provides both age (AE) and grade equivalent (GE) scores. An AE or GE reflects the subject's performance in terms of the age or grade level in the norming sample at which the median score is the same as the subject's score. The WJ III AE and GE scores have advantages over AE or GE scores reported on many other test batteries. One frequently cited criticism of grade (or age) scores is that they are not useful for instructional planning because they do not reflect the student's ability. It is not always recognized that this common criticism of GE's applies to tests that are composed primarily of items with a limited range of difficulty, such as the multilevel tests of many group achievement batteries. For example, if a third-grade student earns a grade equivalent of 6.5 on a test that is designed for students in grade 3, it does not mean that the student will be successful on tasks associated with the mid-sixth-grade level. Rather, it means that the student got a high percentage of the items on a third-grade test correct, the same percentage of items correct that an average sixth-grade student received. In this case, the student's score is a reflection more of the student's accuracy than of the grade level of task difficulty that this student can perform.

The "just say 'no' to grade equivalents" mantra does not apply when test items are: (1) distributed uniformly over a wide range of difficulty, (2) when individuals are administered the subset of items centered on their level of ability, and (3) when the test has been normed on an appropriately selected sample of individuals across a wide grade range (McGrew et al., 1991). The latter three conditions characterize the WJ III AE and GE scores. The WJ III AE and GE scores do, in fact, reflect the level of task difficulty an individual can perform and thus may be useful in instructional planning.

Relative Proficiency Index

The relative proficiency index (RPI), formerly called the RMI (relative mastery index) on the WJ-R, is a valuable score in better understanding a subject's quality of performance relative to peers in the normative sample. The score reads like the index used with Snellen charts to describe visual acuity. A 90 is always written in the denominator. An RPI score of 90/90 means that the subject demonstrated 90% proficiency on tasks where the average person in the comparison group (same age or grade) would also obtain 90%. The Developmental Zone (called the Instructional Zone on the WJ III ACH), is a special application of the RPI provided to help understand the subject's range of functioning on tasks from "easy" (independent level) to "difficult" (frustration level). The Developmental and Instructional Zone profiles are printed when using the *Compuscore and Profiles Program*.

CALP Levels

Cognitive Academic Language Proficiency (CALP) is described by Cummins (1984) as language used in academic situations and those that result from formal schooling. A CALP level can be reported for both the WJ III COG, using the Verbal Ability clusters, and the WJ III ACH, using the Academic Knowledge cluster and several oral and written language tests and clusters. The availability of this CALP score is valuable for examiners working with students who are non-native English speakers. The five CALP levels (1 = Negligible to 5 = Advanced) provide examiners with information useful in describing English language proficiency in academic settings. These levels also may provide information to help a nonbilingual examiner make informed referrals for evaluation by a bilingual evaluator.

Percentile Rank

Percentile ranks are provided for all tests and clusters. They are useful in describing the person's relative standing in the population. A unique feature of the WJ III percentile ranks is the extended percentile ranks at the upper and lower ends of the scale. The extension of the scale adds approximately one and one-half standard deviation units to the range of the traditional percentile rank scale.

Standard Score

The standard score scale for the WJ III is based on a mean of 100 and a standard deviation of 15. The standard score is the score most commonly reported in clinical practice. The WJ III provides standard scores from 1 to greater than 200. In addition, the *Compuscore and Profiles Program* provides an option to select a z-score, T-score, NCE, or stanine.

Types of Profiles

The use of the *Compuscore and Profiles Program* in scoring the WJ III provides the opportunity to plot the *Age/Grade Profile* or the *Standard Score/Percentile Rank Profile.* These profiles provide a visual display of the person's Developmental Zone for the WJ III COG (called Instructional Zone on the WJ III ACH) and normative comparisons. The *Age/Grade Profile* is particularly useful when the examiner needs to explain the person's test performance for instructional planning. The left end of the shaded zone on a graphic bar represents the age or grade level where the subject would perceive the tasks as *easy* (RPI = 96/90). The right end of the zone represents the age or grade level at which the subject would perceive the tasks as *difficult* (RPI = 75/90). An easy level represents a person's independent level for instructional purposes while the difficult level would represent the frustration level. The width of the band will vary, as some zones will appear narrow while others appear wide. The width of the band reflects how rapidly or slowly the underlying skill or ability changes over age or grade (see section on WJ III CHC growth curves later in this chapter). A wide band reflects a slow rate of change while a narrow band indicates a rapid rate of change over time.

The *Standard Score/Percentile Rank Profile* provides a plot of the confidence band surrounding the standard score and percentile for a given test and/or cluster. The confidence band represents the region within which the subject's true score on a test or cluster most likely falls. The software program provides for the option to select three different levels of confidence (68%, 90%, 95%). The 68% confidence interval is recommended for profile interpretation. While statistical procedures are available to interpret differences in scores, guidelines are provided that allow for a visual interpretation of the display.

Clinical and Selected Special Purpose Clusters

Additional cluster scores are available to provide more comprehensive diagnostic information from both the Tests of Cognitive Abilities and the Tests of Achievement. While these clusters are likely not part of the typical test battery an examiner may give for initial evaluations, they do provide valuable special clinical and diagnostic information. Five *clinical clusters* are available on the WJ III COG. The *Phonemic Awareness* cluster measures the ability to attend to the phonemic structure of language by analyzing and synthesizing speech sounds. This ability is important to early reading as well as spelling acquisition. Scores on phonemic awareness measures have a strong relationship with reading achievement (Flanagan et al., 2001). In addition to the two-test Phonemic Awareness cluster available on the WJ III COG, a Phonemic Awareness III cluster is available by combining the Incomplete Words and Sound Blending tests (both from the WJ III COG) and the Sound Awareness test from the WJ III ACH.

Broad Attention is a special clinical cluster that provides a global measure of attention. Frequently, attention and concentration difficulties are measured using external behavior observation measures. However, more recent ADHD research points to the importance of examining cognitive indicators in addition to traditional behavior indicators (Barkley, 1996). The four tests that comprise the Broad Attention cluster—Auditory Working Memory, Numbers Reversed, Auditory Attention, and Pair Cancellation—each measure a qualitatively different aspect of attention (see neuropsychological applications section later in this chapter for more information). The *Working Memory* cluster provides additional information about a person's ability to hold information in immediate awareness while performing operations on it. The two tests comprising the Working Memory cluster (Auditory Working Memory, Numbers Reversed) require divided attention and the management of the limited capacity of short-term memory.

The *Cognitive Fluency* cluster is comprised of three tests that collectively measure a person's ability to quickly and fluently perform simple to complex cognitive tasks: speed of retrieval from stored knowledge (Retrieval Fluency); speed of forming simple concepts (Decision Speed); and speed of lexical (vocabulary) access (Rapid Picture Naming). Finally, the COG battery also provides an *Executive Processing* cluster that provides information regarding a person's ability to effectively control and implement cognitive processes for the purpose of integrating short-term and long-term future goals (Eslinger, 1996). The Planning, Pair Cancellation, and Concept Formation tests tap different aspects of central executive control, namely, strategic planning (forward thinking), proactive interference control, and the ability to repeatedly shift mind-set (cognitive flexibility).

Four special purpose clusters are available on the WJ III ACH. The *Academic Skills* cluster is obtained from three skills tests (one in each academic area). Formerly known as *Basic Skills* on the WJ-R, the Academic Skills cluster provides useful information about the examinee's abilities in basic academic skills, including sight word recognition, spelling, and math calculation. In contrast, *Academic Applications* is a cluster that includes measures of academic reasoning in the three curricular areas. For some persons referred for evaluation, basic academic skills may be well developed in all academic areas, but they have difficulty with the more complex academic applications that require reasoning skills. Others may do well in all academic areas in both skills and applications, but they may have difficulty in their ability to work smoothly and efficiently with their abilities. The *Academic Fluency* cluster measures efficiency and the ability to work in the reading, math, and written language areas with ease. The Academic Fluency cluster may provide particularly relevant information when evaluating special accommodation requests for extended time on tasks. The *Phoneme/Grapheme Knowledge* cluster is a potentially clinically useful feature of the WJ III ACH, especially when this cluster is used in conjunction with the *Phonemic Awareness* cluster from the WJ III COG. The Phoneme/Grapheme Knowledge cluster is comprised of the Word Attack and Spelling of Sounds tests, tests in which examinees are asked to read and spell non-words with regular English spelling patterns.

The Cognitive and Achievement Performance Models

The *Cognitive and Achievement Performance Models (CA-PM)* are another set of frameworks that can assist in interpreting an individual's performance on the WJ III. The CA-PM are based largely on logical and theoretical considerations rather than empirical data (Woodcock, 1993, 1997). Based on the CA-PM frameworks, additional clusters are provided for diagnostic information. The models indicate that four overarching factors can impact a person's "real-world" cognitive and academic performance. Indicators of three of the four CA-PM domains can be obtained from the WJ III. Figure 14.2 illustrates the relationship between the four indicators and cognitive and achievement performance.

The conceptual roots of the CA-PM framework can be traced back to Spearman, who, in *The Abilities of Man*, stated that "the process of cognition cannot possibly be treated apart from those of conation and affection, seeing that all these are inseparable aspects in the instincts and behavior of a single individual, who himself, as the very name implies, is indivisible" (Spearman, 1927, p. 2). David Wechsler was similarly convinced that a variety of *nonintellectual factors* (e.g., persistence, curiosity, and motivation) influenced the expression of intelligent behavior (Zachary, 1990). Snow's (1989) work on *aptitude complexes* and Ackerman's more recent *intelligence-as-process*,

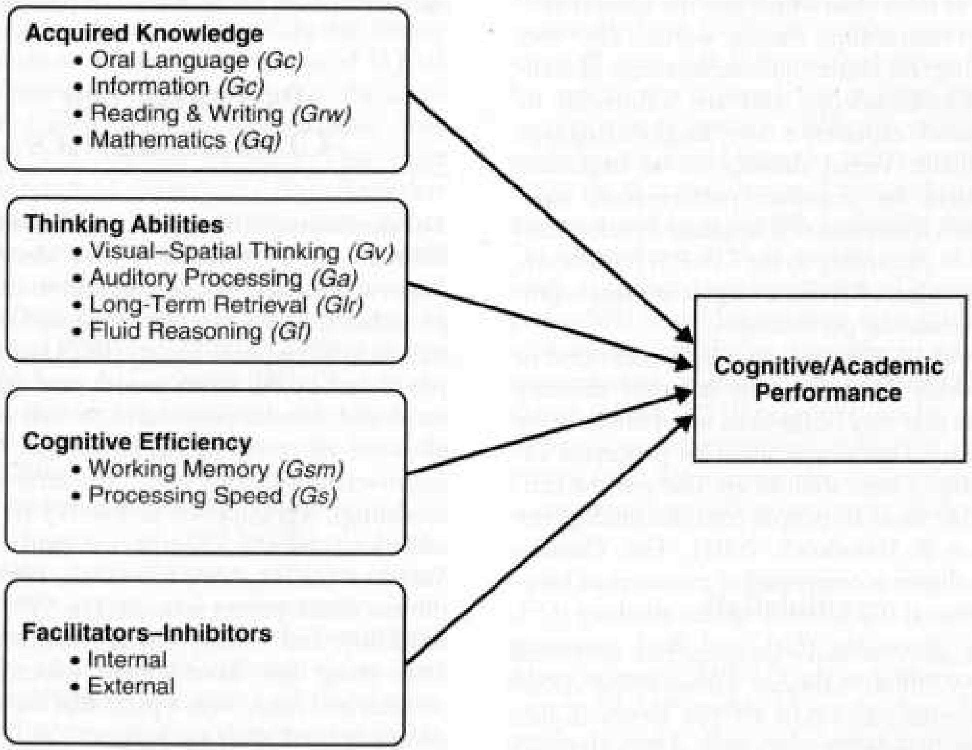

FIGURE 14.2

The WJ III cognitive and achievement performance models

SOURCE: Copyright © 2001 by The Riverside Publishing Company. Adapted from the Woodcock-Johnson®III (WJ III®) by Richard W. Woodcock, Kevin S. McGrew and Nancy Mather, with permission of the publisher. All rights reserved.

personality, interests, intelligence-as-knowledge (PPIK) theory (Ackerman, 1996; Rolfhus & Ackerman, 1999) represent other attempts to integrate noncognitive constructs into a larger theoretical framework to explain variations in cognitive and academic performance. The CA-PM is conceptually useful for bridging psychometric perspectives on test interpretation (such as the CHC model) and information-processing models of cognitive and academic functioning

Briefly, the CA-PM stores of acquired procedural and declarative knowledge are represented by the *Verbal Abilities* clusters (Std or Ext) on the WJ III COG and the respective WJ III ACH math (*Gq*) and reading and written language (*Grw*) clusters. In the cognitive model, the WJ III Verbal Abilities cluster provides a measure of language development that includes the comprehension of individual words and the comprehension of relationships among words. The tests comprising this cluster include measures of comprehension-knowledge and are influenced by the person's experience and English-language development. Verbal abilities are an important requirement for cognitive performance, especially when it involves oral language development in English. According to the CA-PM framework, "things you know" is one area that impacts cognitive and academic performance.

The WJ III *Thinking Abilities* clusters (Std or Ext) provide a sampling of different thinking processes that may be invoked when information in short-term memory cannot be processed automatically. These abilities are likely at the center of what most laypeople consider intelligence (McGrew & Woodcock, 2001). The *Thinking Abilities* cluster is comprised of measures of long-term retrieval (*Glr*), visual–spatial thinking (*Gv*), auditory processing (*Ga*), and fluid reasoning (*Gf*). According to the CA-PM, a person could "know things" yet not be able to "think" or reason with that knowledge well. Thus, thinking abilities impact cognitive performance.

Short-term memory and processing speed influence the efficiency of a person's cognitive performance. The WJ III *Cognitive Efficiency* clusters (Std or Ext) measure this aspect of the CA-PM. These two aspects of automatic cognitive processing (*Gs* and *Gsm*) represent the capacity of the cognitive system to process information automatically. This automatic processing facilitates complex cognitive functioning. In the CA-PM, people must be able to "work efficiently" with their "knowledge" and "thinking" skills to demonstrate their best cognitive and academic performance. Finally, *facilitators–inhibitors* to cognitive performance modify cognitive and academic performance for better or worse. These noncognitive factors may be *internal* to the person (e.g., emotional state, health, and motivation) or *external* to the person in the environment (e.g., auditory distractions) or the result of situational variables (e.g., the teaching method). Woodcock (1993, 1997) and Dean and Woodcock (1999) have presented a more detailed and complex *CHC Information Processing Model* that is not presented here.

PSYCHOMETRIC CHARACTERISTICS

The development of the original WJ and WJ-R followed many traditional test development stages and procedures. In addition to traditional procedures, new concepts introduced in latent-trait or item-response theory (IRT) and the analysis of data by the Rasch model were extensively employed. Similar procedures, as well as the application of recent developments in IRT and multivariate statistics (e.g., structural equation modeling), were applied to the WJ III data. In addition, particular attention was paid to the *Test Standards* (AERA, APA, & NCME, 1999) during the test development process. The WJ III norm, reliability, and validity characteristics resulting from these test development procedures are summarized here, with a particular focus on the adolescent and adult age ranges.

Norms

The WJ III norms were calculated on the largest standardization sample of any individually administered battery of cognitive and achievement tests. A total of 8,818 individuals, from age 24 months to age 95+ years, living in more than 100 geographically and economically diverse communities in the United States, were assessed. The norm subjects were randomly selected via a stratified sampling plan that controlled for 10 specific individual (e.g., parent SES for school-age subjects; occupational status and level for adults) and community (e.g., community size, SES, etc.) variables. The preschool sample includes 1,143 children from 2 to 5 years of age (not enrolled in kindergarten). The kindergarten to 12th-grade sample is composed of 4,783 students, the college/university sample is based on 1,165 students, and the adult sample includes 1,843 individuals.

Although the distribution of norming subjects approximated the U.S. population distribution, individual subject weighting was applied during data analysis to obtain a distribution of WJ III data that was exactly proportioned to the community and individual sampling variables from year 2000 U.S. Census statistics. This exact weighting removed the potential bias effects that might result from having approximate, rather than exact, proportional representation in each cell of the sampling design. The sample and norming procedures are described in detail in McGrew and Woodcock (2001). Continuous age norms (Zachary & Gorsuch, 1985; Woodcock, 1987b) are provided at one-month intervals from 2–0 to 18–11 years of age and one-year intervals from 19 to 95+.

The combined adult and college/university sample is comprised of 3,008 individuals. In addition to the standard census variables of sex, race, Hispanic origin, and census region, individual-level information was also collected (and used in the adult subject weighting) on number of years of education, occupational/employment status, and occupational job category or status (e.g., white collar, service worker, etc.). Unique to the WJ, WJ-R, and WJ III, when compared to traditional norming plans, was the collection and use of community characteristic information in the selection of communities. The community SES characteristics included the distribution of the population in each community according to years of education, household income, labor force characteristics (i.e., percent employed, unemployed, and not in the labor force), and occupational characteristics (i.e., percent in white-collar, service, etc., job categories).

The WJ III provides grade-based norms for subjects in kindergarten through high school and for adults enrolled in postsecondary institutions. New to the WJ III is the provision of *separate* continuous grade norms (13.0 to 18.0) for subjects in two-year and four-year postsecondary institutions.[1] This provides the ability to compare an adult subject enrolled in a postsecondary institution against a representative sample of peers at similar institutions. For example, a 27-year-old with an obtained WJ III Broad Reading *W*-score of 524 would be at the 20th percentile (standard score [SS] = 87) when compared to other 27-year-olds. If this individual was in his or her second year of instruction at a two-year community college (grade = 14.0), that same reading performance would be at the 12th percentile (SS = 82) for that population. This level of performance would be at the 6th percentile rank (SS = 76) for students in their second year of coursework at a four-year institution. This could be particularly useful information for a subject if he or she were considering a transfer to a four-year postsecondary institution. Adult subject performance on the WJ III cognitive can be scored in reference to three different normative reference groups (viz., age norms; two-year college grade norms; four-year college/university grade norms).

Reliability

Reliability and standard error or measurement (SEM) statistics are reported in the technical manual for all WJ III tests and clusters across their range of intended use (McGrew & Woodcock, 2001). The reliabilities for all but the

[1]Grade 18.0 represents the beginning of the year for second-year graduate students.

speeded tests and tests with multiple-point scoring systems were calculated using the split-half procedure (odd and even items) and corrected for publication test length using the Spearman-Brown correction formula. The reliabilities for the multiple-point and speeded tests were calculated using the unique standard error of measurement for each subject from Rasch analysis (McGrew & Woodcock, 2001). The reliability of cluster scores were calculated by a formula from Mosier (1943; McGrew & Woodcock, 2001). One-day interval test-retest reliabilities are reported for 165 subjects on the eight WJ III speeded tests. Test-retest correlation studies with an interval of less than one year to 10 years, involving more than 1,600 subjects, are also reported.

The WJ III test and cluster median reliabilities for eight adolescent and adult age ranges are presented in Tables 14.7 and 14.8.[2] As summarized in Table 14.7, 18 of the cognitive test median reliabilities meet or exceed the .80 level standard and 12 meet or exceed the .90 standard. All 22 achievement test median reliabilities exceed .80 and 11 meet or exceed the more stringent .90 standard. Across the WJ III cognitive and achievement batteries almost all of the 42 clusters meet or exceed the .90 reliability standard (Table 14.8). The WJ III cognitive and achievement clusters, the recommended unit for interpretation and decision making for the WJ III, possess strong reliabilities across the adolescent and adult age ranges.

Validity

Validity is the most important consideration in test development, evaluation, and interpretation and "refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (AERA, 1999). Furthermore, validity is not a single-point event; it involves accumulating multiple sources of validity evidence over time.

[2]A detailed breakdown of reliabilities by specific age groups can be found in the WJ III technical manual (McGrew & Woodcock, 2001).

For many WJ III tests and clusters, validity evidence has accumulated across three different versions of the battery. For example, five of the Standard COG tests (Visual–Auditory Learning, Sound Blending, Concept Formation, Visual Matching, and Numbers Reversed) have retained the same general format across all three versions of the battery. Two additional tests (Spatial Relations and Incomplete Words) have two generations of accumulated research evidence. The remaining test (Verbal Comprehension) is comprised of four subtests for which validity data is present across all three versions.

Summarizing all generations of validity evidence presented in the respective technical manuals (McGrew, Werder, and Woodcock, 1991; McGrew & Woodcock, 2001; Woodcock, 1978) and by independent researchers is beyond the scope of this chapter. Much of this information can be found in the three respective technical manuals and other publications (McGrew, 1986, 1994). Instead, the current chapter will focus on summarizing WJ III-specific validity evidence for the adolescent and adult age ranges. Post-WJ III publication validity research (with subjects from the adolescent and adult age ranges) is also presented.

Content Validity

Although content validity has been a focus in all three editions of the WJ III, content validity evidence derived from a theoretically based test design specification framework received special attention during the WJ III revision. As described previously, the CHC theory served as the test-design blueprint for the WJ III (see Figure 14.1). The use of the strong CHC theory, a theory based on evidence accumulated over nearly 60 years of research, maximizes the substantive (content) validity vis-à-vis the specification of a well-bounded construct domain. Although confirmatory factor analysis and developmental evidence are presented in the technical manual to support the WJ III test narrow ability classifications, the WJ III narrow ability test classifications rest primarily on expert and logical task

TABLE 14.7 Median WJ III cognitive and achievement test reliabilities across eight adolescent/adult age groups (from 14 to 80+ years)

Standard Cognitive Battery Tests	Median Reliability	Standard Achievement Battery Tests	Median Reliability
Test 1: Verbal Comprehension	0.95	Test 1: Letter–Word Identification	0.94
Test 2: Visual–Auditory Learning	0.91	Test 2: Reading Fluency	0.90
Test 3: Spatial Relations	0.83	Test 3: Story Recall	0.88
Test 4: Sound Blending	0.92	Test 4: Understanding Directions	0.89
Test 5: Concept Formation	0.95	Test 5: Calculation	0.89
Test 6: Visual Matching	0.92	Test 6: Math Fluency	0.91
Test 7: Numbers Reversed	0.90	Test 7: Spelling	0.95
Test 8: Incomplete Words	0.88	Test 8: Writing Fluency	0.92
Test 9: Auditory Working Memory	0.86	Test 9: Passage Comprehension	0.86
Test 10: Visual–Auditory Delayed Recall	0.94	Test 10: Applied Problems	0.94
		Test 11: Writing Samples	0.90
		Test 12: Story Recall-Delayed	0.82
Extended Cognitive Battery Tests	**Median Reliability**	**Extended Achievement Battery Tests**	**Median Reliability**
Test 11: General Information	0.94	Test 13: Word Attack	0.87
Test 12: Retrieval Fluency	0.91	Test 14: Picture Vocabulary	0.90
Test 13. Picture Recognition	0.79	Test 15: Oral Comprehension	0.89
Test 14: Auditory Attention	0.88	Test 16: Editing	0.89
Test 15: Analysis–Synthesis	0.94	Test 17: Reading Vocabulary	0.92
Test 16: Decision Speed	0.90	Test 18: Quantitative Concepts	0.94
Test 17: Memory for Words	0.85	Test 19: Academic Knowledge	0.91
Test 18: Rapid Picture Naming	0.97	Test 20: Spelling of Sounds	0.82
Test 19: Planning	0.74	Test 21: Sound Awareness	0.83
Test 20: Pair Cancellation	0.84	Test 22: Punctuation and Capitalization	0.88

Copyright © 2001 by The Riverside Publishing Company. Adapted from the Woodcock-Johnson® III (WJ III™) by Richard W. Woodcock, Kevin S. McGrew and Nancy Mather, with permission of the publisher. All rights reserved.

analyses, including expert consensus-based task analyses of the WJ III tests.

Structural Validity: Norm-Based Studies

The design of both the WJ-R and WJ III paid particular attention to the extent to which the relations among test scores (and/or their components) conformed to the relations implied by the CHC theoretical construct domain. This is typically referred to as internal or structural validity (AEAR, 1999; Benson, 1998). Structural validity research focuses on answering the question: "*Do the observed measures behave in a manner consistent with the theoretical domain definition of intelligence?*" Although it would be possible to report and discuss the patterns of test and cluster correlations

TABLE 14.8 Median WJ III cognitive and achievement cluster reliabilities across eight adolescent/adult age groups (from 14 to 80+ years)

Standard Cognitive Battery Clusters	Median Reliability	Standard Achievement Battery Clusters	Median Reliability
General Intellectual Ability (Std)	0.98	Total Achievement	0.98
Brief Intellectual Ability	0.97	Oral Language (Std)	0.92
Verbal Ability (Std)	0.95	Broad Reading	0.94
Thinking Ability (Std)	0.97	Broad Mathematics	0.96
Cognitive Efficiency (Std)	0.94	Broad Written Language	0.97
Phonemic Awareness	0.94	Academic Skills	0.97
Working Memory	0.93	Academic Fluency	0.94
		Academic Applications	0.96
Extended Cognitive Battery Clusters	**Median Reliability**	**Extended Achievement Battery Clusters**	**Median Reliability**
General Intellectual Ability (Ext)	0.99	Oral Language (Ext)	0.95
Verbal Ability (Ext)	0.97	Oral Expression	0.91
Thinking Ability (Ext)	0.98	Listening Comprehension	0.94
Cognitive Efficiency (Ext)	0.95	Basic Reading Skills	0.95
Comprehension–Knowledge (*Gc*)	0.97	Reading Comprehension	0.94
Long-Term Retrieval (*Glr*)	0.93	Math Calculation Skills	0.94
Visual–Spatial Thinking (*Gv*)	0.85	Mathematics Reasoning	0.97
Auditory Processing (*Ga*)	0.94	Basic Writing Skills	0.95
Fluid Reasoning (*Gf*)	0.97	Written Expression	0.94
Processing Speed (*Gs*)	0.95	Phoneme/Grapheme Knowledge	0.90
Short-Term Memory (*Gsm*)	0.92		
Broad Attention	0.94		
Cognitive Fluency	0.97		
Executive Processes	0.95		
Delayed Recall	0.93		
Knowledge	0.97		
Phonemic Awareness 3	0.95		

Copyright © 2001 by The Riverside Publishing Company. Adapted from the Woodcock-Johnson® III (WJ III™) by Richard W. Woodcock, Kevin S. McGrew and Nancy Mather, with permission of the publisher. All rights reserved.

reported for the adolescent and adult age ranges, factor analysis is typically used to objectively evaluate the structural validity of a battery of ability tests. In this regard, the WJ III structural evidence has the advantage of building on the internal exploratory and confirmatory factor analysis of the 1977 WJ and the WJ-R norm data (McGrew et al., 1991), as well as a series of joint cross-battery analyses with all other major intelligence batteries (see Flanagan & Ortiz, 2001, for the most recent synthesis). The extant structural analyses research supports the CHC broad factor structure of the WJ-R (McGrew, 1994; McGrew et al., 1991). Given that the design of the WJ III extended and refined the previously validated broad CHC ability structure of the WJ-R, confirmatory factor analyses (CFA) studies were used almost exclusively in the evaluation of the structural validity of the WJ III.

Across the entire standardization sample, the WJ III operational CHC measurement model was determined to be the best fitting model when compared to six alternative models (McGrew & Woodcock, 2001). The six alternative models included a null, *g*-only, dichotomous *Gf–Gc*, and three four-factor models that operationalized the intelligence measurement models represented by the PASS, SB-IV, and WAIS-III. The factor loadings for the best fitting CHC models for one adolescent and two adult norm samples are presented in Table 14.9. In all three samples, the WJ III CHC measurement model was the best fitting model (McGrew & Woodcock, 2001). Similar to the WJ-R CFA analyses, the models reported for the oldest samples (ages 40–100) were the poorest fitting models (in an absolute sense). These findings suggest that additional research is needed to determine if alternative models may better represent the CHC structure of cognitive functioning during middle to late adulthood.

The results summarized in Table 14.9 indicate that, in general, most WJ III tests are relatively strong and invariant indicators of the same CHC abilities across the adolescent and adult age ranges. A few exceptions are noted. First, in the *Gc* domain, the Understanding Direction test is a mixed measure of *Gc* (LS, listening ability) and *Gsm* (MW, working memory) during the adolescent and young adult age ranges, but taps more *Gsm* (MW) during middle to late adulthood. Similarly, the Quantitative Concepts test is a consistent mixed indicator of *Gq* (KM, mathematical knowledge) and *Gf* (RQ, quantitative reasoning) during the adolescent and young adult age ranges, with a greater emphasis on *Gf* during middle to late adulthood. Although some age-related changes are observed for a few WJ III reading and writing (viz., Reading Vocabulary and Spelling of Sounds) and *Glr* (Story Recall and Retrieval Fluency) tests, these changes appear unsystematic in nature.

The only *Gv* test that changes noticeably across the three age groups is Planning. The Planning test demonstrates a slight increase in *Gv* (SS, Spatial Scanning) with age. However, it is important to note that this test has more unique than common variance (at all age ranges) and is likely measuring other variables outside the boundaries of the WJ III CHC construct domain (e.g., cognitive style; planning; attention and concentration). With the possible exception of the Auditory Attention test (which appears to increase in *Ga* variance with age), all WJ III *Ga* and *Gf* tests are relatively invariant CHC-domain specific indicators.

The WJ III *Gs* tests demonstrate the most interesting variations with age. In general, the Visual Matching, Decision Speed, and Pair Cancellation tests show increased reliance on *Gs* abilities with increasing age, particularly after age 40. This is particularly noticeable for Pair Cancellation, a test deliberately designed to assess sustained concentration or vigilance. This finding suggests that the Pair Cancellation test may become an increasingly important diagnostic test for detecting changes in cognitive efficiency during the later adult years. In the case of Visual Matching, the increase in *Gs* ability (P, perceptual speed) is associated with a decrease in a minor *Gq* (N, number facility) influence. Finally, Rapid Picture Naming appears to be a relatively moderate indicator of *Gs* (R9, rate-of-test-taking) and a weak indicator of *Glr* (NA, naming facility) during the adolescent and young adult ages, with the *Glr* abilities measured diminishing after age 40. Given the prominent role processing speed (*Gs*) plays in theories of changes in adult intelligence (Bashore, Ridderinkhof, & van der Molen, 1998; Kail, 1991a, 1991b; Kail & Salthouse, 1994; Salthouse, 1996), the findings for Pair Cancellation and Rapid Picture Naming are particularly interesting. Additional research

TABLE 14.9 WJ III broad CHC test factor loadings in adolescent and two adult norm samples

	Broad CHC Factors by Age Groups											
	Gc			*Gq*			*Grw*			*Glr*		
Tests	**14–19**	**20–39**	**40+**	**14–19**	**20–39**	**40+**	**14–19**	**20–39**	**40+**	**14–19**	**20–39**	**40+**
Verbal Comprehension	0.92	0.94	0.95									
General Information	0.89	0.93	0.93									
Understanding Directions (Ach)	0.41	0.30	0.17								0.18	
Picture Vocabulary (Ach)	0.82	0.84	0.89									
Oral Comprehension (Ach)	0.75	0.81	0.79									
Academic Knowledge (Ach)	0.89	0.90	0.93									
Calculation (Ach)				0.85	0.87	0.89						
Math Fluency (Ach)				0.46	0.47	0.41						
Applied Problems (Ach)	0.19	0.23	0.23	0.53	0.57	0.54						
Quantitative Concepts (Ach)				0.51	0.59	0.39						
Letter–Word Identification (Ach)	0.12	0.15	0.12				0.77	0.76	0.80			
Reading Fluency (Ach)							0.45	0.42	0.35			
Passage Comprehension (Ach)	0.48	0.50	0.46				0.32	0.32	0.43			
Word Attack (Ach)							0.73	0.76	0.81			
Reading Vocabulary (Ach)	0.66	0.82	0.63				0.18		0.30			
Spelling (Ach)							0.83	0.85	0.90			
Writing Fluency (Ach)							0.47	0.46	0.52			
Writing Samples (Ach)							0.73	0.78	0.84			
Handwriting (Ach)							0.25	0.25	0.36			
Editing (Ach)							0.75	0.76	0.85			
Spelling of Sounds (Ach)							0.36	0.49	0.32			
Visual–Auditory Learning										0.77	0.71	0.86
Visual–Auditory Learning: DR										0.74	0.68	0.80
Retrieval Fluency										0.31	0.55	0.37
Story Recall (Ach)	0.59		0.41							0.10	0.59	0.28
Story Recall: DR (Ach)	0.57	0.42	0.44									
Memory for Names										0.65	0.62	0.79
Memory for Names: DR										0.61	0.56	0.72

Broad CHC Factors by Age Groups

Gv			*Ga*			*Gf*			*Gs*			*Gsm*		
14–19	20–39	40+	14–19	20–39	40+	14–19	20–39	40+	14–19	20–39	40+	14–19	20–39	40+
												0.35	0.34	0.64
									0.47	0.40	0.51			
						0.25	0.16	0.18						
						0.48	0.42	0.56						
									0.48	0.45	0.54			
									0.36	0.38	0.35			
			0.37	0.26	0.50									
									0.35	0.08	0.40			

(Continues)

TABLE 14.9 (Continued)

Broad CHC Factors by Age Groups

	Gc			*Gq*			*Grw*			*Glr*		
Tests	**14–19**	**20–39**	**40+**	**14–19**	**20–39**	**40+**	**14–19**	**20–39**	**40+**	**14–19**	**20–39**	**40+**
Spatial Relations												
Picture Recognition												
Planning												
Visual Closure												
Block Rotation												
Sound Blending												
Incomplete Words												
Auditory Attention												
Sound Awareness (Ach)												
Sound Patterns												
Analysis–Synthesis												
Concept Formation												
Numerical Reasoning				0.34	0.39							
Visual Matching				0.22	0.21	0.12						
Decision Speed												
Rapid Picture Naming										0.17	0.21	0.08
Pair Cancellation												
Cross Out												
Numbers Reversed												
Auditory Working Memory												
Memory for Words												
Memory for Sentences	0.33	0.42	0.38									
Factor loading on g	0.90	0.91	0.92	0.71	0.66	0.85	0.90	0.91	0.91	0.80	0.95	0.89

NOTE: Italic font designates WJ III Research tests used during the development of the WJ III (McGrew & Woodcock, 2001).
Copyright © 2001 by The Riverside Publishing Company. Adapted from the Woodcock-Johnson® III (WJ III®) by Richard W. Woodcock, Kevin S. McGrew and Nancy Mather, with permission of the publisher. All rights reserved.

is needed to determine if these two WJ III *Gs* tests may have special diagnostic utility (above and beyond the other WJ III *Gs* tests) in the assessment of the effects of aging on cognitive performance.

All WJ III *Gsm* tests are factorially pure and strong indicators of the *Gsm* domain. There is a trend for the Auditory Working Memory (MW, working memory) and Memory for Words (MS, memory span) tests to change slightly in *Gsm* characteristics after adolescence, with the former increasing in *Gsm* and the latter decreasing in *Gsm* starting at age 20.

Broad CHC Factors by Age Groups

Gv			*Ga*			*Gf*			*Gs*			*Gsm*		
14–19	20–39	40+	14–19	20–39	40+	14–19	20–39	40+	14–19	20–39	40+	14–19	20–39	40+
0.71	0.75	0.73												
0.47	0.43	0.55												
0.33	0.43	0.54				0.09								
0.37	0.46	0.55												
0.53	0.65	0.61												
			0.67	0.72	0.66									
			0.51	0.61	0.61									
			0.34	0.44	0.48									
			0.79	0.83	0.85									
			0.35	0.47	0.55									
						0.74	0.80	0.83						
						0.77	0.79	0.79						
						0.56	0.56	0.86						
									0.70	0.71	0.81			
									0.69	0.73	0.78			
									0.48	0.41	0.59			
									0.59	0.73	0.82			
0.24	0.26	0.27							0.66	0.67	0.67			
												0.75	0.71	0.78
												0.73	0.82	0.82
												0.70	0.69	0.64
												0.42	0.38	0.37
0.77	0.79	0.85	0.87	0.96	0.97	0.92	0.92	0.94	0.49	0.51	0.75	0.83	0.86	0.92

Brief Comments on "Good" Factors

As more intelligence test batteries pay closer attention to sampling a greater breadth of narrow abilities within each CHC theoretical construct domain, some of the established "rules of thumb" regarding what constitutes a *strong* or *good* factor loading for a cognitive construct may need to change. Reviewers often relate the strength of a factor (and its indicators) to the absolute magnitude and consistency of the factor loadings. McGrew and Woodcock (2001) argue that, in some situations, excessively high test factor loadings may be a counter indicator of good construct validity.

For example, in the WJ-R, the *Glr* cluster was comprised of the Visual–Auditory Learning and Memory for Names tests. Both tests consistently demonstrated high *Glr* factor loadings. In contrast, the WJ III *Glr* cluster is comprised of the Visual–Auditory Learning and Retrieval Fluency tests. These two WJ III tests demonstrate markedly less consistent *Glr* factor loadings in Table 14.9, with the Retrieval Fluency loading being much lower (.31 to .51) than Visual–Auditory Learning (.71 to .86). Did the WJ-R *Glr* factor (and its two tests) represent a more valid factor than the WJ III *Glr* factor? McGrew and Woodcock (2001) argue that, when *all* forms of validity evidence are combined, the WJ III *Glr* factor is more valid than the WJ-R *Glr* construct. Why?

It is not often recognized that, when indicators of different aspects of a broad construct domain are sampled adequately, a factor with all moderate factor loadings may, in fact, be a more valid factor. This is referred to as the *attenuation paradox* in the reliability literature (Boyle, 1991; Clark & Watson, 1995; Loevinger, 1954), where it has long been recognized that if the inter-item correlation increases among items too much, a test may become very homogeneous and reliable, but at the expense of narrow content coverage that compromises validity. The same principle holds for the correlation of test factor indicators. Maximizing the absolute magnitude of indices of factor homogeneity (e.g., factor loadings) may occur at the expense of factor breadth (construct validity). One may end up with a very narrow factor with test indicators that may tap very similar abilities, rather than a factor that more adequately samples different aspects of the construct domain. Post-WJ-R content validity and hierarchical CFA studies that included narrow and broad factors suggested that the relatively high and tight WJ-R *Glr* factor loadings were reflecting a narrow ability (viz., MA, Associative Memory) and not the intended broad *Glr* ability (Flanagan et al., 2000; McGrew & Flanagan, 1998; McGrew & Woodcock, 2001). McGrew and Woodcock (2001) present logical and empirical evidence (e.g., content validity; hierarchical narrow and broad CFA analyses; differential developmental growth curve trajectories) that supports the divergent WJ III *Glr* factor loadings for Visual–Auditory Learning and Retrieval Fluency as reflecting a more valid factor and set of indicators. A review of the factor loadings in Table 14.9, in the context of other forms of validity evidence, suggests that the WJ III *Glr*, *Ga*, and *Gv* clusters are also exemplars of broader and more valid CHC construct measures.

Structural Validity: Special Study Analysis

The WJ III technical manual describes a study (Gregg/Hoy 1985 study) based on a sample of 204 university students who were administered a wide array of cognitive and achievement measures. Recently, McGrew et al., (2001) subjected the WJ III, WAIS-III, WMS-III, and KAIT test scores from this study to a set of CHC-organized broad and broad plus narrow ability CFAs. The broad and narrow CHC classifications for the most viable broad factor CHC model are presented in Table 14.10.

The pattern of findings summarized in Table 14.10 is nearly identical to those summarized in Table 14.9 and will not be repeated here. The cross-battery CFA results provide additional structural validity evidence for the WJ III tests included in the analysis. A number of findings in Table 14.10 are of particular interest with regard to the structural validity of the WJ III and cross-battery applications (to be discussed later). More detailed discussion and interpretation of this CFA study can be found in McGrew et al. (2001).

First, the modification of four prior WJ-R tests to serve as four *subtests* for the Verbal Comprehension composite test appears to have produced a strong single test indicator of *Gc*. McGrew et al. (2001) report that the WJ III Verbal Comprehension test demonstrated the highest *Gc* factor loading (.85), followed next by the

TABLE 14.10 WJ III, WAIS-III, WMS-III, KAIT broad and narrow classifications based on Gregg/Hoy university sample CFA (McGrew et al., 2001)

Battery	Tests	*Gc*	*Gq*	*Grw*	*Glr*	*Gv*	*Ga*	*Gf*	*Gs*	*Gsm*
WJ III	**Verbal Comprehension**	**LD/VL**								
WAIS-III	Information	K0								
WAIS-III	Comprehension	LD/K0								
WAIS-III	Vocabulary	VL								
WAIS-III	Similarities	LD/K0								
WMS-III	Logical Memory I	LS								
WMS-III	Logical Memory II	ls			mm					
KAIT	Definitions	vl/ld		SG/RD						
KAIT	Double Meanings	VL		v						
KAIT	Auditory Comprehension	LD/LS								
WJ III	**Math Fluency (Ach)**		**A3**							
WAIS-III	Arithmetic		A3							
WJ III	**Letter–Word Identification (Ach)**			**rd**			**pc**			
WJ III	**Reading Fluency (Ach)**			**RS**					**r4/r9**	
WJ III	**Passage Comprehension (Ach)**			**cz**				**rq**		
WJ III	**Visual–Auditory Learning**				**MA**					
WJ III	**Retrieval Fluency**				**fi**				**r4/r9**	
WJ III	*Memory for Names*				*MA*					
KAIT	Rebus Learning				MA					
WMS-III	Family Pictures I				ma	mv				
WMS-III	Family Pictures II				ma					
WMS-III	Spatial Span					mv				ms
WMS-III	Faces I				mm					
WMS-III	Faces II				mm					
WMS-III	Verbal-Paired Associates I				MA					
WMS-III	Verbal-Paired Associates II				MA					
WJ III	**Spatial Relations**					**Vz/SR**				
WJ III	**Picture Recognition**					**mv**				
WJ III	*Block Rotation*					*Vz/SR*				
WJ III	*Visual Closure*					*cs*				
WAIS-III	Picture Completion					CF				
WAIS-III	Picture Arrangement					vz				
WAIS-III	Block Design					SR/Vz				

(Continues)

TABLE 14.10 (Continued)

Battery	Tests	*Gc*	*Gq*	*Grw*	*Glr*	*Gv*	*Ga*	*Gf*	*Gs*	*Gsm*
WJ III	**Sound Blending**						**PC:S**			
WJ III	**Incomplete Words**						**PC:A**			
WJ III	**Auditory Attention**						**us/ur**			
WJ III	*Sound Patterns*						*U3*			
WJ III	**Analysis-Synthesis**							**RQ**		
WJ III	**Concept Formation**							**I**		
WAIS-III	Matrix Reasoning							I		
KAIT	Mystery Codes							I		
KAIT	Logical Steps							RQ		
WJ III	**Visual Matching**								**P**	
WJ III	**Decision Speed**								**R4**	
WJ III	**Rapid Picture Naming**								**na**	
WJ III	*Cross Out*					*ss*			*P*	
WAIS-III	Digit Symbol-Coding								R9	
WAIS-III	Symbol Search								P/R9	
WJ III	**Numbers Reversed**									**MW**
WJ III	**Auditory Working Memory**									**MW**
WJ III	**Memory for Words**									**MS**
WJ III	*Memory for Sentences*	*ld*								*Ms*
WAIS-III	Letter–Number Sequencing									MW
WMS-III	Letter–Number Sequencing									MW

NOTE: Bold font = WJ III tests; italic font = WJ III Research tests (McGrew & Woodcock, 2001). Significant factor loadings have been replaced with the narrow CHC ability classifications proposed by McGrew et al. (2001). See Table 14.1 for the names and definitions corresponding to each narrow ability code/abbreviation. Capitalized narrow ability codes designate significant factor loadings > .49. Lower-case narrow ability codes designate significant factor loadings < .50. Significant residuals were reported between WMS-III Family Pictures I & II, WMS-III Logical Memory I & II, WAIS-III/WMS-III Letter–Number Sequencing; WMS-III Faces I & II; WMS-III Verbal Paired-Associates I & II; WJ III Decision Speed & Rapid Picture Naming; WJ III Visual Matching and Math Fluency; WJ III Math Fluency & WAIS-III Digit Symbol; WJ III Math Fluency and Reading Fluency.

WAIS-III Vocabulary test (.83). The WJ III Verbal Comprehension test has a distinct advantage over the other *Gc* tests reported in Table 14.10 in that it is comprised of four tests that together tap lexical knowledge (VL) and general language development (LD, which subsumes VL). Unfortunately, the WJ III General Information test was not included in this study. Second, consistent with a prior joint WJ-R/KAIT CFA study (Flanagan & McGrew, 1998), two of the KAIT *Gc* tests (Definitions and Double Meanings) were found to be mixed measures of *Gc* and *Grw*. The influence of reading and writing achievement variance on these two *Gc* tests, and the Definitions test in particular (*Grw* loading = .80), although consistent with Carroll's definition of

Gc, suggests that these two *Gc* tests must be interpreted with caution when assessing adolescents and adults with learning difficulties in reading or writing skills.

Third, the results continue to reinforce the conclusion that two of the original 1977 WJ tests (Analysis–Synthesis and Concept Formation) have withstood the test of time as two of the best available tests of *Gf* abilities. The results in Table 14.10 also indicate that the KAIT Mystery Codes and Logical Steps are also excellent tests for the assessment of *Gf* abilities. Fourth, given the historical prominence of the WAIS series in the assessment of adolescent and adult intelligence, it is important to note that the results in Table 14.10 indicate that the long-overdue addition of a *Gf* test (viz., Matrix Reasoning) to the WAIS-III has successfully plugged the WAIS-III *Gf* hole (McGrew, 1997; McGrew & Flanagan, 1996, 1998; Woodcock, 1990), at least with one test. More importantly, the strong loading of the WAIS-III Block Design on *Gv* (.80; McGrew et al., 2001), in the absence of any significant loading on the robust *Gf* factor, adds one more nail to the coffin of the traditional clinical interpretation of WAIS-III Block Design as a measure of "reasoning," particularly abstract fluid reasoning. These results, plus the extant CHC cross-battery factor-analysis research of the Wechsler series of tests (see Flanagan et al., 2001, for the most recent synthesis), should put to rest the notion that the WAIS-III (and WISC-III) Block Design test can be used to draw inferences about *Gf* abilities. The WJ III and KAIT *Gf* tests appear to be much more valid tests for measuring the *Gf* abilities of adolescents and adults.

Fifth, the interpretation of the WJ III Retrieval Fluency test as a good indicator of *Glr* requires the use of a hierarchical model in which Retrieval Fluency and Rapid Picture Naming form a narrow naming facility (NA) factor, which, in turn, has a moderate loading on *Glr*. The broad *Glr* factor also subsumes the narrow abilities of meaningful memory (MM) and associative memory (MA) (see McGrew and Woodcock, 2001, for a detailed explanation). Sixth, the KAIT Rebus Learning (MA, Associative Memory) and many WMS-III tests provide for additional coverage of *Glr*. Two of the WMS-III tests (Family Pictures I and Spatial Span) appear to be factorially complex measures, a situation that clouds their diagnostic interpretation. Seventh, the WJ III is the only major intelligence battery for use with adolescents and adults that includes strong indicators of *Ga*.

Finally, although imposing near-identical task requirements on subjects, the WJ III Auditory Working Memory test (MW, working memory) appears to be a stronger *Gsm* (.81) indicator than the WAIS-III and WMS-III Letter-Number Sequencing tests (with .67 *Gsm* loadings for both) (McGrew et al., 2001). The less complex WJ III *Gsm* tests of memory span (MS, Memory for Words and Memory for Sentences) displayed more moderate factor loadings when compared to the working memory tasks. Of note for the WAIS-III Working Memory Index (a combination of Digit Span, Letter-Number Sequencing, and Arithmetic) is the finding that Arithmetic loads on a *Gq* factor and *not* on the *Gsm* factor with other tests of working memory. These findings suggest that the WJ III Working Memory cluster is a more valid measure of working memory (MW) than is the WAIS-III Working Memory Index. The WAIS-III Working Memory Index is confounded with construct irrelevant *Gq* variance.

External Validity

External validity focuses on the external relations among the focal constructs and their measures and other constructs and/or subject characteristics (AERA, 1999; Benson, 1998). The question posed is "*Do the focal constructs and observed measures fit within a network of expected construct relations (i.e., the nomological network)?*" The adolescent- and adult-specific external validity evidence for the WJ III COG presented in the WJ III technical manual is summarized here. The reader is referred to McGrew and Woodcock (2001) for external validity evidence at other age groups. In addition to summarizing the studies presented in the WJ III technical manual, a recently completed group differentiation analysis with the special Gregg/Hoy LD/non-LD university sample will be presented later in this chapter.

Concurrent Validity: WJ III Ability Cluster Correlations with Achievement and Other Intelligence Batteries

The WJ III GIA scores displayed concurrent correlations in the .70s with the general composite scores across all age samples and instruments (WPPSI-R, WISC-III, WAIS-III, DAS, KAIT, and SB-IV; McGrew & Woodcock, 2001). These studies provide concurrent validity evidence that the WJ III GIA-Std and GIA-Ext clusters are valid indicators of general intelligence, as operationalized by other intelligence batteries.

The single adult-specific concurrent study that presented correlations with the composite scores of other intelligence batteries was the Gregg/Hoy university sample (1985) described previously. Given the selective nature of the two adult subgroups included in this study, the combined sample exhibited significant restriction of range in scores, a situation that dampened the resultant correlations. Thus, the relative comparisons of concurrent correlations are important here, not the absolute magnitude of the correlations. Given this caveat, the .67 and .75 correlations (reported in Table 14.11) with the WAIS-III Full Scale and KAIT Composite IQ, respectively, reinforce the validity of the WJ III COG as a valid measure of general intellectual functioning with adults.

Also presented in Table 14.11 are select correlations that allow for the comparison of the relative concurrent validity of the WJ III, WAIS-III, and KAIT composite scores in the prediction of basic reading, math, and writing achievement. Across and within achievement domains, the WJ III Predicted Achievement and GIA-Standard clusters outperformed both the WAIS-III Full Scale and KAIT composite IQs in the concurrent prediction of achievement. The superiority of the WJ III Predicted Achievement option was again borne out in these data. It is particularly interesting to note that the KAIT demonstrated a median correlation (.51) with achievement much closer to the WJ III GIA-Standard (.56) and WJ III Predicted Achievement (.60) options than did the WAIS-III (.36). It is hypothesized that this is because the KAIT (although only providing *Gf* and *Gc* composite scores) includes a much greater breadth of CHC abilities than does the WAIS-III, but not as great a breadth as does the WJ III. The prior discussion of the joint WJ III, WAIS-III, WMS-III, KAIT CFA study and the Flanagan and McGrew (1998) study support this interpretation. These data suggest that CHC-designed intelligence batteries, and the WJ III in particular, may hold a distinct advantage over the venerable WAIS-III when investigating the academic functioning of adolescents and adults.

A more circumscribed study described in the WJ III technical manual (Norton Study; see McGrew & Woodcock, 2001) involved 50 adults attending a California community college. As part of a study focused on understanding math achievement, the adult subjects were administered a select set of eight WJ III Cognitive tests, five WJ III research tests, and two WJ III math achievement tests. They were also administered select tests from the KAIT and WAIS-III. The specific cognitive tests selected from each battery were those from the CHC ability domains (i.e., *Gf, Gv, Gs, Glr*) that prior research suggested were the most related to math achievement (Flanagan et al., 2000; McGrew & Flanagan, 1998).

In the Norton Community College sample, the simple correlations between all *Gf* tests (WJ III Analysis–Synthesis and Concept Formation; WAIS-III Matrix Reasoning; KAIT Logical Steps and Mystery Codes) and math achievement were consistently higher than the correlations between the *Glr, Gs,* and *Gv* tests and math achievement. Stepwise multiple regression found the combination of the WJ III Concept Formation and Analysis–Synthesis tests (regression

TABLE 14.11 Select WJ III, WAIS-III, and KAIT concurrent validity correlations in LD/Not-LD university sample

Cognitive Measures	Reading Achievement				Writing Achievement			Fluency	
	WJ III Broad Reading	*WJ III Basic Reading Skills*	*WRAT-III Reading*	*Nelson Denny Reading Comp.*	*WJ III Basic Writing Skills*	*OWLS Written Expression*	*WRAT-III Spelling*	*WJ III Academic Fluency*	**Median Correlations**
WJ III General Intellectual Ability—Standard	.56	.56	.61	.61	.62	.49	.47	.48	**.56**
WJ III Predicted Achievement	.66	.62	.60	.60	.63	.52	.49	—	**.60**
Wechsler Adult Intelligence Scale III Full Scale	.35	.39	.43	.47	.38	.36	.27	.27	**.36**
Kaufman Adolescent & Adult Intelligence Composite	.57	.53	.58	.59	.51	.47	.37	.49	**.51**
N	206	185	204	191	200	102	204	205	
Mean	94.7	92.2	107.6	219.1	95.9	93.2	104.5	103.6	
Standard Deviation	12.4	12.9	9.6	23.9	14.0	13.4	13.7	13.3	

NOTE: WJ III General Intellectual Ability Standard correlated .67 and .75 with WAIS-III Full Scale and KAIT Composite scores, respectively.
Copyright © 2001 by The Riverside Publishing Company. Adapted from the Woodcock-Johnson® III (WJ III™) by Richard W. Woodcock, Kevin S. McGrew and Nancy Mather, with permission of the publisher. All rights reserved.

weights = .41 to .36, respectively), followed to a lesser extent by KAIT Mystery Codes (weight = .22), provided the best prediction of adult math achievement (R^2 = .66). This study suggests that the WJ III *Gf* tests of Concept Formation and Analysis–Synthesis may be particularly useful for predicting and explaining mathematics achievement in a postsecondary education setting. Together, the Gregg/Hoy and Norton postsecondary education studies provide support for the concurrent validity of the WJ III COG tests and ability clusters, particularly when compared to other intelligence batteries suitable for this population.

Concurrent Validity: WJ III Ability and Achievement Cluster Correlations

The average concurrent WJ III ability–achievement correlations across the adolescent and adult age ranges (McGrew & Woodcock, 2001) are reported in Table 14.12. A review of Table 14.12 leads to a number of conclusions. First, the three WJ III Cognitive ability options (Predicted

TABLE 14.12 Average (median) correlations between WJ III ability cluster options and WJ III achievement clusters across the adolescent and adult norm samples

	Median Correlations			
Achievement Cluster	**Predicted Achievement**	**General Intellectual Ability (GIA) Standard**	**General Intellectual Ability (GIA) Extended**	**Oral Language**
Reading				
Broad Reading	0.85	0.81	0.84	0.76
Basic Reading Skills	0.79	0.76	0.78	0.74
Reading Comprehension	0.82	0.76	0.79	0.81
Mathematics				
Broad Mathematics	0.77	0.75	0.76	0.67
Math Calculation Skills	0.72	0.69	0.61	0.57
Math Reasoning	0.79	0.77	0.78	0.73
Written Language				
Broad Written Language	0.83	0.81	0.83	0.73
Basic Writing Skills	0.81	0.77	0.79	0.74
Written Expression	0.80	0.78	0.80	0.69
Language & Knowledge				
Oral Language (Std.)	0.77	0.82	0.86	
Oral Language (Ext.)	0.77	0.80	0.83	
Oral Expression	0.67	0.73	0.77	
Listening Comprehension	0.77	0.81	0.83	
Academic Knowledge	0.87	0.77	0.80	0.84

Copyright © 2001 by The Riverside Publishing Company. Adapted from the Woodcock-Johnson® III (WJ III™) by Richard W. Woodcock, Kevin S. McGrew and Nancy Mather, with permission of the publisher. All rights reserved.

Achievement; GIA-Std; GIA-Ext) are consistently better predictors of achievement than the Oral Language ability option. The three WJ III COG ability options are the preferred measures when making predicted/actual achievement comparisons during the adolescent and adult age ranges. Second, as expected, the ability option that uses optimal differential weights of the seven standard COG tests to predict achievement (Predicted Achievements) outperforms a combination of the same seven tests when they are differentially weighted to approximate *g* or general intelligence (GIA). Third, although there is a minor trend for the GIA-Ext cluster to correlate slightly higher with achievement than does the GIA-Std, these differences are not practically significant.

When in-depth diagnostic information is not necessary, the seven-test GIA-Std offers prediction equal to that of the more in-depth 14-test GIA-Ext. Finally, when predictions are necessary regarding an adolescent or adult oral language functioning, the GIA clusters are consistently better predictors. This is not surprising given that the GIA clusters include measures of *Gc*, a construct domain that subsumes oral language. The design of the oral language Predicated Achievement score option eliminates this predictor/criterion overlap by fixing the *Gc* test weights to zero. The decision on which ability score to use when making oral language expected/actual performance comparisons depends on the nature of the specific referral questions (McGrew & Woodcock, 2001).

Group Membership Differentiation: LD/Normal Mean Score Comparisons

The final set of external validity evidence available for the adolescent and adult population was a comparison of test performance of the LD and Not-LD (Normal) subjects in the Gregg/Hoy university sample (McGrew & Woodcock, 2001). Mean score LD/Not-LD comparisons were presented for 16 WJ III COG clusters and 6 WJ III ACH clusters. Even after the application of the Bonferroni adjustment to control for overall experiment-wise error rate, all but 3 of the 22 *t*-tests were significant at the .05 level of significance. As would be expected, given the prominence achievement plays in the identification and classification of individuals with learning disabilities, the largest mean score differences were on five of the six achievement clusters. With the Normal subject mean scores ranging primarily in the average to above average ranges (98.2 for Basic Reading Skills to 112.0 for Academic Fluency), the LD subjects scored approximately one standard deviation lower on Basic Writing Skills (–17.8), Academic Fluency (–17.3), Broad Reading (15.8), Basic Reading Skills (–14.0), and Phoneme/Grapheme Knowledge (–13.5). Oral Expression scores were not significantly different.

The LD subjects were also –11.8 points lower on the GIA-Std cluster. The largest differences on the cognitive clusters occurred in domains related to the efficiency of cognitive processing and *Ga* abilities. In particular, the largest mean score differences were noted for the Cognitive Efficiency (–11.8 for Standard and –10.4 for Extended), Auditory Processing (–11.7), Phonemic Awareness (–11.3), and Working Memory (–11.1) clusters. The only cognitive clusters that did not differentiate the two groups were the Long-term Retrieval (*Glr*), Visual–Spatial Thinking (*Gv*), and Cognitive Fluency (*Gs*) clusters. Collectively, the mean score comparisons suggest that the WJ III COG and ACH batteries provide useful information for the differentiation of adult university subjects with and without learning disabilities. These largely descriptive findings suggest that adult university students with learning disabilities, as a group, are characterized by: significantly lower achievement (across all areas); lower general intellectual functioning: and relatively larger specific cognitive deficits in the auditory processing domains and the efficiency of cognitive processing (particularly working memory). More refined analyses of these data with specialized group-differentiation methodology are described in the next section of this chapter.

Special Applications and Use with Special Populations

Assessment of Learning Disabilities

The WJ III includes a variety of cluster scores and interpretive options that can be useful in the assessment and identification of learning disabilities (LD). As described previously in this chapter, the breadth of measures included in the WJ III provides diagnosticians with a large theory-based tool chest for surveying many of the cognitive and achievement abilities associated with comprehensive LD assessments. The WJ III also provides a set of discrepancy-based interpretive features that facilitate the norm-based identification of within-person cognitive and achievement strengths and weaknesses, a practice that has been at the core of LD assessment and identification since the field's inception. Three different discrepancy models for evaluating the strengths and weaknesses between and among WJ III cluster scores are presented. These discrepancy models are then followed by the presentation of preliminary research that identifies potentially important WJ III clusters for the identification of adults with LD.[3]

General Intellectual Ability–Achievement Discrepancy Model

Ninety-eight percent of all states, and most federal agencies servicing adolescents and adults, have followed the lead of federal law and have incorporated the notion of a learning disability being defined by a discrepancy between a person's actual (measured) and expected achievement (usually predicted from general intellectual ability) (Flanagan et al., 2001). The WJ III provides reliable and valid procedures that can contribute useful information for use in these procedures. The GIA Ability–Achievement Model and the Predicted Achievement Model are both presented in Figure 14.3.

It is important to note that *neither the WJ III GIA or PA discrepancy procedures were designed for the diagnosis of LD*, if the intent is to identify a specific learning disorder. The GIA and PA approaches are intended to answer the question, "*Given the person's present cognitive abilities, is he/she achieving as well as could be expected?*" The WJ III GIA Model is straightforward. Either the differentially weighted WJ III GIA-Std or GIA-Ext *g* standard scores are used to provide a prediction of what a person's achievement standard score would be, given their level of general intellectual ability and age or grade. As portrayed in Figure 14.3, the individual's expected achievement (e.g., predicted Basic Reading Skills standard score) is then subtracted from the person's actual Basic Reading Skills standard score, producing an ability–achievement standard score discrepancy. The WJ III is unique with respect to three major features of this model.

- The WJ III provides a true *g* score for use in the ability–achievement calculations. Other major intelligence batteries rely on an arithmetic average of test scores, an average that implies an equally weighted general intelligence score. Thus, a more theoretically sound general ability index is used in the WJ III GIA ability–achievement discrepancy procedures.
- The WJ III predicted (sometimes referred to as "expected") achievement score accounts for regression-to-the-mean in a manner that captures the developmental changes in ability–achievement correlations. The regression-to-the-mean effect is greatest for predicted achievement scores that diverge the farthest

[3]It is important to note that space limitations do not allow a detailed explanation of all caveats related to the use of the different WJ III discrepancy procedures in the "art and science" of LD decision making and classification. Comprehensive models that encompass a broader array of variables are required. A particularly interesting comprehensive model grounded in the CHC theory has recently been outlined by Flanagan, Ortiz, Alfonso, and Mascolo (in press).

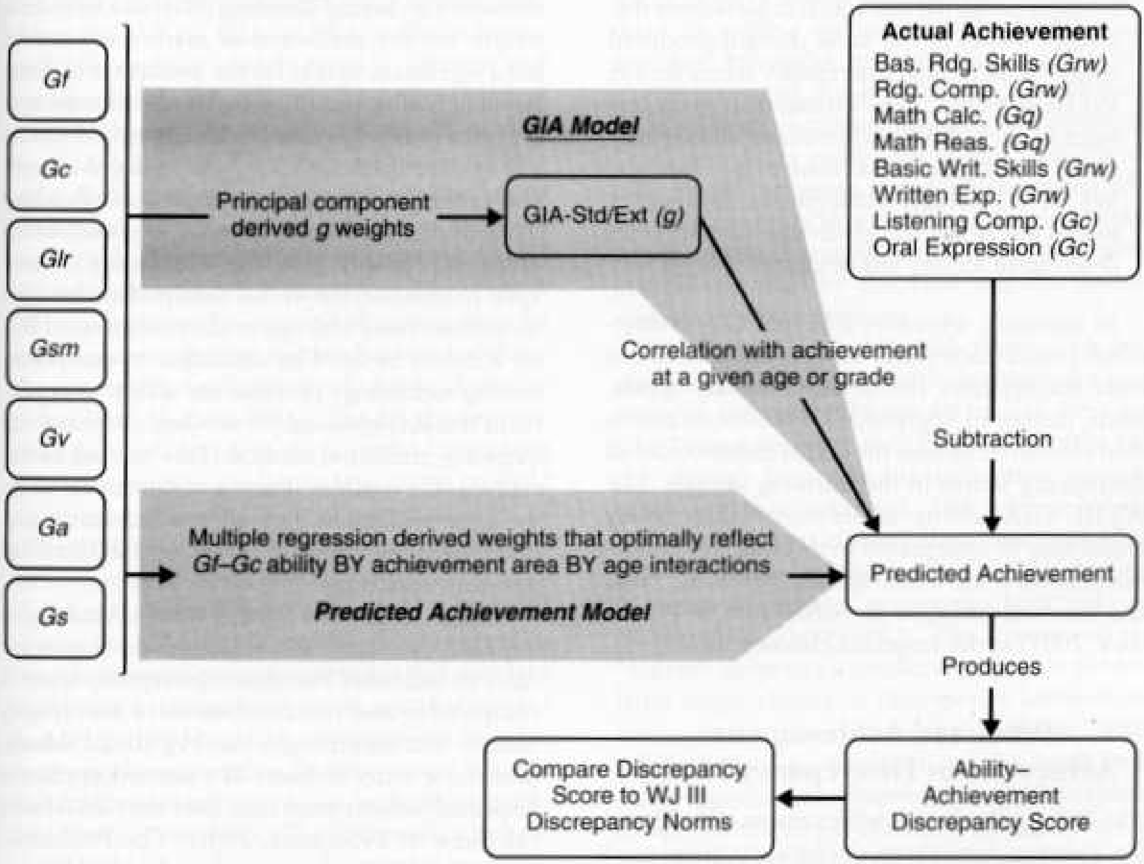

FIGURE 14.3

WJ III GIA (*g*) and predicted achievement ability–achievement discrepancy

from the mean. This occurs because the correlation between any combination of ability and achievement measures is less than perfect.

- The WJ III capitalizes on the co-norming of the WJ III COG and ACH to provide slightly different regression adjustments across age or grade when generating the predicted achievement score. The developmental changes in the ability–achievement correlations observed in the WJ III norm data are incorporated into the prediction equations. A common practice, particularly when an intelligence test is not co-normed with the administered achievement test, is to select a single point correlation (e.g., .65) and use this value to correct for regression effects via a formula for all ages. The predicted score in the WJ III GIA Ability–Achievement Model is based on a developmentally sensitive regression-to-the-mean adjustment procedure.
- The WJ III GIA ability–achievement discrepancy scores are compared against *real* discrepancy norms (end box in Figure 14.3; McGrew, 1994; McGrew et al., 1991; McGrew & Woodcock, 2001; Mather & Schrank, 2001; Woodcock, 1978). Predicted achievement scores in all relevant achievement domains were generated for all norming subjects. Each

subject's predicted and actual achievement difference scores in the same domain produced ability–achievement discrepancy scores for the WJ III norm file. The distributional characteristics of each ability–achievement discrepancy score distribution were used to produce the WJ III GIA ability–achievement discrepancy norms just as age- or grade-specific norms are determined for any test or cluster.

In summary, whenever a WJ III GIA ability–achievement discrepancy score is calculated, this score incorporates the developmentally appropriate degree of regression-to-the-mean and is then compared against the actual distributions of discrepancy scores in the norming sample. The WJ III GIA ability–achievement discrepancy scores may be interpreted in three different metrics (standard score discrepancy, percentile rank, and standard deviation units; McGrew & Woodcock, 2001) as the examiner chooses.[4]

Predicted Achievement–Achievement Discrepancy Model

The WJ III Predicted Achievement (PA) Model for ability–achievement discrepancy calculation is also portrayed in Figure 14.3. A similar model was present in the WJ and WJ-R in which the differential Scholastic Aptitude clusters were used as the predictor measures. The Scholastic Aptitude clusters were used to provide predicted achievement scores based on the best combination of four cognitive tests that predicted different achievement domains (McGrew, 1986, 1994; McGrew et al., 1991; Woodcock, 1978).

As portrayed in Figure 14.3, in the WJ III PA Model the standard seven WJ III COG tests are placed into equations that differentially weight each tests' contribution to the prediction of the target achievement domain. Not only do the weights for the same test differ by achievement domain (e.g., Sound Blending [*Ga*] has a near zero weight for the prediction of math achievement but a significant weight for the prediction of Basic Reading Skills), but the weights also change systematically as a function of developmental status. For example, although Sound Blending is important in the prediction equation for Basic Reading Skills during the formative years, its contribution drops appreciably past the elementary school ages. In contrast, the Verbal Comprehension (*Gc*) weight increases with age in the prediction of Basic Reading Skills. The utilization of computer-scoring technology provides the ability to implement this *developmental/purpose-focused/optimal test weighting* prediction method. This method better captures the complex nuances of human development via predictions that reflect achievement domain, by developmental status, by CHC ability domain interactions.

Similar to the GIA Model, the PA Model also implicitly accounts for regression-to-the-mean. Also, an individual's resultant discrepancy score is compared against real distributions of discrepancy norms. Not surprisingly, the WJ III PA Model provides a better estimate of a person's predicted (expected) achievement than does the GIA Model (McGrew & Woodcock, 2001). The PA Model optimally weights the seven standard COG tests to "wring out" as much variance as possible in the prediction of achievement. In comparison, the WJ III GIA score is developed to "wring out" as much general intelligence variance as is possible from the 7- or 14-test combinations of tests; optimal weighting for the prediction of achievement is not included in the differential GIA *g*-weighting.[5]

[4]An additional ability–achievement discrepancy procedure that is identical to the GIA Model, with the exception being the substitution of the WJ III ACH Oral Language (*Gc*) cluster as the ability measure, is not described here.

[5]Although true, as described by McGrew and Woodcock (2001), one of the major criteria used to select tests for the WJ III COG-Std was which respective CHC test was a better predictor of achievement. For example, both the Analysis–Synthesis and Concept Formation *Gf* were found to be equally strong indicators of *Gf*. Concept Formation was found to be a slightly better predictor of achievement across all domains and ages and, therefore, was selected to be the featured *Gf* test in the WJ III COG-Std. This insured that the GIA-Std score, although not weighted to best predict achievement, included those tests from each CHC ability domain (when all other psychometric factors were judged to be relatively equal) that best predicted achievement.

Given the different design goals and philosophies, the WJ III GIA and PA ability–achievement discrepancy models provide different and complementary information. The calculation of ability–achievement discrepancies with either the GIA-Std or GIA-Ext may be useful when a generalized measure of cognitive functioning or intelligence is required for eligibility purposes (Schrank & Mather, 2001). In contrast, the WJ III PA option is intended to determine if a person is performing as well as one would expect, *given his or her measured levels of associated cognitive abilities*, not necessarily to diagnose a learning disability (Schrank & Mather, 2001). The PA discrepancy procedure will be particularly useful for making the most accurate predictive statements possible concerning an individual's anticipated levels of current achievement. It should be informative for setting short-term goals.

Because the strong PA prediction is achieved via the inclusion and higher weighting of certain tests that measure cognitive abilities that may be a significant weakness for a person, and that may reflect an intrinsic cognitive or processing disorder, it may not be appropriate (in many cases), nor was it ever intended to be used, for determining a specific learning disability. The PA's predict how an individual will perform, on the *average*, in a variety of situations requiring that particular subset of abilities, otherwise known as "aptitude."[6]

Intra-Ability Discrepancy Model

Unique to the WJ III are three types of norm-based within-person discrepancy score procedures that have the potential to better identify the unique patterns of cognitive abilities and achievements in individuals with LD. Collectively the intra-cognitive, intra-achievement, and intra-individual (cognitive and achievement combined) discrepancies are called the *intra-ability discrepancies*. The intra-ability discrepancies allow examiners to analyze an individual's cognitive and academic strengths and weaknesses across the cluster scores of the WJ III COG and WJ III ACH. These discrepancies, and the combined COG and ACH procedure in particular (intra-individual), can assist in the identification of a learning disability by providing information that complements (but does not supplant) the information provided by the GIA and PA ability–achievement discrepancies.

The logic of intra-ability discrepancies is simple. Using the intra-cognitive discrepancies as an example, the standard score for each of the seven CHC cognitive clusters is first isolated (the target cluster) from the six remaining clusters, which are then averaged. The average of the "other" cognitive clusters then functions similarly to the ability cluster in the GIA Ability–Achievement Model previously described and presented in Figure 14.3. The average of the "others" generates a predicted score for the isolated target cluster. A discrepancy between the target cluster standard score and the predicted standard score is calculated and the result compared against the distribution of discrepancy norms for the target cluster. The same discrepancy evaluation scores (percentile rank and standard deviation units) provided for the GIA and PA Models are used to interpret the importance of the intra-cognitive discrepancy. The intra-achievement and intra-individual (achievement and cognitive clusters combined) discrepancy norms are derived from similar procedures.

Schrank and Mather (2001) believe that the intra-individual discrepancy procedure may be particularly useful in the identification of a specific learning disability when the examiner needs to determine what is "specific" about the problem. The intra-individual procedure is conceptually similar to recent recommendations to identify an individual with a learning disability via the evaluation of domain-specific achievement skills conjointly with their related cognitive abilities (Brackett & McPherson 1996; Fletcher et al., 1998). For example, an adolescent referred for long-standing problems with math

[6]The term *aptitude* has come to be misunderstood in much of psychological practice. Snow (1991) provides an excellent summary of the history of how the original connotation has changed (for the worse) over time. The term *aptitude* is used here in the classic sense as described by Snow.

who demonstrates relative intra-individual weaknesses (less than -1 *SD*) on the WJ III ACH Math Calculation and Math Reasoning clusters with concurrent deficits on the WJ III COG Working Memory (*Gsm*), Long-term Retrieval (*Glr*), Processing Speed (*Gs*), and Fluid Reasoning (*Gf*) clusters would be exhibiting a constellation of deficits consistent with a domain-specific disability in mathematics (Carroll, 1996; Geary, 1993; Geary, Hamson, & Hoard, 2000; Geary, Hoard, & Hamson, 1999). The intra-individual discrepancy procedure can be used with several combinations of clusters depending on which sets of WJ III COG and ACH tests are administered.

Preliminary Research on Adult LD Identification with the WJ III

Preliminary research evidence on potentially important diagnostic patterns of WJ III scores in adult university subjects was recently extracted from the Gregg/Hoy LD/Normal university sample previously described in this chapter. The results, briefly summarized here, demonstrate the potential of the WJ III in the identification of adolescents and adults with or without learning disabilities.

Select WJ III COG and ACH data from the Gregg/Hoy university sample was subjected to the *Classification and Regression Tree (CART)* program, a robust set of decision-tree procedures for "data mining" and predictive modeling (Salford Systems, 1999, 2000). Briefly, CART uses computer-intensive and complex data-searching algorithms to identify important patterns and relations in data. CART can uncover hidden structure in very large and highly complex data, even data that may be difficult to analyze with traditional statistical methods (e.g., when a set of variables are highly multicollinear).

Using the dependent categorical variable of LD versus Not-LD (Normal), the CART analyses "grew" a large decision-making tree that resulted in the optimal classification of 204 subjects with scores from the complete set of WJ III COG and ACH clusters. The initial LD/Not-LD decision-tree was set aside by CART and a tenfold internal cross-validation procedure produced 10 new independent trees that were then combined and used to "prune" the original tree. The complete sample was then classified based on the final pruned tree. The results of these analyses, which identify two Normal (Not-LD) and three LD classification groups (called terminal nodes), are presented in Figure 14.4.

The first decision point in the tree indicates that the WJ III Academic Fluency cluster is the single most important variable in differentiating LD from Not-LD university students in this sample. Subjects with WJ III Academic Fluency scores greater than 107 are most likely not LD. Terminal Node 5 includes 81 normal subjects and 13 LD subjects. Thus, this first decision rule results in 86.2% of these 94 subjects correctly classified as Not-LD, and 13.8 % of the LD subjects misclassified (included in Terminal Node 5). Conversely, a WJ III Academic Fluency score less than or equal to 107 produces three other decision-making points and four other terminal nodes (Nodes 1 through 4). Nodes 1 through 4 have classification accuracy figures ranging from 81.0% (Node 2) to 95.4% (Node 1). The complete decision tree indicates the following for university subjects:

- A subject with a WJ III Academic Fluency score greater than 107 is most likely not previously diagnosed as LD. This classification was accurate 86.2% of the time.
- A subject with a WJ III Academic Fluency score less than 107 and a WJ III Basic Writing Skills cluster score less than or equal to 96 is most likely LD (95.4% accuracy).
- A subject with a WJ III Academic Fluency score less than or equal to 107, a WJ III Basic Writing Skills cluster score greater than 96, and a WJ III Verbal Ability-Std cluster (same as the Verbal Comprehension test) score less than or equal to 104 is most likely LD (81.0% accuracy).

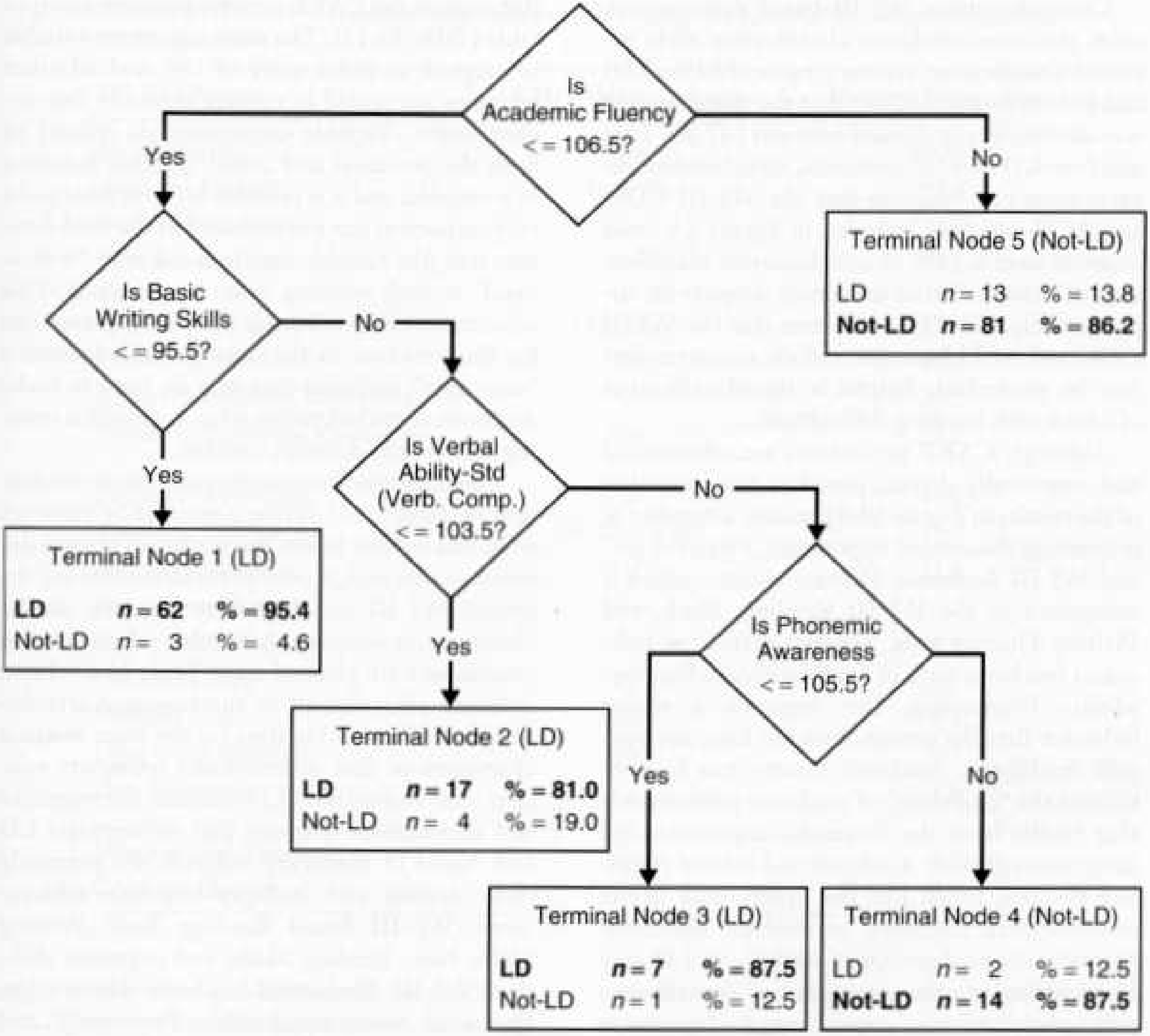

FIGURE 14.4

WJ III university subject LD/Not-LD classification and decision tree (Gregg/Hoy study, 2001)

- A subject with a WJ III Academic Fluency score less than or equal to 107, a WJ III Basic Writing Skills cluster score greater than 96, a WJ III Verbal Comprehension test score greater than 104, and a WJ III Phonemic Awareness (*Ga*) cluster score less than or equal to 106 is most likely LD (87.5% accuracy).
- Finally, a subject with a WJ III Academic Fluency score less than or equal to 107, a WJ III Basic Writing Skills cluster score greater than 95.5, a WJ III Verbal Comprehension test score greater than 104, and a WJ III Phonemic Awareness (*Ga*) cluster score greater than 106 is most likely not LD (87.5% accuracy).

Using the above WJ III-based decision-tree rules, the cross-validation classification table revealed classification accuracy rates of 81% (LD) and 85% (Normal). Given that the initial sample was almost equally divided between LD (n = 101) and Not-LD (n = 103) subjects, this classification agreement rate suggests that the WJ III COG and ACH variables included in Figure 14.3 can improve over a 50% chance base-rate classification of newly referred university subjects by approximately 30%. This indicates that the WJ III COG and ACH batteries include measures that may be particularly helpful in the identification of adults with learning difficulties.

Although CART procedures are atheoretical and empirically driven, post-hoc interpretation of the results in Figure 14.4 presents a number of interesting theoretical hypotheses. First, the pivotal WJ III Academic Fluency cluster, which is comprised of the WJ III Reading, Math, and Writing Fluency tests, suggests that, if an individual reaches a state of automaticity in basic academic functioning, this may be a strong indicator that the person does not have any specific disabilities. Academic fluency can be considered the "end state" of academic performance that results from the successful acquisition and integration of basic academic and related cognitive abilities, much like the expert state in the novice/expert cognitive psychology literature. Second, three subgroups of adults with LD may be identified via the inspection of domain-specific constellations of cognitive and achievement abilities. One group (Terminal Node 1) may be subjects with poor automaticity in general academic functioning and low basic skills in writing, with no apparent associated cognitive deficits. The second group (Terminal Node 2) may display poor academic automaticity associated with a weakness in general verbal knowledge or comprehension (*Gc*). The third group also displays poor academic automaticity, but, instead of relative *Gc* deficits, they have associated cognitive problems in phonemic awareness (*Ga*).

An additional source of useful information regarding the WJ III COG and ACH clusters in this study is the CART *variable importance* output table (Table 14.13). The most important variable is assigned an index score of 100, and all other variables are scaled in relative terms to this anchor point. Variable importance is related to both the potential and actual splitting behavior of a variable, and it is possible for a variable to be very important but not included in the final decision tree (the variable may be a constant "bridesmaid" at each splitting point in the data). This information helps identify possible explanations for the structure in the data and also identifies "surrogate" variables that may be used to make decisions at critical points when a subject is missing data on the decision variable.

The most obvious conclusion from an inspection of Table 14.13 is that a number of variables were lurking just below the surface at critical decision points and, if used as replacements for the critical WJ III variables, may produce similar classification accuracy. A number of substantive conclusions are gleaned from Table 14.13. First, deficient achievement in the language arts domains (reading and writing) is the most obvious characteristic that differentiates university subjects with and without LD. Second, the cognitive and achievement domains that differentiate LD and Not-LD university subjects are primarily those dealing with auditory–linguistic achievement (WJ III Broad Reading, Basic Writing Skills, Basic Reading Skills) and cognitive abilities (WJ III Phoneme/Grapheme Knowledge, Phonemic Awareness, Auditory Processing), and efficient/automatic cognitive and achievement functioning (WJ III Academic Fluency, Cognitive Efficiency, Processing Speed, Short-Term Memory, and Working Memory). This information, in addition to the practical decision-rule tree, suggests that the WJ III COG and ACH clusters that tap these abilities should receive significant attention when evaluating adults for possible learning difficulties. The high level of accuracy achieved via these analyses is most likely a function of the power of the CART procedures combined with a battery of cognitive and achievement measures (WJ III) that cover a wide

TABLE 14.13 WJ-III relative importance variable ratings for LD/Not-LD university sample CART analysis

WJ III Cluster	CHC Ability Domain	Relative Importance
Broad Reading	*Grw*	100.0
Academic Fluency	*Grw, Gq*	97.8
Basic Writing Skills	*Grw*	92.1
Basic Reading Skills	*Grw*	78.5
Cognitive Efficiency-Std	*Gs, Gsm*	37.1
Processing Speed	*Gs*	27.3
General Intellectual Ability-Std	*g*	27.1
Phoneme/Grapheme Knowledge	*Grw*	23.3
Phonemic Awareness	*Ga*	19.8
Thinking Abilities-Std	*Glr, Gv, Ga, Gf*	12.7
Auditory Processing	*Ga*	11.6
Short-Term Memory	*Gsm*	9.7
Verbal Ability-Std	*Gc*	9.3
Fluid Reasoning	*Gf*	8.7
Working Memory	*Gsm*	8.4
Long-Term Retrieval	*Glr*	7.3
Oral Expression	*Gc*	5.6
Visual–Spatial Thinking	*Gv*	0.0
Cognitive Fluency	*Gs, Glr*	0.0

and theoretically valid breadth of human abilities. These results also demonstrate the strong research potential of the WJ III.

Neuropsychological Applications

One distinguishing characteristic of neuropsychological assessment is its emphasis on the identification and measurement of psychological deficits.... Neuropsychological assessment is also concerned with the documentation and description of preserved functions—the patient's behavioral competencies and strengths. (Lezak, 1995, p. 97)

Neuropsychological assessment is concerned with evaluating brain–behavior relations. Historically, there have been two approaches to neuropsychological assessment. The *quantitative* (structural) approach has focused on the development of psychometric test batteries that allow for the identification of aberrant neurological conditions. In contrast, a more *qualitative* approach, often associated with Luria (1966), has focused more on "pathognomonic signs" (Dean & Woodcock, 1999). Regardless of the research tradition, contemporary neuropsychological assessment has been largely an atheoretical approach that has employed an eclectic mix of measurement tools that have been normed at different dates and with different subjects (Dean & Woodcock, 1999; Flanagan et al., 2000; Wilson, 1992).

This section provides a brief glimpse at current and future applications of the WJ-R/WJ III to neuropsychological assessment. Although the WJ III does not cover all aspects required for a comprehensive neuropsychological evaluation, it provides more coverage for the assessment and description of deficits and preserved neurocognitive functions than any other single nationally standardized and norm-referenced source battery. Furthermore, the WJ III provides a coherent and empirically based framework (viz., CHC theory) that can be used to ground contemporary neuropsychological cognitive assessment and interpretation. The WJ III also provides for evaluation of performance on a common integrated set of norms (Dean & Woodcock, 1999).

In this section, the WJ III CHC-classified tests are cross-referenced to a more traditional neuropsychological taxonomy that includes the constructs of attention, visual and auditory perception/processing, memory and learning, language, reasoning and problem solving, and academic achievement. The breadth of coverage and wide age range of the WJ III make it particularly suited for use as a neuropsychological instrument, either as the primary battery or as a resource of supplemental measures. Efforts are currently underway to integrate the WJ-R and WJ III batteries into a comprehensive neuropsychological assessment system that would also include measures of sensory and motor functioning, a structured interview, and a mental status exam. The *Dean-Woodcock Neuropsychological Assessment System (D-WNAS)* (Dean & Woodcock, 1999) uses the WJ-R/WJ III to integrate CHC and neuropsychological constructs within an information-processing framework.

Several features of the WJ III enhance its usefulness as a neuropsychological instrument:

- The wide age range of application spans from 2 to 90+ years.
- All tests are normed and used across almost the entire life span.
- All major areas of cognitive functioning and academic achievement can be assessed without the need to "cross" batteries.
- The wide range of item difficulty within each test allows the documentation of strengths and superior performance as well as deficits.
- To foster standardized administration, critical auditory stimuli tests (e.g., auditory memory span, phonological processing, and listening comprehension) are presented by recorded audiotape. The WJ III COG is the only norm-referenced intelligence battery to provided co-normed measures of auditory processing.
- The selective or focused testing principle allows any single test or combination of tests to be selected for use and interpreted on a common norm base.
- Special college- and university-level norms are provided that differentiate performance by type of postsecondary education institution.
- Significance of any aptitude–achievement, intra-cognitive, or intra-achievement discrepancies is reported and is based on real discrepancy norms.
- Computerized scoring and a narrative report program are available.
- Parallel test batteries are available in English and Spanish.
- Equivalent forms of the achievement battery are available for situations requiring frequent retesting.

Application of the WJ III to Neuropsychological Assessment

Why should a practicing neuropsychologist spend time studying new models of cognitive abilities and, in particular, CHC theory? As outlined earlier in this chapter, a primary reason is that the CHC model represents the best of current research into the structure of intellect. A second reason is that the CHC organization offers an empirically derived taxonomic classification of cognitive abilities that is characterized by a high level of functional independence among the categories. Traditional ability classifications used in neuropsychology have evolved primarily from clinical practice and have broadly defined areas of special interest. As a result, the classification constructs sometimes overlap and, at other times, may be a mix of two or three distinctly different types of functions as defined by current cognitive science. The efforts of Dean and Woodcock (1999) coincide with other recent attempts to integrate neuropsychological assessment paradigms with CHC theoretical constructs (Pallier, Roberts, & Stankov, 2000).

Toward the goal of urging the integration of CHC theory into current neuropsychological thought, the following material describes how the CHC-organized WJ III can measure deficits and preserved functions in the context of a traditional neuropsychological organization. Table 14.14 illustrates the relationships between these two systems of classifying neurocognitive functions. As presented in Table 14.14, the WJ III provides measures across a wide spectrum of the traditional neuropsychological functional categories. The reader will note that several tests appear in more than one category of functions. This reflects the less precise and largely atheoretical nature of a traditional neuropsychological classification.

Assessment of Attention (*Gs, Gsm, Ga, Gf*)

During the past several decades, researchers have recognized that attention is not a unitary construct. This has resulted in the specification of multidimensional models of the construct of attention (Carroll, 1993; Lezak, 1995; Mirsky, Anthony, Duncan, Ahearn, & Kellam, 1991). The WJ III does not measure all of the important aspects of this broad construct; however, four aspects—selective attention, divided attention, sustained attention, and attention capacity—are measured by eight WJ III tests listed in Table 14.15.

Briefly, *attentional capacity* is the ability to hold information in immediate awareness while performing some action on the information. The task requirements (see Table 14.4) of the WJ III Numbers Reversed test, and to a lesser degree the Memory for Words test, suggest these two tests may shed light on an individual's attentional capacity. *Sustained attention,* or the capacity to stay on task in a vigilant manner, is measured by Pair Cancellation. The WJ III Pair Cancellation test requires a subject to rapidly identify a certain instance of a repeated pattern under conditions in which he or she must maintain a constant focus on the target condition in the presence of similar distracting stimuli. The clerical speed ability required by Visual Matching may also tap sustained attention, but under conditions that require less vigilance.

The WJ III Auditory Attention test, a test that requires making simple sound discriminations under increasing stimulus background noise distor-

TABLE 14.14 Comparison of traditional neuropsychological ability with CHC broad ability categories

	CHC Broad Abilities								
Traditional Neuropsychological Categories	***Gsm***	***Gs***	***Gc***	***Gq***	***Grw***	***Gv***	***Ga***	***Glr***	***Gf***
Attention	•	•					•		•
Visual Perception/Processing						•			
Auditory Perception/Processing							•		
Memory and Learning	•		•			•		•	
Language			•						
Reasoning and Problem Solving				•					•
Academic Achievement			•	•	•				

TABLE 14.15 Attention dimensions tapped by eight WJ IIII COG tests

		Attention Dimension			
WJ III Test	**Type of Stimuli**	**Selective Attention**	**Shifting/ Divided Attention**	**Vigilance/ Sustained Attention**	**Attentional Capacity**
Auditory Attention (*Ga*)	Auditory	•			
Auditory Working Memory (*Gsm*)	Auditory		•		
Concept Formation (*Gf*)	Visual		•		
Pair Cancellation (*Gs*)	Visual	•		•	
Decision Speed (*Gs*)	Visual		•		
Visual Matching (*Gs*)	Visual			•	
Numbers Reversed (*Gsm*)	Auditory				•
Memory for Words (*Gsm*)	Auditory				•

NOTE: Bold font designates tests comprising the WJ III Broad Attention cluster.

tion, measures *selective attention*, or the ability to focus attention when distracting stimuli are present. The sheer mass and repeated occurrences of the same visual stimuli on the Pair Cancellation test also may require the "selective" filtering of relevant and irrelevant stimuli for successful performance. Finally, the WJ III Auditory Working Memory test requires an individual to retain and rearrange information placed in short-term memory to form two distinct sequences (a form of mental "juggling"). The Auditory Working Memory and Numbers Reversed tests may also be combined to form a *Working Memory* cluster. This cluster measures the ability to hold information in immediate awareness while performing a mental operation on it.

Assessment of Visual Perception/Processing (Gv)

Visual–spatial measures have had a long and prominent history in neuropsychological assessment. Visual perceptual–spatial skills are part of everyday life and enable individuals to receive, process, integrate, and synthesize information that is seen or manipulated "in the mind's eye." As presented in Table 14.14, the WJ III provides three measures of visual perception–processing. These include Spatial Relations (Vz, visualization), Picture Recognition (MV, visual memory), and, to a lesser extent, Planning (SS, Spatial Scanning).

Assessment of Auditory Perception/Processing (Ga)

Neuropsychologists have long recognized the importance of assessing auditory perception/processing. Auditory processing involves the ability to perceive, discriminate, process, and synthesize both speech and nonspeech sounds. The WJ III is the only intelligence battery to provide for comprehensive assessment of certain aspects of the auditory domain recognized by neuropsychologists. The Sound Blending and Incomplete Words tests are measures of phonological awareness or phonemic knowledge (PC, or phonetic coding as per the CHC taxonomy). A third WJ III COG test, Auditory Attention, is a measure of the ability to discriminate speech sounds (US) under distracting conditions (UR). The WJ ACH battery also includes the Sound Awareness test, a *Ga* test that may prove to be a particularly good *Ga* screener because of the diversity of auditory skills required (e.g., sound deletion, substitution, and rhyming).

Assessment of Memory and Learning (*Gsm, Gv, Glr, Gf,* and *Gc*)

Memory and learning tests constitute the broadest category of tests in a traditional neuropsychological classification. From the perspective of CHC theory, three factorially distinct cognitive abilities (*Gsm, Glr, Gc*) fall within this broad category. The clinical assessment of memory deficits typically involves evaluation of the ability to actively learn and remember new material presented in both auditory and visual modalities. The adequacy of both short-term memory (immediate recall) and long-term retention (delayed recall) are typically assessed. Indexes of remote memory may also be helpful with persons of advanced age and other clinical populations. Eleven WJ III tests are identified as good measures of various aspects of memory or learning. Tests of auditory short-term memory (*Gsm*) include Numbers Reversed (MW), Auditory Working Memory (MW), and Memory for Words (MW). Picture Recognition (*Gv*) is an indicator of immediate visual recall (MV).

Three other tests are identified as measures of long-term retrieval (*Glr*). Visual–Auditory Learning is a visual–auditory associational learning task. The task requires learning new material with corrective feedback provided whenever the examinee makes an error. There is a delayed recall version of the test, Delayed Recall-Visual–Auditory Learning, which measures the ability to recall, from 30 minutes to 8 days later, the just-learned associations. This test is among the few clinical memory tests that include standardized and normed delay procedures extending more than 24 hours beyond initial administration. Retrieval Fluency (*Glr*) is a measure of ideational fluency (FI) or the ability to fluently recall related items from memory within a short time. Finally, the Story Recall and Delayed Recall-Story Recall tests from the WJ III ACH battery can provide information regarding meaningful memory (MM).

Two other WJ III tests can be characterized as new learning tasks. Concept Formation (*Gf*-I) and Analysis–Synthesis (*Gf*-RG) are controlled learning paradigms that both require learning a series of procedures to solve inductive and deductive logic problems. Corrective feedback for errors and reinforcement for correct responses are provided, two essential characteristics of many real-world learning situations. These two tests will be identified again in the discussion of Reasoning and Problem-Solving tasks.

Some neuropsychologists include tests of learned or acquired information (sometimes called long-term memory or remote memory) among their assessment procedures. The WJ III General Information (*Gc*-K0) and Academic Knowledge (K0, K1, K2, A5) tests are of this type.

Assessment of Language (*Gc*)

The ability to communicate through language is typically assessed through an examination of both receptive and expressive language. The three traditional broad divisions of language are oral language, reading, and writing. Verbal Comprehension (*Gc*), consisting of four subtests (Picture Vocabulary, Synonyms, Antonyms, and Analogies), is the primary measure of oral language (LD, language development; VL, lexical knowledge) in the WJ III COG. Several other tests of oral language, as well as the tests of reading and writing, are included in the WJ III ACH (see Table 14.5). Reading and writing are generally considered skills that are learned primarily through formal schooling. Therefore, those tests will be mentioned again in a later section on academic achievement. Additionally, as reflected in the test descriptions in Table 14.5, the four WJ III ACH oral language tests can provide language-related information regarding receptive language (Understanding Directions, LS, listening ability; Oral Comprehension, LS, listening ability) and expressive language (Picture Vocabulary, VL, lexical knowledge; Story Recall, LD, language development).

Assessment of Reasoning and Problem Solving (*Gf, Gv*)

Problem solving, or the ability to arrive at solutions in novel and unpracticed situations, involves a complex set of cognitive processes. Abstract thinking and adequate concept formation are required to formulate flexible ideas and strategies and to apply them across a variety of situations. Neuropsychological test batteries have employed a wide variety of tasks (e.g., Halstead Categories; Tower of London; Wisconsin Card Sorting Test) to tap different aspects of reasoning and problem solving.

Two WJ III tests are strong measures of abstract reasoning (*Gf*-Concept Formation), primarily a measure of inductive reasoning (I), and Analysis–Synthesis, primarily a measure of sequential or deductive reasoning (RG). Planning (*Gv*) is also a measure of sequential reasoning (RQ), although it also taps an aspect (SS, spatial scanning) of visual processing (*Gv*). The Quantitative Concepts test on the WJ ACH battery, which consists of 50% number series items, can provide additional insights into quantitative reasoning (RQ).

Assessment of Academic Achievement (*Grw*, *Gq*, *Gc*)

An important advantage of the WJ III, when the goal is to provide for psychometrically sound neuropsychological assessment, is the inclusion of a number of co-normed tests that measure learned skills associated with formal schooling. These tests all appear in the WJ III ACH and are only mentioned briefly (see Table 14.5 for additional information). The five tests of reading measure a spectrum of reading abilities from identifying letters and words in isolation to the comprehension of written text. The five tests of writing ability measure several writing abilities, ranging from spelling to the writing of sentences, that must meet certain requirements. The four tests of mathematics measure skills from basic calculation to mathematics reasoning. The WJ III ACH also contains a test of Academic Knowledge (Science, Social Studies, and Humanities). All of these tests were described earlier as possible measures of remote memory.

Assessment of Handwriting

The WJ III includes a normed scale of handwriting legibility. An individual's quality of handwriting may provide useful information about fine-motor hand coordination. This may be particularly useful if premorbid samples of the patient's handwriting are also available for evaluation. Although this procedure is usually applied to the written output from the WJ III Writing Samples test, the scale can be applied to any handwritten product.

Interpretation

Focused norms are currently being prepared for use in neuropsychological applications of the WJ III (Dean & Woodcock, 1999). Focused norms allow an individual's performance to be compared to others of the same age, education, and gender in the norming sample. This scoring system adjusts the WJ III age-based standard scores into standard scores based jointly on age, education, and gender. Thus, these demographics for individual patients can be taken into account when classifying patients as "normal" or "impaired." Conceptually, the *D-WNAS* focused norms are similar to recent efforts to incorporate both development (age) and the effects of schooling into norms for a Hebrew-language version of the WISC-R (Cahan, 2000).

Examples of score adjustment via the focused norms procedure follow. In one instance, a 40-year-old woman with 6 years of college obtained a standard score of 85 on the *Gs* (processing speed) cluster. A score of 85 is one standard deviation below the mean for all persons in the norming sample of the same age. When the score is adjusted by also taking into account gender and education, the focused norm standard score is 78. The adjusted score indicates that her performance is 1.6 standard deviations below the mean when compared to others in the norming sample most like herself. Thus, the deficit in performance is now seen as more significant than if age alone had been the basis for the standard score. In another instance, an 80-year-old man with only one year of schooling had a standard score of 76 on the *Gs* cluster. After adjustment via the focused norms procedure, his standard score was 103. These two examples are somewhat extreme; however, such cases may be encountered in neuropsychological practice.

Table 14.16 presents a suggested modification for the verbal labels used in Table 4-2 of the Examiner's Manuals for the WJ III. These labels are more appropriate for reporting levels of deficit or preserved function in neuropsychological reports. The functional level indices, derived from Rasch scaling (McGrew & Woodcock, 2001), are particularly useful in neuropsychological settings for describing the degree of deficit or preservation of functions demonstrated by the patient.

WJ-R Neuropsychological Research Data

Dean and Woodcock (1999) have presented a preliminary report on the validity of the *D-WNAS*.[7] This report includes information on: (1) descriptive statistics for a wide variety of clinical groups organized as per *DSM-IV* or *ICD-9* (total $N = 1,315$ subjects from 5 to 81 years of age), (2) factor analysis of the WJ-R cognitive tests in these samples, (3) factor analysis of the *D-WNAS* sensory and motor batteries, (4) predictive validity studies focused on predicting the presence and location of brain damage, and (5) four clinical case studies. Space limits the presentation of these extant data in detail in this chapter. Instead, a sample of the type of data and samples on which these data are being gathered is presented in Table 14.17. Inspection of the relative ordering of WJ-R *Gf–Gc* cluster scores within each clinical sample in Table 14.17 suggest is a number of interesting hypotheses. The reader is referred to the original report (Dean & Woodcock, 1999) for more detailed analysis and interpretation of these data.

The available *D-WNAS* validity evidence suggests possibilities for improving the "state-of-the art" of neuropsychological assessment through the combination of the CHC model of cognitive abilities, a co-normed battery of cognitive and achievement tests designed as per the CHC model (i.e., the WJ III), and supplementary and traditional neuropsychological (sensory, motor, interview, and mental status) assessments. Additional research data are currently being gathered and analyzed, and will be forthcoming.

CHC Abilities across the Life Span

A basic premise in science is that meaningful comparison of any two instances of a phenomenon requires use of the

[7]A copy of the report (*The WJ-R and Bateria-R in Neuropsychological Assessment: Research Report Number 3*) can be downloaded free at *www.IAPsych.com*.

TABLE 14.16 Functional level/deficit descriptions for WJ III scores used in neuropsychological assessments

RMI	W Difference Score	Functional Level	Patient Will Find the Demands of Related Age Level Tasks:
97/90 to 100/90	+11 and above	Advanced	Very Easy
75/90 to 96/90	−10 to +10	Adequate	Manageable
25/90 to 74/90	−30 to −11	Mildly Impaired	Very Difficult
4/90 to 24/90	−50 to −31	Moderately Impaired	Extremely Difficult
0/90 to 3/90	−51 and below	Severely Impaired	Impossible

TABLE 14.17 WJ-R *Gf–Gc* cluster score pattern by type of sample as reported for the *D-WNAS:* Ages 5 to 81

Sample	n			***Gf–Gc*** **Cluster by Standard Score Order**						
			BCA	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Reference Samples:										
WJ-R Norming Sample	5470	Cluster:	BCA	*Gv*	*Gc*	*Gf*	*Ga*	*Gs*	*Glr*	*Gsm*
		Mdn:	100	100	100	100	100	100	100	100
		SD:	16	16	16	15	15	16	16	16
Total Clinical Sample	1315	Cluster:	BCA	*Gs*	*Glr*	*Gc*	*Ga*	*Gf*	*Gsm*	*Gv*
		Mdn:	90	87	91	92	93	93	94	98
		SD:	18	18	15	18	15	17	18	17
Gifted	84	Cluster:	BCA	*Gv*	*Gsm*	*Ga*	*Glr*	*Gf*	*Gs*	*Gc*
		Mdn:	120	105	110	111	112	116	118	120
		SD:	11	13	15	13	16	11	14	13
Clinical Samples:										
Deficits in Acquired Knowledge										
Knowledge <70	56	Cluster:	BCA	*Gc*	*Gf*	*Gs*	*Gsm*	*Glr*	*Ga*	*Gv*
		Mdn:	56	58	65	68	70	72	73	76
		SD:	11	10	12	15	12	16	11	16
Math <70	122	Cluster:	BCA	*Gs*	*Gc*	*Gf*	*Gsm*	*Glr*	*Ga*	*Gv*
		Mdn:	64	68	68	72	77	78	80	82
		SD:	14	15	16	12	14	15	13	16
Oral Language <70	63	Cluster:	BCA	*Gc*	*Gsm*	*Gf*	*Gs*	*Glr*	*Ga*	*Gv*
		Mdn:	59	60	70	70	71	73	74	77
		SD:	10	10	11	11	12	12	11	15
Reading <70	133	Cluster:	BCA	*Gc*	*Gs*	*Gsm*	*Gf*	*Glr*	*Ga*	*Gv*
		Mdn:	66	69	72	75	76	77	82	89
		SD:	15	16	13	15	14	13	13	16
Written language <70	164	Cluster:	BCA	*Gs*	*Gc*	*Glr*	*Gsm*	*Gf*	*Ga*	*Gv*
		Mdn:	70	75	76	78	78	80	83	89
		SD:	15	14	16	12	15	14	13	16
Anxiety Spectrum Disorders	100	Cluster:	BCA	*Gs*	*Glr*	*Ga*	*Gc*	*Gf*	*Gsm*	*Gv*
		Mdn:	95	91	94	94	96	97	97	100
		SD:	16	17	15	15	17	16	16	15
Attention Deficit/Hyperactivity Disorders, Mixed	494	Cluster:	BCA	*Gs*	*Glr*	*Ga*	*Gc*	*Gf*	*Gsm*	*Gv*
		Mdn:	95	90	93	94	96	96	97	100
		SD:	16	17	14	14	16	15	17	15
Brain Tumors, Mixed	32	Cluster:	BCA	*Gs*	*Gc*	*Glr*	*Gsm*	*Ga*	*Gf*	*Gv*
		Mdn:	90	90	92	93	94	94	96	97
		SD:	15	20	17	12	14	11	14	16

TABLE 14.17 (Continued)

				Gf–Gc Cluster by Standard Score Order						
Sample	***n***		**BCA**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Clinical Samples:										
Depressive Spectrum Disorder	150	Cluster:	BCA	*Gs*	*Ga*	*Gf*	*Gsm*	*Glr*	*Gc*	*Gv*
		Mdn:	95	92	94	96	96	97	98	100
		SD:	16	18	13	14	17	14	17	15
Hydrocephalus	18	Cluster:	BCA	*Gs*	*Gc*	*Gf*	*Glr*	*Ga*	*Gsm*	*Gv*
		Mdn:	62	66	69	76	78	81	82	89
		SD:	19	18	20	14	22	14	16	21
Impulsive/Disruptive	73	Cluster:	BCA	*Gs*	*Gc*	*Gf*	*Ga*	*Gsm*	*Glr*	*Gv*
Spectrum Disorders		Mdn:	87	86	87	90	91	92	94	98
		SD:	16	19	14	16	14	17	14	17
Infectious Processes	23	Cluster:	BCA	*Gs*	*Gc*	*Ga*	*Gsm*	*Glr*	*Gf*	*Gv*
		Mdn:	79	68	82	82	85	87	89	93
		SD:	20	20	20	12	16	17	21	22
Language Disorders	48	Cluster:	BCA	*Gsm*	*Gc*	*Ga*	*Gs*	*Gf*	*Glr*	*Gv*
		Mdn:	78	81	82	82	84	86	88	100
		SD:	15	14	16	11	15	17	15	14
Learning Disorders, Mixed	584	Cluster:	BCA	*Gs*	*Glr*	*Gc*	*Ga*	*Gf*	*Gsm*	*Gv*
		Mdn:	88	86	89	91	92	93	93	98
		SD:	15	17	14	16	14	15	17	16
Mental Retardation, Mild to Profound	81	Cluster:	BCA	*Gc*	*Gf*	*Gs*	*Gsm*	*Ga*	*Glr*	*Gv*
		Mdn:	56	62	66	68	71	74	75	80
		SD:	13	12	13	16	13	13	15	17
Motor Impairment	52	Cluster:	BCA	*Gs*	*Glr*	*Ga*	*Gf*	*Gv*	*Gsm*	*Gc*
		Mdn:	93	90	90	95	96	96	101	102
		SD:	17	16	18	13	16	18	20	18
Neurofibromatosis	11	Cluster:	BCA	*Gsm*	*Ga*	*Gc*	*Glr*	*Gs*	*Gf*	*Gv*
		Mdn:	84	85	87	87	88	89	89	97
		SD:	14	10	14	11	9	19	13	19
Pervasive Developmental Disorders	13	Cluster:	BCA	*Gs*	*Gf*	*Ga*	*Gsm*	*Gc*	*Glr*	*Gv*
		Mdn:	75	72	80	80	81	87	88	93
		SD:	20	29	16	11	19	18	16	20
Seizure Disorders/Epilepsy	57	Cluster:	BCA	*Gc*	*Gs*	*Glr*	*Gf*	*Gsm*	*Ga*	*Gv*
		Mdn:	83	84	85	89	89	91	92	93
		SD:	17	18	17	15	14	16	16	15
Traumatic/Closed Head	170	Cluster:	BCA	*Gs*	*Gc*	*Glr*	*Gsm*	*Ga*	*Gf*	*Gv*
Injury		Mdn:	92	89	94	95	95	96	96	96
		SD:	19	21	17	18	16	15	16	18

(Continues)

TABLE 14.17 (Continued)

Sample	*n*			*Gf–Gc* Cluster by Standard Score Order						
			BCA	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Samples with Known Lesion Localization:										
Left Hemisphere Only	56	Cluster:	BCA	*Gs*	*Glr*	*Gc*	*Ga*	*Gsm*	*Gf*	*Gv*
		Mdn:	85	84	86	87	92	92	95	97
		SD:	18	19	15	15	17	14	18	18
Right Hemisphere Only	64	Cluster:	BCA	*Gs*	*Glr*	*Gc*	*Gsm*	*Gv*	*Gf*	*Ga*
		Mdn:	88	83	90	92	92	93	93	93
		SD:	18	23	19	17	18	18	14	14
Bilateral Diffuse Brain Damage	36	Cluster:	BCA	*Gs*	*Glr*	*Gc*	*Gf*	*Ga*	*Gv*	*Gsm*
		Mdn:	89	88	89	90	94	95	95	96
		SD:	20	21	15	20	19	16	14	15
Anterior Cortical Lesions, Mixed	68	Cluster:	BCA	*Gs*	*Gc*	*Glr*	*Ga*	*Gsm*	*Gf*	*Gv*
		Mdn:	90	86	91	92	95	96	96	98
		SD:	20	24	18	20	16	17	17	21
Posterior Cortical Lesions, Mixed	78	Cluster:	BCA	*Gs*	*Glr*	*Gc*	*Gsm*	*Gf*	*Ga*	*Gv*
		Mdn:	88	82	87	89	91	92	93	94
		SD:	18	20	15	17	17	16	17	17
Frontal Lobe Lesions, Mixed	22	Cluster:	BCA	*Gs*	*Gc*	*Gv*	*Glr*	*Gf*	*Ga*	*Gsm*
		Mdn:	85	76	85	85	88	88	89	90
		SD:	24	34	18	29	24	17	17	22
Temporal Lobe Lesions, Mixed	52	Cluster:	BCA	*Gs*	*Glr*	*Gc*	*Gf*	*Gsm*	*Ga*	*Gv*
		Mdn:	88	85	87	87	93	93	93	97
		SD:	15	17	14	16	13	15	16	15
Parietal Lobe Lesions, Mixed	20	Cluster:	BCA	*Gs*	*Glr*	*Gc*	*Gsm*	*Ga*	*Gf*	*Gv*
		Mdn:	78	72	78	78	83	84	86	92
		SD:	16	22	14	12	16	13	15	19
Subcortical/Brain Stem Lesions	17	Cluster:	BCA	*Gs*	*Glr*	*Gv*	*Gf*	*Ga*	*Gc*	*Gsm*
		Mdn:	83	73	86	86	86	88	89	95
		SD:	16	18	14	15	9	11	17	17

Key to Cluster Abbreviations: BCA = Broad Cognitive Ability; *Gsm* = Short-Term Memory; *Gs* = Processing Speed; *Glr* = Long-Term Retrieval; *Gv* = Visual Processing; *Gc* = Comprehension–Knowledge; *Ga* = Auditory Processing; *Gf* = Fluid Reasoning.

Copyright © 2001 by The Riverside Publishing Company. Adapted from the Woodcock-Johnson® III (WJ III™) by Richard W. Woodcock, Kevin S. McGrew and Nancy Mather, with permission of the publisher. All rights reserved.

same measurement scale. (Strauss, Spreen & Hunter, 2000, p. 242)

The study of developmental patterns of cognitive growth and change has fascinated cognitive and developmental psychologists for decades (see Chapter 5). Although longitudinal data are typically viewed as the best source for investigating developmental changes in intelligence, large-scale longitudinal data sets are expensive and are few and far between. Not unexpectedly, cross-sectional normative data from nationally standardized measures of cognitive abilities have played a prominent role in this research. As noted by Gustafsson and Undheim (1996), most of the research on changes in cognitive abilities, particularly during adulthood, "has been tied largely to *Gf–Gc* theory" (p. 221). This section highlights some of the limitations of the extant cross-sectional intelligence research and then describes how the WJ III can help overcome these limitations. Finally, potential new insights into the growth and decline of human intelligence are presented via select cross-sectional WJ III CHC growth curves.

Some Limitations of the Available Intelligence Growth Curve Research

Various iterations of the Wechsler series' norm data have played a central role in the analysis of intellectual development, primarily because of the relatively similar format of most of the individual tests in the three separate Wechsler batteries. Notwithstanding the important contribution of the analyses of the various Wechsler standardization samples, these data suffer from a number of significant limitations that raise questions about the accuracy of some of the conclusions derived from this research.

First, and foremost, despite tests with common names and test formats (e.g., WPPSI-R, WISC-III, and WAIS-III Similarities test), "substantial differences in content are present, and the pattern of performance for a given participant across subtests may not be particularly consistent between tests" (Strauss et al., 2000, p. 238). Attempts to apply decision rules derived from the analyses of test scores and subtest patterns on one version of the Wechsler may not be applicable when using a different version (Bornstein, 1987; Chelune, Eversole, Kane, & Talbott, 1987).

Second, as outlined in this chapter and by others (Flanagan et al., 2001; McGrew & Flanagan, 1998; Woodcock, 1990), most of the Wechsler-based developmental research has used tests that are "impure from the perspective of Horn's theory" (Kaufman, Kaufman, Chen, & Kaufman, 1996, p. 161). The questioning of the validity of some of the Wechsler tests as indicators of CHC abilities has significant implications. For example, the confidence that has been placed in the interpretation of the Wechsler Performance Scale as an indicator of *Gf* is now being questioned. Wang and Kaufman (1993) pointed out the significant implications of this now-recognized Performance Scale misinterpretation when they observed:

> *analyses conducted by Woodcock (1990) and Stone (1992) that offer empirical evidence that Wechsler's Performance Scale may, indeed, be primarily a measure of* Gv*.... The possibility remains that the numerous research investigations of aging and intelligence that have involved the WAIS and WAIS-R may have attested to the rapid and early decline of* Gv*—or, more likely, an amalgam of* Gv *and* Gf*—instead of just* Gf*. (p. 30)*

The aforementioned confounded interpretation of the Wechsler Performance Scale is problematic as it is axiomatic in science that the measurement of a phenomenon requires a unidimensional measurement scale in which the resultant scores reflect individual differences on a single common dimension (Hattie, 1985; Lumsden, 1961; Reise, Waller, & Comrey, 2000). "If a scale is multidimensional (i.e., has multiple correlated dimensions), then not only is the total score more challenging to interpret but different aspects of the scale (e.g., its content facets) may have different correlations with external variables" (Reise et al., 2000, p. 293).

Third, as described in the cross-battery section of this chapter, the Wechsler batteries, as well as most other major intelligence batteries, have been unable to shed light on a number of important CHC abilities because of inadequate construct representation. With the exception of the WAIS-III Matrix Reasoning tests, none of the Wechsler batteries have been able to shed valid light on the developmental patterns for *Gf*. Other constructs that have not been reflected in the extant Wechsler cognitive developmental literature are *Glr* and *Ga*. Fortunately, recent analyses of the developmental change in valid *Gf* and *Glr* test scores from other nationally normed instruments (viz., K-BIT, K-FAST, K-SNAP, and KAIT) have been reported (Kaufman et al., 1996; Wang & Kaufman, 1993). Fourth, even for the Wechsler tests that are valid indicators of a CHC ability (e.g., Arithmetic as an indicator of *Gq*), the availability of only one test or indicator for a CHC construct limits the generalizability of some of the research findings (Kaufman et al., 1996).

Finally, the lack of an equal-interval measurement scale *across* the three Wechsler batteries has necessitated some creative methodological "trickery" to analyze scores across ages. For example, Kaufman (1990) employed a procedure with the WAIS-R data in which the individual test scores for all seven adult age groups in the standardization sample were equated to a reference norm group. All WAIS-R norm subject individual test raw scores were converted to subtest scaled scores (M = 10; SD = 3) using the "target" norm age group of 25–34. This provided for the ability to analyze the change in standard scores in reference to a common standard.

Another creative approach reported by Kaufman et al. (1996) was to convert the raw scores for all norm subjects between ages 15 and 94 on seven tests from the Kaufman family of instruments to z-scores (M = 100; SD = 15) calculated on the entire sample (N = 1,193). The resultant standard scores, which are referenced to the mean and standard deviation of the entire sample, were then analyzed. At the level of individual tests, the latter approach provides for more precision in the measurement of change than the WAIS-R approach. The WAIS-R subtest scaled scores only provide for three points of measurement for every standard deviation on the scale, regardless of which normative reference group is used. In contrast, the Kaufman et al. (1996) approach placed the individual tests on a scale that allowed for five times the degree of ability differentiation (15 points are covered for each standard deviation on the scale). All subjects with a scaled score of 12 on a WAIS-R subtest are not all likely to be at the same ability level and would cover a range of scores on a scale with a standard deviation of 15.

Regardless of the creative methods used to obtain a score suitable for analyses across age groups, this has only provided a partial *within*-battery metric solution. Without a common equal-interval growth scale, the analyses of cognitive change *across* similar batteries from the same family of tests (e.g., Wechsler or Kaufman family of related instruments) is extremely difficult and fraught with potential error. These measurement limitations result in lost opportunities for more comprehensive and informative analyses of cross-sectional CHC-based data.

Advantages of the WJ III in Measuring Growth and Change

The WJ III is particularly well suited for the measurement of growth and change both in clinical practice and for developmental research. A number of characteristics of the WJ III address the previously described limitations of measures.

First, the WJ III includes the same tests across all developmental age groups. Although only certain tests provide norms below age five, almost all of the 20 WJ III COG and 24 ACH tests provide measurement starting at age 5 or 6 and extending up through 95+ years of age. The use of the same tests across most of the life span removes the potential of "method" effects (i.e., different test content across tests in different batteries) confounding the interpretation of the resultant change scores. Second, when the focus is on changes in CHC abilities across the life span, the WJ III provides two-test clusters that maximize construct relevant variance. As described previously, each WJ III CHC COG cluster is comprised of two tests of qualitatively different narrow abilities (therefore insuring adequate construction representation) within each respective *Gf–Gc* domain. No other individually administered battery provides empirically validated cluster scores for the major cognitive constructs included in contemporary CHC theory.

Finally, and probably more important in the context of the current discussion, is the fact that all WJ III tests are grounded in unidimensional and equal-interval *growth* scales. All WJ III tests incorporate the *W*-scale, a transformation and application of the Rasch measurement model (Woodcock, 1978; Woodcock & Dahl, 1971). Each test's *W*-score is centered on a value of 500, which is the approximate average performance of 10-year-olds. Cluster scores represent the arithmetic mean (average) of the tests comprising the cluster. Although the *W*-scores are test- or cluster-specific (*W*-scores cannot be compared *across* measures), changes in scores can be compared. That is, a change of 1 *W* point represents the same amount of unit change within any of the WJ III tests or clusters. More importantly, within a test or cluster, growth can be measured from the preschool years through late adulthood on a single common scale.[8]

For the above reasons, the WJ III battery is particularly well suited to the measurement and evaluation of cognitive growth and change. Examples of the potential research benefits accrued from using a battery designed like the WJ III have been demonstrated in research with the WJ-R. For example, McArdle and Woodcock (1997; also see test-retest study by McArdle & Woodcock reported in McGrew et al., 1991) presented a series of longitudinal test-retest designs with developmental time-lag components that focused on decomposing the sources of change in test scores over time (e.g., test score variance due to practice and retention, growth or maturation, trait stability, and test unreliability). Using the WJ-R standardization data, Salthouse (1998a) investigated the extent to which age-related differences in cognitive abilities should be interpreted as reflecting either a general developmental mechanism or an ability-specific mechanism.

WJ III CHC Growth Curves

The norm-based growth curves for 11 WJ III clusters are presented in Figures 14.5a–k.[9] Included are the curves for the GIA-Ext (Figure 14.5a), seven CHC cognitive clusters (*Gc, Glr, Gv, Ga, Gf, Gs, Gsm*, Figure 14.5b–h), and three broad achievement clusters (reading, math, and written language, Figure 14.5i–k). Each figure includes three smoothed curves (average score and standard deviations) based on the WJ III norms. We believe these figures represent the first time a complete set of *Gf–Gc* growth curves based on measures with strong construct validity (adequate construct representation) have been presented across most of the life span. The following discussion of the curves will be descriptive. Appropriate data-analytic methods need to be applied to these data to empirically evaluate the trends and to compare the results with the extant literature on the development of CHC abilities. These growth curves should be systematically compared to the analyses of the WAIS-III and other measures (WAIS, WAIS-R, Kaufman tests) presented in detail in Chapter 5, especially regarding the different aging patterns for different abilities within Horn's expanded *Gf–Gc* framework. However, note that the data are not directly comparable because (1) the WJ III analyses use *W*-scores and the Chapter 5 analyses use

[8]See Woodcock (1978) for a thorough treatment of the development and application of the Rasch-based *W*-score metric.

[9]See McGrew and Woodcock (2001) for a description of how the growth curves were centered at the same starting point to allow for a comparison of relative changes across the curves. The *W*-scores on the x-axis do *not* represent the normative values as a different constant has been subtracted from all values for each curve. Furthermore, ideally it would be optimal to present an additional set of curves of the same data using a logarithmic transformation of the age scale. This would allow for a closer examination of the changes occurring during the early years. Space limitations preclude the presentation of both sets of figures.

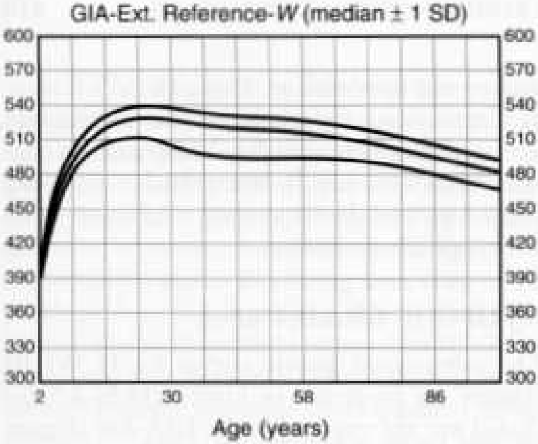

FIGURE 14.5a

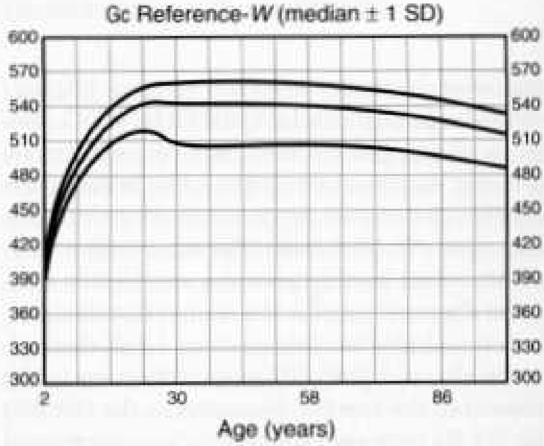

FIGURE 14.5b

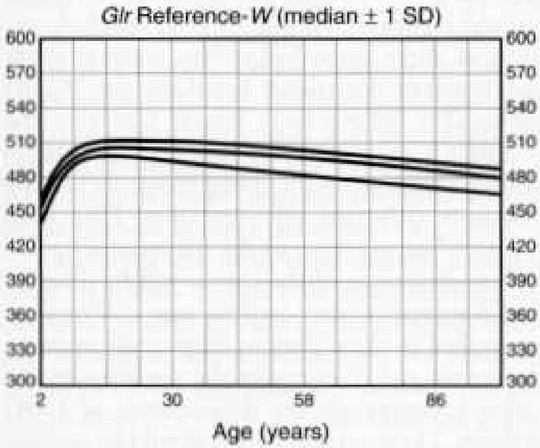

FIGURE 14.5c

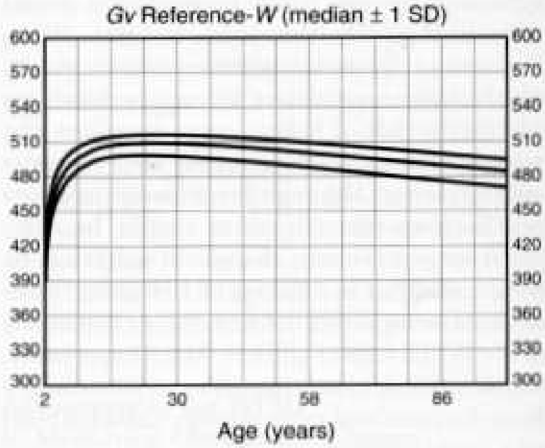

FIGURE 14.5d

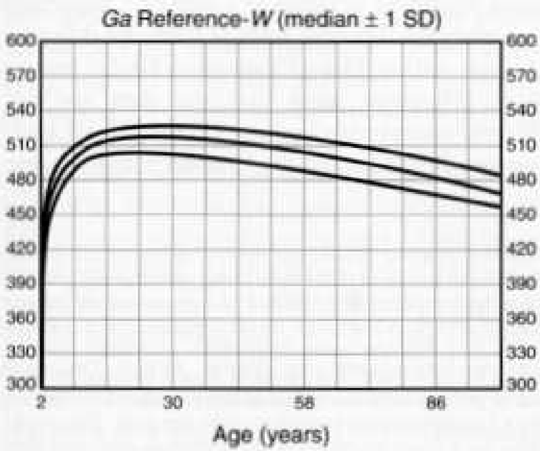

FIGURE 14.5e

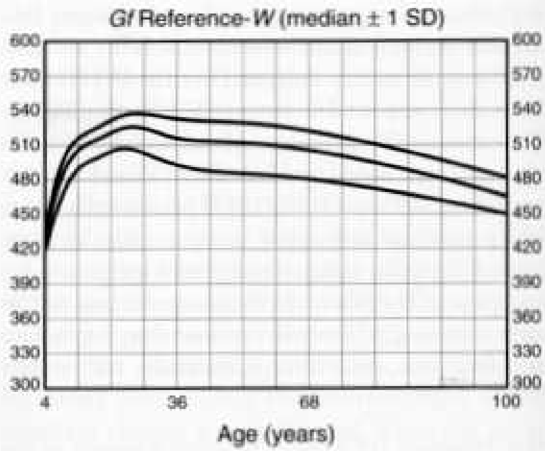

FIGURE 14.6f

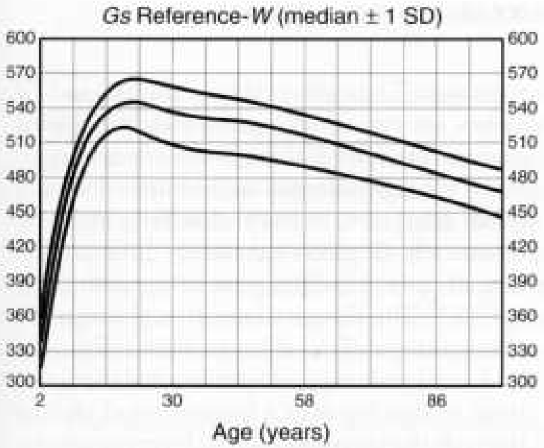

FIGURE 14.5g

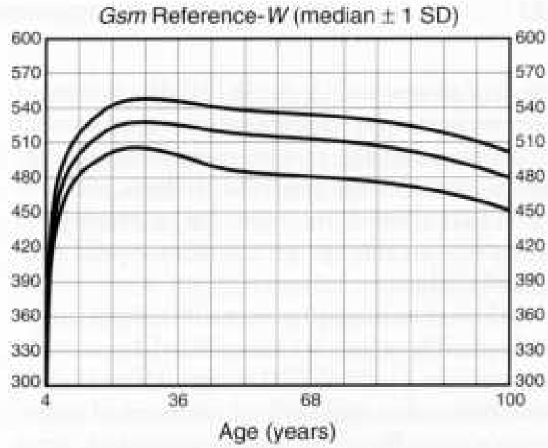

FIGURE 14.5h

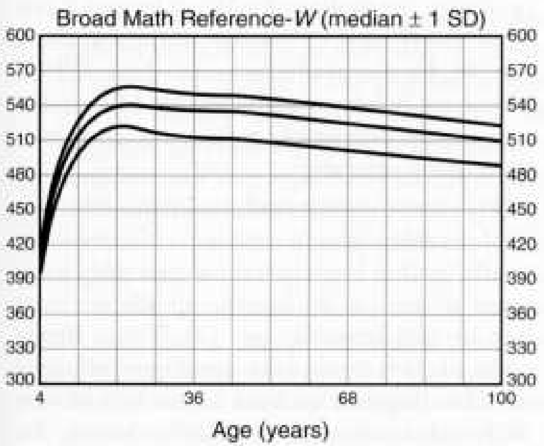

FIGURE 14.5i

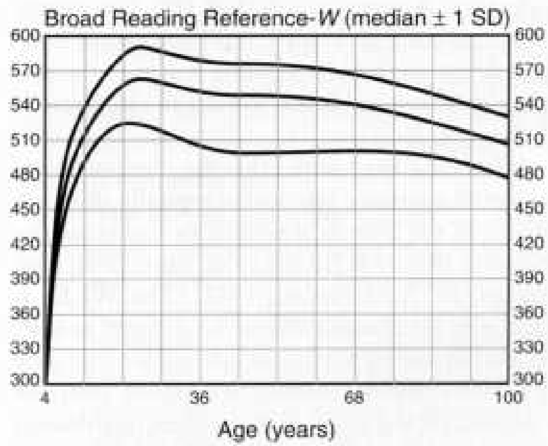

FIGURE 14.5j

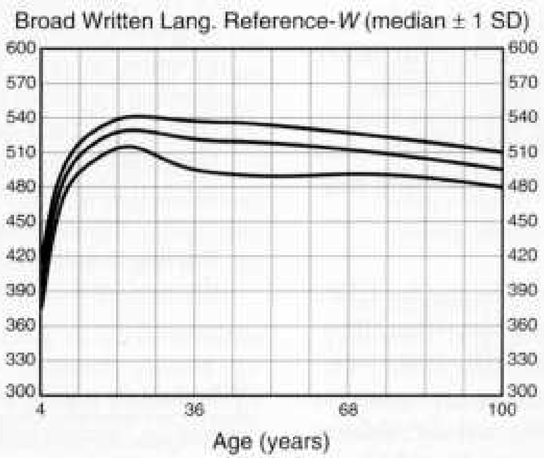

FIGURE 14.5k

standard scores; and (2) the WJ III data presented here are not corrected for educational attainment across the adult age groups, raising the possibility that some of the observed declines with age for WJ III abilities are more due to educational differences in the age groups than in real age-related decline.

Before examining the curves, it is important to understand how they are constructed, because the construction of the WJ III norms differs slightly from most other standardized measures of cognitive abilities. First, the middle smoothed curve represents the average (median) *W*-score for an age group. These are referred to as the WJ III Reference *W*-score values (Ref *W*). The curves above and the Ref *W* curve represent *unique* standard deviations for the separate halves of the distribution above and below the Ref *W* (McGrew & Woodcock, 2001).

It has long been noted that the *observed* distribution of cognitive test scores typically does not adhere to the *theoretical* normal or Gaussian distribution (Brody & Brody, 1976; Burt, 1963; Matarazzo, 1972; McNemar, 1942). The WJ III is unique in psychoeducational assessment in that, via a set of special norm procedures (see McGrew et al., 1991, and McGrew & Woodcock, 2001, for explanation), the shape of the original distribution of traits is *not* transformed to conform to the normal curve. Instead, the use of separate standard deviations above and below the Ref *W* retains the "real-world" distribution of ability traits. We believe this procedure provides scores that better mirror the reality of human cognitive abilities.

Inspection of the curves in Figure 14.5a–k, as well as review of the actual Ref *W* and SD values (not reported here), suggest the following general conclusions:

- The pattern of growth and decline for CHC abilities differs markedly, an observation attesting to the uniqueness of the CHC constructs.
- Most CHC abilities reach an asymptote at approximately age 25. *Glr* and *Gf* abilities reached an earlier apex at ages 20 and 22, respectively. The early peaks for the *Glr* and *Gf* curves are typical of cognitive abilities based more on processing abilities that are developed more through informal and indirect learning (often referred to as *process*-dominant abilities). Conversely, *Gc* shows a markedly different pattern of growth, with a peak at approximately age 45.[10] The *Gc* curve is more representative of cognitive abilities that are more influenced by formal training and learning and that continue to develop over a longer period of time through the crystallization of learning experiences (*product*-dominant abilities). The peak ages for abilities such as *Glr* and *Gf* accord well with the findings for Wechsler and Kaufman tests (Chapter 5), as does the peak at about age 45 for *Gc*; again, however, these WJ III data have not been corrected for educational attainment differences across the adult portion of the age range.
- All cognitive abilities influenced more by formal learning and instruction (*Gc*, reading, math, written language) show extremely steep rates of growth during the childhood years (up to approximately age 12). These trends most likely represent the significant influence of schooling on the rapid acquisition of new acquired knowledge and skills during the early school years. Conversely, a number of more process-dominant abilities (*Glr*, *Gv*, *Ga*, *Gf*) display a much briefer and less rapid rate of growth during the formative years. It is interesting to note that two largely process-dominant abilities classified as cognitive efficiency constructs in the CPM model (*Gs* and *Gsm*) also demonstrate steep rates of development during the childhood and adolescent years (especially *Gs*). This observation is intriguing in light of the information-processing-theory-based *developmental cascade* model,

[10]The slight dip at age 30 in Figure 14.5b is a result of a less than perfect curve smoothing process used in the development of the figures for this chapter and does not represent a real change in abilities or the actual WJ III *Gc* norms.

which posits that, with increasing age and maturation, processing speed (*Gs*) and working memory (*Gsm*, MW) may be responsible for most increases in general intellectual functioning (Fry & Hale, 1996, 2000; Miller & Vernon, 1996). According to Fry and Hale (1996, p. 30), "virtually all of the effect of the age-related increase in speed on intelligence appears to be mediated through the effect of speed on working memory."

- The variability (as reflected in the two standard deviation curves) for CHC cognitive abilities is markedly different. The process-dominant abilities of *Glr*, *Gv*, and *Ga* show a much narrower range of variability. Conversely, reading abilities demonstrate an extremely wide range of normal variability in the population, particularly starting and continuing after adolescence. In contrast, math achievement, an acquired trait that is probably the most linked to formal instruction, is much less variable in comparison to *Grw*. One possible explanation for the significant difference in *Grw*/Math (*Gq*) variability may be that many people continue to read during their everyday life experiences once they leave school and, thus, continue to develop better reading abilities, while fewer people continue to learn new math skills via everyday learning experiences.
- One of the most significant WJ III contributions to the developmental literature is the presentation of the asymmetrical SD values in Figures 14.5a–k. These unique SD values provide potential insights into the distribution of human cognitive abilities not previously recognized. Inspection of the exact ratio of the two SDs at each age (not reported here) reveals a number of interesting findings that will benefit from additional research and study. Of interest are the observations that:
  - With a few exceptions, all CHC abilities are characterized by positively skewed distributions. There is a greater portion of the population below the average or median value than above.
  - *Glr* shows significant swings in distributional characteristics. The distribution is positively skewed until approximately age 6, after which the shape shifts to a more normal distribution. However, starting at approximately the beginning of adolescence, the distribution systematically shifts in the direction of positive skew and becomes extremely skewed after approximately age 30. Given the prominence of memory decline in descriptive and theoretical studies of aging, this finding warrants further exploration.
  - Acquired knowledge abilities (*Gc*, *Grw*, *Gq*) also display a significant shift toward more individuals being below than above average after approximately age 25. There is an observed swing back toward normality for *Grw* (reading and writing) starting at approximately age 75.
  - The process-dominant *Gf* and *Ga* abilities show interesting patterns that vary from other CHC abilities. Although approximating normality beginning at age 8, the *Ga* distribution begins a systematic and monotonically increasing trend toward a negatively skewed distribution starting at approximately age 30. Around age 65 to 70, the *Ga* distribution becomes increasingly negatively skewed in shape. The relation between this observation and the age-related changes in hearing acuity warrants further exploration. The trend for *Gf* abilities roughly mirrors that for *Ga*, although the observed SD ratio suggests a positively skewed distribution of abilities from approximately age 13 through age 75, after which it systematically shifts into a more normal distribution.

In summary, the WJ III CHC growth curves presented here provide intriguing insights into the growth and development of human cognitive abilities. Additional research with appropriate analytic methods is necessary to "tease out" the possible explanations of the observations noted above (for example, controlling for educational attainment within the adult age groups to rule out the possible role of educational differences as an explanation for a portion of the apparent age-related declines; see Chapter 5). Although there is a risk in simply presenting the WJ III CHC growth curves in the absence of structured data analyses, we believe that the presentation of the WJ III CHC ability curves may serve to stimulate new research, dialogue, and insights into understanding and explaining the growth and development of human intelligence. In particular, the presentation of data that maintains the "real-world" distributional characteristics of CHC constructs has the potential to impact thinking on how human cognitive abilities should be properly measured and described. Similar to recent research that has simultaneously examined the population distributions at the extremes (viz., individuals with mental retardation or who are gifted), the WJ III approach to maintaining the asymmetrical characteristics of human cognitive abilities is consistent with the suggestion that we should "reevaluate our concept of intelligence as necessarily conforming to the expected normal curve distribution" (Robinson, Zigler, & Gallagher, 2000, p. 1415). In other words, it may be time to "redraw the normal curve" (Robinson et al., 2000).

CHC Cross-Battery Applications

CHC Cross-Battery (CB) assessment is a "*time efficient method of intellectual assessment that allows practitioners to measure validly a wider range (or a more in-depth but selective range) of cognitive abilities than that represented by any one intelligence battery in a way consistent with contemporary psychometric theory and research on the structure of intelligence*" (McGrew & Flanagan, 1998, p. 357; italics original). The goal of CB assessment is to guide practitioners, via a set of systematic principles, steps, and procedures, to design assessments that are organized vis-à-vis the CHC theory of intelligence. For most test batteries, this serves as a form of *CHC post-hoc validity repair* (Flanagan, McGrew, & Ortiz, 2001). That is, with the exception of the WJ-R and WJ III, all available individually administered intelligence batteries have not used the CHC theory as their test-design blueprint. This results in the need to supplement the other intelligence batteries with additional measures to increase their coverage of CHC abilities. The primary objective of CB assessments is to combine two or more tests of different narrow CHC abilities to provide a composite score to represent a broad CHC ability. In this regard, tests that are relatively factorially pure measures of constructs are favored.

The birth of CHC CB assessments can be traced to Woodcock's (1990) CHC-organized joint confirmatory factor analysis of the then-available major intelligence batteries. Woodcock's analyses resulted in the classification of the individual tests from each intelligence battery into the broad *Gf–Gc* abilities of the CHC model. McGrew (1997) extended this work by classifying all tests at both the narrow *and* broad ability strata. This led to a detailed specification of the CB approach (Flanagan & McGrew, 1997; McGrew & Flanagan, 1998). A Wechsler-specific application of the CB approach followed (Flanagan, McGrew, & Ortiz, 2001). The most recent refinement of the CB approach is that articulated by Flanagan and Ortiz (2001).

The purpose of this section is threefold. First, the focus is on the application of the CB approach with adolescents and adults. The second focus is on how the WJ III battery can be used as the primary "tool chest" for supplementing other major intelligence batteries. Finally, approaches to supplementing the WJ III cognitive battery will be discussed.

The Big Three

Only three intelligence batteries provide coverage of the complete adolescent and adult age ranges. The Kaufman Adolescent and Adult Intelligence Test (KAIT; Kaufman & Kaufman, 1993) can be used from ages 11 through 85+. The WAIS-III/

WMS-III (Wechsler, 1997a, 1997b) span ages 16 to 89. Finally, as described previously, the WJ III spans ages 2 through 95+. Although other cognitive batteries (i.e., CAS, DAS, SB-IV) provide coverage of the adolescent age range, norms are not provided for assessing adults. The KAIT, WAIS-III/WMS-III, and WJ III have the most current set of test norms with publication dates of 1993, 1997, and 2001, respectively. For these reasons, plus the desire to adhere to the CB principle of minimizing the number of norm groups "crossed" when designing a CB assessment, the current discussion will be limited to these three batteries. Information on the selective and judicious use of individual tests from the remaining intelligence batteries for adolescents can be gleaned from other sources (Flanagan et al., 2000; Flanagan & Ortiz, 2001; McGrew & Flanagan, 1998).

Supplementing the WAIS-III/WMS-III with the WJ III[11]

Table 14.10 provides a summary of the broad and narrow CHC ability classifications of the individual tests in the WAIS-III, WMS-III, KAIT, and WJ III based on McGrew et al. (2001). A review of Table 14.10 indicates that the WAIS-III has adequate construct coverage (i.e., at least two tests of two qualitatively different narrow CHC abilities) of *Gc*, *Gv*, *Gs*, and *Gsm*. Supplementing the WAIS-III in these broad ability domains is not necessary, unless in-depth assessment of specific narrow abilities is suggested. For example, if a subject's fund of general information (K0) is a concern, the WJ III General Information test could be administered. The WJ III General Information and WAIS-III Information tests could be combined, as per CB procedures (see Flanagan & Ortiz, 2001), into a general information (K0) composite score. The WJ III should also be considered when additional information is required in the domain of listening ability (LS), particularly given that the WMS-III Logical Memory I test is a factorially complex measure (and, thus, is a diagnostically indeterminate measure of *Gc* and *Glr*). The WAIS-III Verbal Comprehension Index would appear to be a good score to use as an indicator of the broad ability of *Gc*.

The WAIS-III Processing Speed Index (PSI) can be interpreted as a valid measure of the broad *Gs* given its coverage of the narrow abilities of perceptual speed (P) and rate-of-test-taking (R9). The WJ III battery includes measures of the same narrow *Gs* abilities. In addition, the WJ III Decision Speed test is believed to be a measure of an aspect of semantic processing speed (R4).

The WAIS-III version of the Perceptual Organization Index (POI), when viewed from the perspective of CHC theory, is a step backwards. The WAIS-III POI consists of one good indicator each of *Gv* (Block Design, SR/Vz) and *Gf* (Matrix Reasoning, I), and a factorially complex indicator of *Gv* (CF) and *Gc* (Picture Completion).[12] The interpretation of the POI index is, therefore, diagnostically complex and indeterminate. If valid coverage of the broad *Gv* ability is required, it is suggested that the WAIS-III Block Design test be supplemented with the WAIS-III Object Assembly test (CS) and the WJ III Picture Recognition test (MV). This would provide for coverage of three different narrow *Gv* abilities (viz., Vz, CS, and MV). The WJ III Spatial Relations test would be useful if in-depth measurement of the highly related SR/Vz abilities is necessary.

The WAIS-III provides for adequate coverage of *Gsm* via measures of memory span (MS) and working memory (MW). However, the CHC-organized CB research consistently suggests that the WAIS-III Working Memory Index (WMI) contains a significant proportion of construct-irrelevant *Gq* variance. The WAIS-III

[11]The interpretations provided in this section also draw on the extant CB factor-analysis research for Wechsler as summarized by Flanagan et al. (2001), particularly for the WAIS-III tests not included in the analysis reported in Table 14.10.

[12]Although not reported in Table 14.10, the extant Wechsler factor-analysis research has consistently found the Picture Completion test to load on both a visual (*Gv*) and verbal (*Gc*) factor.

Arithmetic test is considered a good indicator of *Gq*, not *Gsm*. In addition to providing additional measures of memory span and working memory from the WMS-III and WJ III, the WJ III has the advantage of providing a relatively factorially pure norm-based composite score for working memory (Working Memory cluster).

The greatest benefit offered by the WJ III in WAIS-III CB assessments is the number of factorially pure indicators of the broad CHC abilities in the WJ III. The WJ III is also the only significant source for the assessment of *Ga* in adolescents and adults. *Glr* abilities are also completely underrepresented on the WAIS-III. As presented in Table 14.10, clinicians can turn to either the WMS-III or the WJ III to measure *Glr* abilities. The unique contribution of the WJ III comes from the Retrieval Fluency and Rapid Picture Naming tests, tests that provide for the measurement of ideational fluency (FI) and naming facility (NA; often referred to as RAN, or rapid automatic naming, in the reading literature).

The addition of Matrix Reasoning to the WAIS-III was a useful move that addresses prior criticisms that the various Wechsler batteries have never contained an appreciable measure of *Gf* (Flanagan et al., 2000; Flanagan & Ortiz, 2001; McGrew & Flanagan, 1996; McGrew & Flanagan, 1998; Woodcock, 1990). The WJ III Analysis–Synthesis test, which is classified as a measure of general sequential (deductive) reasoning (RG), would be an ideal CB supplement to Matrix Reasoning, which measures induction (I). However, this combination does not provide a score based on actual norms. If a norm-based *Gf* score is desired, then both the WJ III Concept Formation (I) and Analysis–Synthesis (RG) tests can be administered to obtain the broad WJ III Fluid Reasoning cluster.[13]

[13]The KAIT Logical Steps and Mystery Codes tests can provide a norm-based *Gf* score. The KAIT can also be viewed as a battery to use to supplement the WAIS-III. This function, as well as the possibility of using the WAIS-III to supplement the KAIT, is not addressed directly in this chapter. Table 14.10 can be inspected to perform these functions.

Supplementing the KAIT with the WJ III

Table 14.10 can also be used in a similar manner for KAIT CB assessments. The WJ III can address the KAIT's construct underrepresentation of *Gv*, *Ga*, *Gs*, and *Gsm*. In contrast, the KAIT provides for valid measurement of *Gf*. The WJ III *Gf* tests could help examiners if there is a need for in-depth narrow ability assessment of inductive and deductive abilities or if a sampling of quantitative reasoning (RQ) is necessary. Similarly, the WJ III Visual–Auditory Learning and Delayed Recall-Visual–Auditory Learning could be used to supplement the KAIT Rebus Learning test in the pursuit of an associative memory (MA) narrow ability composite score. Similar to the case of WAIS-III/WMS-III CB assessment, the WJ III can make a unique contribution to KAIT CB assessments in the measurement of ideational fluency (FI) and naming facility (NA).

Supplementation of the KAIT *Gc* tests would depend on the specific circumstances of the assessment. As reported in Table 14.10, the KAIT Definitions and Double Meanings tests, although displaying *Gc* factor loadings, also contain construct irrelevant *Grw* (reading and spelling) variance. If a strict Carroll (1993) model were followed, this extraneous *Grw* variance would not necessarily be considered a weakness of these two tests. Carroll includes *Grw* abilities under *Gc*. From Carroll's theoretical perspective, the KAIT *Gc* tests would be viewed as good indicators of *Gc* for most adolescents and adults. However, if an examinee has problems with reading and spelling, and more importantly, if the referral is for academic/learning problems, the KAIT *Gc* tests may be inappropriately impacted by reading and writing deficiencies. In such situations, the WJ III *Gc* tests could be added to a KAIT CB assessment.

Supplementing the WJ III

As previously described, the WJ III COG was designed to provide two or more qualitatively different narrow abilities within each broad

CHC domain. When viewed from the perspective of CHC CB assessment, the WJ III cognitive battery requires little, if any, supplementing in order to provide adequate ability coverage at the broad *Gf–Gc* level. This can be ascertained by reviewing the number and variety of narrow abilities measured by the WJ III in Table 14.10.

The WAIS-III/WMS-III and KAIT could be used to broaden the WJ III's *Gv* coverage via the addition of a measure of closure speed (WAIS-III Object Assembly). The WJ III *Glr* coverage might also benefit from a measure of free recall memory (WMS-III Word Lists I & II). Additionally, the WJ III Quantitative Concepts achievement test may not be an optimal indicator of quantitative reasoning (*Gf*, RQ), given that half of the test measures math knowledge (*Gq*, KM). The primary area in which the WJ III may benefit from CB supplemental testing is the addition of tests for in-depth exploration at the narrow ability level, a topic not discussed in detail here (see Flanagan & Ortiz, 2001, for a detailed explanation of these procedures).

Summary

This chapter provides an overview of one of the newest revisions of a comprehensive test of intelligence: the Woodcock-Johnson III. A brief history of the family of Woodcock-Johnson tests is provided, which helps to explain how the changes for the test's third edition came into existence. The WJ III's theoretical foundation is the Cattell-Horn-Carroll (CHC) theory of cognitive abilities. This theory is described in some detail, as well as its application to the WJ III.

The WJ III is a comprehensive collection of individually administered co-normed tests organized as two distinct test batteries. The Woodcock-Johnson Tests of Cognitive Abilities (WJ III COG) and the Woodcock-Johnson Tests of Achievement (WJ III ACH) are designed to measure a wide array of cognitive, oral language, and academic achievement abilities for individuals age 2 through the geriatric population. Although the focus of this book and this chapter is on cognitive abilities, the WJ III ACH tests play a significant role in the conceptual framework of the overall WJ III; thus, they are included here. The WJ III includes 20 Tests of Cognitive Abilities and 22 Tests of Achievement, each of which are described and categorized within tables in the chapter. The cognitive tests are organized by both the broad CHC clusters and by three broader categories related to cognitive performance: verbal ability, thinking ability, and cognitive efficiency. The achievement tests are organized by curricular area (reading, mathematics, written language, and academic knowledge) and oral language and by clusters within these areas, with additional groupings for special purpose clusters.

Interpretation of the WJ III involves careful examination of the multiple scores that the test yields. A variety of scores are provided by the WJ III, including age and grade equivalents, relative proficiency indexes (RPI), cognitive–academic language proficiency (CALP) levels, percentile ranks, and standard scores. The WJ III's scores were derived from an exceptionally large standardization sample: a total of 8,818 individuals, from age 24 months to 95+ years, living in more than 100 geographically and economically diverse communities in the United States, were assessed. The psychometric properties of the WJ III are quite strong. Eighteen of the cognitive test median reliabilities meet or exceed the .80 level standard and 12 meet or exceed the .90 standard. All 22 achievement test median reliabilities exceed .80 and 11 meet or exceed the more stringent .90 standard. Across the WJ III cognitive and achievement batteries almost all of the 42 clusters meet or exceed the .90 reliability standard. Support for the validity of many WJ III tests and clusters has accumulated across three different versions of the battery. The evidence that specifically pertains to WJ III is summarized in this chapter.

The latter half of the chapter presents special applications and use with special populations, including applications with individuals with learning disabilities, neuropsychological applications, CHC abilities across the life span, and CHC cross-battery applications. The cross-battery section focuses on how the WJ III may be used in conjunction with other comprehensive intelligence tests such as the WAIS-III and the KAIT.