Woodcock Johnson IV

Assessment Service Bulletin Number 2

WJ IV™ Technical Abstract

Erica M. LaForte, PhD

Kevin S. McGrew, PhD

Fredrick A. Schrank, PhD

This bulletin provides a summary of the procedures followed in developing and validating the Woodcock-Johnson IV *(WJ IV) (Schrank, McGrew, & Mather, 2014a) as a comprehensive measure of individuals' cognitive abilities, oral language abilities, and academic achievement. Throughout the development and design of the WJ IV, the test standards outlined in the* Standards for Educational and Psychological Testing *(American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014) were followed carefully. Information in this bulletin is abstracted from the* Woodcock-Johnson IV Technical Manual *(McGrew, LaForte, & Schrank, 2014) and is intended as an overview to highlight important aspects of the WJ IV test design, reliability, and validity. Readers who are interested in more detailed information should consult the WJ IV Technical Manual.*

Houghton Mifflin Harcourt

Reference Citation

- To cite this document, use:

  LaForte, E. M., McGrew, K. S., & Schrank, F. A. (2014). *WJ IV Technical Abstract* (Woodcock-Johnson IV Assessment Service Bulletin No. 2). Rolling Meadows, IL: Riverside.

For technical information, please visit www.riversidepublishing.com or call Riverside Customer Service at 800.323.9540.

WJ IV™ Technical Abstract

The *Woodcock-Johnson®IV* (WJ IV) (Schrank, McGrew, & Mather, 2014a) consists of three distinct, co-normed batteries: the *Woodcock-Johnson IV Tests of Cognitive Abilities* (WJ IV COG) (Schrank, McGrew, & Mather, 2014b), the *Woodcock-Johnson IV Tests of Oral Language* (WJ IV OL) (Schrank, Mather, & McGrew, 2014b), and the *Woodcock-Johnson IV Tests of Achievement* (WJ IV ACH) (Schrank, Mather, & McGrew, 2014a). Together, these batteries form a comprehensive system for measuring general intellectual ability (*g*), specific cognitive abilities, oral language abilities, and academic achievement across a wide age range. Normative data are based on a large, nationally representative sample of 7,416 individuals ranging in age from 2 to 90+ years.

Overview of the WJ IV

The WJ IV is a theoretical, structural, and interpretive revision of the *Woodcock-Johnson III* (WJ III®) (Woodcock, McGrew, & Mather, 2001) designed to provide measures of general intellectual ability; broad and narrow cognitive abilities as defined by contemporary Cattell-Horn-Carroll (CHC) theory, (Schneider & McGrew, 2012) including oral language, reading, mathematics, writing, and academic domain-specific aptitudes; and academic knowledge. The following WJ IV revision goals and design objectives maintain the traditional Woodcock-Johnson focus on quality, while advancing CHC theory from its initial articulation in the *Woodcock-Johnson Psycho-Educational Battery–Revised* (WJ-R®) (Woodcock & Johnson, 1989) and WJ III in ways that provide more administration and interpretive options to meet contemporary assessment needs.

1. The WJ IV is designed to provide the most contemporary measurement model of an evolving CHC theory of human cognitive abilities by:
    a. Creating new tests and interpretive clusters, based on extant research and professional practice needs, to measure the most important cognitive, oral language, and academic abilities;
    b. Focusing on the ecological and predictive validity of key interpretive clusters in the cognitive, oral language, and achievement batteries by increasing the cognitive complexity of selected tests; and
    c. Offering a new fluid-crystallized (*Gf-Gc*) cognitive composite in the WJ IV COG for comparison to other measures of cognitive processing, linguistic competency, and academic achievement to determine relative strengths and weaknesses across all domains.
2. The WJ IV is organized into three distinct, co-normed batteries that can be used independently or in any combination to provide greater flexibility for professional examiners by:
    a. Recognizing the importance of oral language abilities as essential correlates of cognitive and academic functioning and making the WJ IV OL available to examiners who conduct cognitive ability, language proficiency, or academic achievement evaluations;

b. Including an overall index of oral language ability in Spanish as well as in English and suggesting a practical option for administering the Spanish tests; and
c. Offering three parallel forms of the WJ IV ACH Standard Battery to avoid overexposure to items on any given form.

3. The WJ IV provides new and useful options for comparing abilities within and across batteries by:
   a. Offering options to explore individual strengths and weaknesses across cognitive, linguistic, and academic abilities;
   b. Organizing each battery for ease of use, leading with a core set of tests in each battery that can be used as a predictor pool for calculations identifying relative strengths and weaknesses among administered tests and clusters; and
   c. Creating new domain-specific scholastic aptitude clusters that allow for efficient and valid predictions of academic achievement.
4. The WJ IV retains the focus on psychometric quality that has been associated with the previous editions of Woodcock-Johnson batteries by:
   a. Providing a new, large, and nationally representative norming sample drawn from the U.S. population;
   b. Updating items and simplifying test administration and interpretation procedures;
   c. Augmenting the underlying scaling of speeded tests; and
   d. Utilizing state-of-the-art data collection, test development, and data analytic methods as models to facilitate progress in the field of applied test development.

Theoretical Foundation of the WJ IV

The WJ IV represents the fourth generation of a comprehensive battery of psycho-educational tests that originated in 1973 and was first published as the *Woodcock-Johnson Psycho-Educational Battery* (WJ) (Woodcock & Johnson, 1977). The theoretical foundation of the WJ IV is derived from the CHC theory of cognitive abilities, which stems from the psychometric factor-analytic work of Raymond Cattell (1941, 1943, 1950), John Horn (1988, 1991), and John Carroll (1993, 1998).

CHC theory has evolved beyond its initial specifications (Schneider & McGrew, 2012) through both simplification and elaboration. In addition, throughout the development of the WJ IV, other venues of research have been examined to cross-validate, modify, or add clarity to some of the theoretical constructs proposed by Cattell, Horn, Carroll, Woodcock, and their colleagues. The interpretive model for the WJ IV reflects the most contemporary specification of CHC theory at the time of publication. Analysis of the WJ-R, WJ III, and WJ IV norming samples provided three large, multi-ability data sets that were used to either confirm or revise initial construct specifications. Support for changes to the interpretive constructs also was gleaned from other sources of neuroscience research. The most significant changes to the WJ IV interpretive model are found in the contemporary constructs of working memory, auditory processing, speed of lexical access, and memory for sound patterns.

Battery Organization

The complete WJ IV is organized into three distinct batteries to facilitate a broad range of tailored and comprehensive assessments by one or more professionals. Depending on the purpose of the assessment, the three batteries may be used alone or in combination with tests and clusters from one or both of the other batteries. The complete system of WJ IV tests and clusters, designed for individually administered assessment of important abilities in a variety of settings, provides a wide age range and breadth of coverage that allows the tests and clusters to be used for educational, clinical, or research purposes from the preschool to the geriatric level. Tables 1 through 3 summarize the tests and interpretive cluster configurations for the WJ IV COG, WJ IV OL, and WJ IV ACH batteries.

Table 1.
Selective Testing Table for the WJ IV COG Showing Tests and Interpretive Clusters

			Cognitive Composites			CHC Factors								Narrow Ability and Other Clinical Clusters				
			General Intellectual Ability (GIA)	Brief Intellectual Ability	Gf-Gc Composite	Comprehension-Knowledge (Gc)	Fluid Reasoning (Gf)	Short-Term Working Memory (Gwm)	Cognitive Processing Speed (Gs)	Auditory Processing (Ga)	Long-Term Retrieval (Glr)	Visual Processing (Gv)	Quantitative Reasoning (RQ)	Auditory Memory Span (MS)	Number Facility (N)	Perceptual Speed (P)	Vocabulary (VL/LD)	Cognitive Efficiency
Standard Battery	COG 1	Oral Vocabulary	■	■	■	■										■		
	COG 2	Number Series	■	■	■		■						■					■
	COG 3	Verbal Attention	■	■				■										□
	COG 4	Letter-Pattern Matching	■						■						■			■
	COG 5	Phonological Processing	■							■								
	COG 6	Story Recall	■								■							
	COG 7	Visualization	■									■						
	COG 8	General Information			■	■												
	COG 9	Concept Formation			■		■											
	COG 10	Numbers Reversed						■						■				■
Extended Battery	COG 11	Number-Pattern Matching												■	■			□
	COG 12	Nonword Repetition								■								
	COG 13	Visual-Auditory Learning									■							
	COG 14	Picture Recognition										■						
	COG 15	Analysis-Synthesis					□						■					
	COG 16	Object-Number Sequencing						□										
	COG 17	Pair Cancellation							■									
	COG 18	Memory for Words												■				
Oral Language Battery	OL 1	Picture Vocabulary			□											■		
	OL 5	Sentence Repetition												■				

■ Tests required to create the cluster listed.
□ Additional tests required to create an extended version of the cluster listed.

Table 2.
Selective Testing Table for the WJ IV OL Showing Tests and Interpretive Clusters

			Oral Language Clusters										OL + COG
			Oral Language	Broad Oral Language	Oral Expression	Listening Comprehension	Phonetic Coding	Speed of Lexical Access	Lenguaje oral	Amplio lenguaje oral	Comprensión auditiva	Vocabulary (VL/LD)	Auditory Memory Span (MS)
Oral Language Battery	OL 1	Picture Vocabulary	■	■	■							■	
	OL 2	Oral Comprehension	■	■		■							
	OL 3	Segmentation					■						
	OL 4	Rapid Picture Naming						■					
	OL 5	Sentence Repetition			■								■
	OL 6	Understanding Directions		■		■							
	OL 7	Sound Blending					■						
	OL 8	Retrieval Fluency						■					
	OL 9	Sound Awareness[1]											
	OL 10	Vocabulario sobre dibujos							■	■			
	OL 11	Comprensión oral							■	■	■		
	OL 12	Comprensión de indicaciones								■	■		
Cognitive Abilities Battery	COG 1	Oral Vocabulary										■	
	COG 18	Memory for Words											■

■ Tests required to create the cluster listed.

[1] This is a screening test and does not contribute to a cluster.

Table 3.
Selective Testing Table for the WJ IV ACH Showing Tests and Interpretive Clusters

			Reading						Mathematics				Writing						Cross-Domain Clusters			
			Reading	Broad Reading	Basic Reading Skills	Reading Comprehension	Reading Fluency	Reading Rate	Mathematics	Broad Mathematics	Math Calculation Skills	Math Problem Solving	Written Language	Broad Written Language	Basic Writing Skills	Written Expression	Academic Skills	Academic Fluency	Academic Applications	Academic Knowledge	Phoneme-Grapheme Knowledge	Brief (or Broad) Achievement
Standard Battery	ACH 1	Letter-Word Identification	■	■	■												■					■
	ACH 2	Applied Problems							■	■		■							■			■
	ACH 3	Spelling											■	■	■		■					■
	ACH 4	Passage Comprehension	■	■		■													■			●
	ACH 5	Calculation							■	■	■						■					●
	ACH 6	Writing Samples											■	■		■			■			●
	ACH 7	Word Attack			■																■	
	ACH 8	Oral Reading					■															
	ACH 9	Sentence Reading Fluency		■			■	■									■					●
	ACH 10	Math Facts Fluency								■	■						■					●
	ACH 11	Sentence Writing Fluency												■		■	■					●
Extended Battery	ACH 12	Reading Recall				■																
	ACH 13	Number Matrices									■											
	ACH 14	Editing												■								
	ACH 15	Word Reading Fluency					■															
	ACH 16	Spelling of Sounds																			■	
	ACH 17	Reading Vocabulary				□																
	ACH 18	Science																	■			
	ACH 19	Social Studies																	■			
	ACH 20	Humanities																	■			

■ Tests required to create the cluster listed.

□ Additional test required to create an extended version of the cluster listed.

● Additional tests required to create the Broad Achievement cluster.

Norming Study

The WJ IV norming study included data collected between December 2009 and January 2012 from 7,416 individuals from geographically diverse communities representing 46 states and the District of Columbia. The norming sample contained 664 children ages 2 through 5 years who were not enrolled in kindergarten, 3,891 examinees enrolled in kindergarten through grade 12, 775 college undergraduate and graduate students, and 2,086 adults (ages 18 and up) who were not enrolled in high school or college. Table 4 displays the distribution of the WJ IV norming sample by age and grade. The higher density of examinees in the school-age population (kindergarten through grade 12) reflects the need to collect more concentrated data from examinees in this age range, where the abilities measured by the WJ IV undergo the greatest rate of growth.

Table 4.
Distribution of the WJ IV Norming Sample by Age and Grade

Age	Number	Grade	Number
2	173	Kindergarten	308
3	203	1	334
4	223	2	303
5	205	3	312
6	308	4	327
7	310	5	328
8	336	6	330
9	306	7	294
10	314	8	313
11	329	9	289
12	317	10	269
13	307	11	256
14	299	12	228
15	277		
16	284	College and University	
17	254	13	205
18	276	14	190
19	295	15	104
20–29	759	16	104
30–39	492	17+ (graduate students)	172
40–49	462		
50–59	274		
60–69	164		
70–79	132		
80+	117		
Total	7,416	**Total**	4,666

The WJ IV norming sample was selected to be representative, within practical limits, of the U.S. population of individuals from ages 2 to 90+ years. Examinees were randomly selected within a stratified sampling design that controlled for several community and examinee variables. Table 5 contains the sampling variables and their distribution, both in the U.S. population according to the 2010 census projections and in the WJ IV norming sample, for the school-age sample. The *Woodcock-Johnson IV Technical Manual* (McGrew, LaForte, & Schrank, 2014) provides similar information for the other major age groupings (preschool, college/university, and adult). Some variables were not relevant at all age levels of the norming sample. For example, occupational information

was applied only to the adults in the sample, and type of college was applied only to the individuals enrolled in college.

Table 5.
Distribution of Sampling Variables in the U.S. Population and in the WJ IV Norming Sample–Grades K Through 12

Sampling Variable	Percentage in U.S. Population	Number Obtained	Percentage in Norming Sample	Partial Examinee Weight
Census Region				
Northeast	17.4	652	16.8	1.039
Midwest	21.8	991	25.5	0.854
South	37.2	1,246	32.0	1.163
West	23.6	1,002	25.8	0.916
Community Type				
Metropolitan	83.7	3,323	85.4	0.980
Micropolitan	10.0	372	9.6	1.048
Rural	6.3	196	5.0	1.253
Sex				
Male	51.0	1,924	49.4	1.032
Female	49.0	1,967	50.6	0.969
Country of Birth				
United States	95.0	3,802	97.7	0.972
Other	5.0	88	2.3	2.209
Race/Ethnicity				
White, Not Hispanic	63.7	2,460	63.2	0.984
Black, Not Hispanic	12.5	537	13.8	0.886
AIANAT[c], Not Hispanic	0.8	21	0.5	1.446
ASIPAC[d], Not Hispanic	5.2	164	4.2	1.209
Other, Not Hispanic	—[a]	7	0.2	1.000[b]
White, Hispanic	16.6	591	15.2	1.071
Black, Hispanic	0.7	12	0.3	2.205
AIANAT[c], Hispanic	0.3	6	0.2	1.835
ASIPAC[d], Hispanic	0.2	11	0.3	0.598
Other, Hispanic	—[a]	82	2.1	1.000[b]
Parent Education				
< High School	13.7	502	12.9	1.060
High School	22.7	1,179	30.3	0.747
> High School	63.6	2,198	56.5	1.122
School Type				
Public	85.7	3,483	89.5	0.957
Private	11.4	314	8.1	1.413
Home	2.9	92	2.4	1.227

[a] No reliable population information could be obtained.

[b] Null partial weights of 1.000 were assigned to cells for which reliable population information could not be obtained or for which the sample counts were so low that they inappropriately skewed examinees' overall weights.

[c] AIANAT = American Indian or Alaska Native.

[d] ASIPAC = Asian, Native Hawaiian, or Other Pacific Islander.

Because it was not practical for all WJ IV norming study participants to be administered all 51 tests in the norming edition, a planned incomplete data collection design was used in the study. Planned incomplete (missing) data collection methods (Graham, Taylor, Olchowski, & Cumsille, 2006; McArdle, 1994; McKnight, McKnight, Sidani, & Figueredo, 2007; Rhemtulla & Little, 2012; Schafer, 1997; Wolf, 2006) have been developed as a statistically sound technique for gathering data in large studies where design constraints preclude complete data collection. WJ IV norming study

participants were each randomly assigned to one block of tests (either a unique Block A, Block B, or Block C, or a linking Block D), each of which contained between 15 and 19 tests. Tests that were unique to Block A, Block B, or Block C were administered to approximately 1,500 to 2,200 examinees in the WJ IV norming study, whereas the Block D tests were administered to a larger sample, ranging from 3,500 to 3,800 examinees per test, depending on the age range of the test. Best practice approaches to generating plausible *W* scores for tests not taken by norming study participants were then utilized to generate a "complete record" for all norming study participants. Details regarding the block design criteria, study constraints, and data imputation methods can be found in the WJ IV Technical Manual (McGrew et al., 2014).

Data from the 7,416 norming study participants were summarized for each test and cluster. Individual examinee weights were applied during the norms construction process to ensure that the test, cluster, and difference score norms were based on a sample with characteristics proportional to the U.S. population distribution. The weight for each norming study participant was obtained by calculating the product of several partial weights, each corresponding to a demographic variable for the applicable sampling group (preschool, kindergarten through grade 12, college/university, or adult). For each demographic variable, if an examinee belonged to a category of the variable that was overrepresented in the WJ IV norming study sample, the examinee's partial weight for that variable was less than 1.00. Likewise, if the examinee belonged to a category of the variable that was underrepresented in the WJ IV norming study sample, the examinee's partial weight for that variable was greater than 1.00. Table 5 contains the partial weights assigned for each demographic variable value within the kindergarten through grade 12 sample of examinees. If demographic information was missing for a particular examinee on a particular variable, that examinee was assigned a null (1.00) partial weight for that variable. A partial weight of 1.00 is considered null because when it is multiplied together with the other partial weights to compute a total norming study participant weight, a value of 1.00 has no effect on the overall weight. For some variables (indicated with superscript *b* in Table 5), null partial weights of 1.00 also were assigned to cells for which reliable population information could not be obtained or for which the sample counts were so low that they inappropriately skewed examinees' overall weights (e.g., the "Other, Not Hispanic" values of the race/ethnicity variable).

Calculation of WJ IV Cluster Scores

With the exception of the WJ IV COG General Intellectual Ability (GIA) cluster, all WJ IV COG, WJ IV OL, and WJ IV ACH cluster scores are based on the arithmetic average of the *W* scores of the tests that contribute to the cluster score.

General Intellectual Ability Cluster

The GIA is a general intelligence (*g*) score; it represents the first principal component obtained from principal component analysis (PCA). In PCA, the optimal weighted combination of tests that account for the largest proportion of the variance in a collection of tests is extracted as the first component (similar to a factor in factor analysis). Table 6 presents the GIA average smoothed *g* weights by 25 technical age groups.[1] A review of Table 6 reveals that the weights for the individual tests fluctuate little as a function of age.

[1] Technical age groups for the calculation of the GIA *g* weights included examinees in one-year age intervals from age 2 through 19, and 10-year age intervals from age 20 through 79. The 80+ group included all norming examinees older than age 79.

Table 6.
General Intellectual Ability Average (Smoothed) g *Weights by Technical Age Groups*

WJ IV COG Test	CHC Domain	AGE								
		2	3	4	5	6	7	8	9	10
Test 1: Oral Vocabulary	*Gc*	0.16	0.16	0.16	0.17	0.17	0.17	0.18	0.18	0.18
Test 2: Number Series	*Gf*	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18
Test 3: Verbal Attention	*Gwm*	0.13	0.13	0.13	0.13	0.13	0.14	0.14	0.14	0.14
Test 4: Letter-Pattern Matching	*Gs*	0.17	0.16	0.16	0.15	0.14	0.12	0.11	0.11	0.10
Test 5: Phonological Processing	*Ga*	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.17
Test 6: Story Recall	*Glr*	0.11	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
Test 7: Visualization	*Gv*	0.07	0.07	0.07	0.08	0.08	0.09	0.10	0.10	0.11

WJ IV COG Test	CHC Domain	AGE								
		11	12	13	14	15	16	17	18	19
Test 1: Oral Vocabulary	*Gc*	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18
Test 2: Number Series	*Gf*	0.18	0.18	0.17	0.17	0.17	0.17	0.17	0.17	0.17
Test 3: Verbal Attention	*Gwm*	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
Test 4: Letter-Pattern Matching	*Gs*	0.10	0.10	0.10	0.10	0.11	0.11	0.11	0.11	0.11
Test 5: Phonological Processing	*Ga*	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
Test 6: Story Recall	*Glr*	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
Test 7: Visualization	*Gv*	0.11	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12

WJ IV COG Test	CHC Domain	AGE							
		20–29	30–39	40–49	50–59	60–69	70–79	80+	Median
Test 1: Oral Vocabulary	*Gc*	0.18	0.18	0.18	0.18	0.17	0.16	0.16	0.18
Test 2: Number Series	*Gf*	0.16	0.16	0.15	0.15	0.15	0.15	0.15	0.17
Test 3: Verbal Attention	*Gwm*	0.13	0.14	0.14	0.14	0.15	0.15	0.15	0.14
Test 4: Letter-Pattern Matching	*Gs*	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
Test 5: Phonological Processing	*Ga*	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.17
Test 6: Story Recall	*Glr*	0.11	0.11	0.12	0.12	0.12	0.12	0.13	0.12
Test 7: Visualization	*Gv*	0.13	0.12	0.12	0.12	0.12	0.12	0.12	0.12

Scholastic Aptitude Clusters

A major function of the WJ IV is to provide statements regarding a person's predicted performance in different achievement domains and to make comparisons between predicted and actual achievement within these domains. The six WJ IV Scholastic Aptitude (SAPT) cluster scores (two for each achievement domain) are designed to provide optimal and efficient prediction of expected achievement in each domain. Each SAPT cluster score is based on a combination of four tests, each from a different CHC domain, that together produce the strongest and most efficient prediction for the selected achievement area. Although SAPTs were included in the WJ and the WJ-R, the WJ IV SAPTs represent an advance over those from the earlier batteries because they were constructed from statistical prediction combined with research and theoretical

considerations. The WJ IV SAPTs also differ by academic areas, providing the best prediction of achievement skills. These advances were incorporated into the WJ IV SAPT clusters based on research indicating that predictor tasks vary within broad achievement areas and that these predictors change developmentally (McGrew, 2012; McGrew & Wendling, 2010; Schneider & McGrew, 2012).

Calculation of Norms

The development of test norms and derived scores requires the establishment of the normative (average) score for each measure for individuals at each specific age (age norms) or grade (grade and college/university norms) where normative interpretations are intended. In the WJ family of instruments, this normative score is called the Reference *W* (REF *W*) score. When plotted as a function of chronological age or grade, the REF *W* curves serve as the foundation for the age- and grade-equivalent scores, relative proficiency index (RPI), and instructional range interpretation features of the WJ IV. In addition, when the standard deviations (*SD*s) of the scores at each age or grade are plotted as a function of age or grade, the resultant curves represent the *SD* values that, when combined with REF *W* values, provide the foundation for the calculation of all other norm-referenced score metrics (e.g., standard scores and percentile ranks).

Bootstrap Resampling Procedures

The innovative *bootstrap resampling* procedures (Efron & Tibshirani, 1993), first implemented and described for the *WJ III Normative Update* (McGrew, Dailey, & Schrank, 2007), were used to calculate the WJ IV norms. The use of bootstrap resampling procedures allows for the incorporation of estimates of uncertainty and potential bias (in the sample data) in the calculation of the WJ IV norms. When compared to more traditional norm development procedures (such as those used in the WJ [Woodcock & Johnson, 1977], WJ-R [Woodcock & Johnson, 1989], WJ III [Woodcock et al., 2001], and most other individually administered cognitive ability, oral language, and academic achievement batteries), the bootstrap-based procedures used to calculate the WJ IV norms produce more precise estimates of an examinee's ability.

Difference Score Norms

Difference scores allow users to make data-based predictions and comparisons among selected test or cluster scores derived from the WJ IV batteries, which then can be used to describe performance patterns that may be useful for diagnostic decision making and educational planning. The two most common uses for difference scores in assessment practice are as follows:

1. To determine whether an examinee's relative standing in a group on an individual test or cluster (e.g., WJ IV COG Test 2: Number Series) is significantly different from the examinee's relative standing in the same group on another individual test or cluster (e.g., WJ IV COG Test 7: Visualization).
2. To determine whether an examinee's score on an individual test or cluster is significantly different from what would be expected or predicted, given his or her score on some predictor test or cluster.

The first example above is a *standard score/percentile rank profile difference*. The second example above relies on the distribution of actual differences between *predictor* and *criterion* scores in the WJ IV norming study group.

One benefit of co-norming the WJ IV COG, WJ IV OL, and WJ IV ACH batteries is that it allows computation of actual differences between predictor and criterion variables for each individual in the WJ IV norming sample, which can then be used to model these differences in the population. In the WJ IV, this type of difference score takes two forms: *variations* and *comparisons*. All WJ IV variation and comparison procedures are based on a common statistical model. What distinguishes variations from comparisons is the score that is used as the predictor in the model. While variations rely on a predictor score that is an average of the (noncriterion) scores from a pool of tests that excludes the criterion measure, comparisons rely on a single predictor, such as the GIA or *Gf-Gc* Composite cluster score. The scoring algorithms for these WJ IV difference-based variations and comparisons were constructed using a regression-based procedure similar to that used to calculate discrepancy scores in the WJ-R and WJ III batteries (McGrew, Werder, & Woodcock, 1991; McGrew & Woodcock, 2001).

Because all tests in the WJ IV are co-normed, the variation and comparison difference scores do not contain error that is inherent in measures based on different samples. Additionally, examiners who use the WJ IV variation and comparison difference norms can evaluate the significance of a difference in the population by inspecting either the percentile rank of the difference score (discrepancy PR) or the difference between the achievement score and the predicted achievement score in standard error of estimate units (discrepancy *SD*). This feature allows a professional, school district, or state to define a criterion of significance in terms of either the discrepancy *SD* or the discrepancy PR. The discrepancy *SD* allows the criterion to be defined in terms of the distance of an individual's score from the average score for that subgroup of the norming sample (i.e., individuals of the same age or same grade). The discrepancy PR allows the criterion to be defined in terms of the percentage of the population identified as possessing a discrepancy of a specified direction and magnitude (i.e., the base rate).

Reliability

The reliability coefficient can be thought of as an index of precision by which relative standing or position in a group is measured. High reliability implies that an individual's relative standing in the group would be similar across repeated administrations of the test.

WJ IV Speeded Tests

The WJ IV speeded tests were calibrated using a rate-based metric, whereby each examinee's score was converted into a rate of correct response, and each minute of testing was treated as an item. While this rate-based metric is useful for calibrating items and rank-ordering examinees, it provides inflated standard errors for ability measures due to the limited number of possible scores for each time interval. For this reason, the procedures for calculating WJ IV nonspeeded test reliability coefficients were not appropriate for the speeded tests. Instead, a test-retest study was conducted for all WJ IV speeded tests. Details of the study design can be found in the WJ IV Technical Manual (McGrew et al., 2014). Table 7 contains the median test-retest reliability coefficients (r_{11}) across the three age groups (ages 7 through 11, ages 14 through 17, and ages 26 through 79) from the study.

Table 7.
Median Test-Retest Reliability Coefficients for WJ IV Speeded Tests Across Three Age Groups (7 Through 11, 14 Through 17, and 26 Though 79)

Test	Median r_{11}	Test	Median r_{11}
WJ IV COG Standard Battery		**WJ IV ACH Standard Battery**	
Test 4: Letter-Pattern Matching	0.91	Test 9: Sentence Reading Fluency	0.93
WJ IV COG Extended Battery		Test 10: Math Facts Fluency	0.95
Test 11: Number-Pattern Matching	0.85	Test 11: Sentence Writing Fluency	0.83
Test 17: Pair Cancellation	0.89	**WJ IV ACH Extended Battery**	
WJ IV OL Battery		Test 15: Word Reading Fluency	0.92
Test 4: Rapid Picture Naming	0.90		

WJ IV Nonspeeded Tests

Reliability coefficients were calculated for all WJ IV nonspeeded tests across the age ranges of intended use and included data from all norming study participants. Internal-consistency reliabilities for all dichotomously scored WJ IV nonspeeded tests and subtests were calculated using the split-half procedure. Raw scores were computed for the norming study participants based on the odd and even items in these tests. Correlation coefficients were then computed between the two item sets. These coefficients were corrected for published test length using the Spearman-Brown correction formula. The reliability coefficients for the tests containing multiple-point items (e.g., WJ IV ACH Test 6: Writing Samples and Test 8: Oral Reading) were calculated using information provided by the Rasch model.

Composite Tests

For tests with subtests, Mosier's (1943) formula was used to compute composite test reliabilities using the individual subtest reliabilities obtained through either the test-retest method for speeded tests or the split-half method for dichotomously scored nonspeeded tests.

Table 8 reports the median test reliability coefficients (r_{11}) and the standard errors of measurement in standard score units (*SEM* SS) for all WJ IV nonspeeded tests, obtained using the procedures described above. A review of the median test reliability coefficients reveals the extent to which the reliabilities fall at the desired level of .80 or higher. Of the 39 median test reliability coefficients reported in Table 8, 38 are .80 or higher and 17 are .90 or higher. Although these are strong reliabilities for individual tests, the WJ IV cluster scores are recommended for making important decisions about an individual due to the higher reliabilities of those scores.

Table 8.
Median Test Reliability Coefficients and Standard Errors of Measurement in Standard Score Units for WJ IV Nonspeeded Tests

Test	Median r_{11}	Median *SEM* (SS)	Test	Median r_{11}	Median *SEM* (SS)
WJ IV COG Standard Battery			**WJ IV ACH Standard Battery**		
Test 1: Oral Vocabulary	0.89	4.97	Test 1: Letter-Word Identification	0.94	3.78
Test 2: Number Series	0.91	4.64	Test 2: Applied Problems	0.92	4.27
Test 3: Verbal Attention	0.86	5.70	Test 3: Spelling	0.92	4.13
Test 5: Phonological Processing	0.84	6.00	Test 4: Passage Comprehension	0.89	5.00
Test 6: Story Recall	0.93	3.90	Test 5: Calculation	0.93	3.86
Test 7: Visualization	0.85	5.81	Test 6: Writing Samples	0.90	4.74
Test 8: General Information	0.88	5.20	Test 7: Word Attack	0.90	4.75
Test 9: Concept Formation	0.93	4.04	Test 8: Oral Reading	0.96	3.00
Test 10: Numbers Reversed	0.88	5.15			
WJ IV COG Extended Battery			**WJ IV ACH Extended Battery**		
Test 12: Nonword Repetition	0.91	4.55	Test 12: Reading Recall	0.92	4.34
Test 13: Visual-Auditory Learning	0.97	2.65	Test 13: Number Matrices	0.92	4.31
Test 14: Picture Recognition	0.74	7.70	Test 14: Editing	0.91	4.60
Test 15: Analysis-Synthesis	0.93	4.02	Test 16: Spelling of Sounds	0.88	5.13
Test 16: Object-Number Sequencing	0.89	4.95	Test 17: Reading Vocabulary	0.88	5.19
Test 18: Memory for Words	0.82	6.39	Test 18: Science	0.84	5.91
WJ IV OL Battery			Test 19: Social Studies	0.87	5.48
Test 1: Picture Vocabulary	0.88	5.27	Test 20: Humanities	0.87	6.24
Test 2: Oral Comprehension	0.82	6.32			
Test 3: Segmentation	0.94	3.68			
Test 5: Sentence Repetition	0.83	6.10			
Test 6: Understanding Directions	0.87	5.35			
Test 7: Sound Blending	0.89	4.94			
Test 8: Retrieval Fluency	0.80	6.71			
Test 9: Sound Awareness	0.82	6.36			

Clusters

The reliability coefficients for the WJ IV clusters were computed using Mosier's (1943) formula. Table 9 reports the median reliability coefficients and *SEM* SSs for all WJ IV clusters. A review of the median reliability coefficients for each cluster reveals that most are .90 or higher.

Table 9.
Median Reliability Coefficients and Standard Errors of Measurement in Standard Score Units for WJ IV Clusters

Cluster	Median r_{11}	Median *SEM* (SS)	Cluster	Median r_{11}	Median *SEM* (SS)
General Intellectual Ability	0.97	2.60	Oral Expression	0.89	4.97
Brief Intellectual Ability	0.94	3.67	Listening Comprehension	0.90	4.74
Gf-Gc Composite	0.96	3.00	Phonetic Coding	0.95	3.35
Comprehension-Knowledge	0.93	3.97	Speed of Lexical Access	0.89	4.97
Comprehension-Knowledge–Extended	0.94	3.67	Vocabulary	0.93	4.97
Fluid Reasoning	0.94	3.67	Reading	0.95	3.35
Fluid Reasoning–Extended	0.96	3.00	Broad Reading	0.97	2.80
Short-Term Working Memory	0.91	4.50	Basic Reading Skills	0.95	3.35
Short-Term Working Memory–Extended	0.93	3.97	Reading Comprehension	0.93	3.97
Cognitive Processing Speed	0.94	3.67	Reading Comprehension–Extended	0.96	3.18
Auditory Processing	0.92	4.24	Reading Fluency	0.96	3.00
Long-Term Retrieval	0.97	2.60	Reading Rate	0.96	3.00
Visual Processing	0.86	5.61	Mathematics	0.96	3.00
Quantitative Reasoning	0.94	3.67	Broad Mathematics	0.97	2.60
Auditory Memory Span	0.90	4.74	Math Calculation Skills	0.97	2.60
Number Facility	0.90	4.74	Math Problem Solving	0.95	3.35
Perceptual Speed	0.93	3.97	Written Language	0.94	3.67
Cognitive Efficiency	0.93	3.97	Broad Written Language	0.95	3.35
Cognitive Efficiency–Extended	0.95	3.35	Basic Writing Skills	0.95	3.35
Reading Aptitude A	0.89	4.97	Written Expression	0.92	4.24
Reading Aptitude B	0.90	4.86	Academic Skills	0.97	2.60
Math Aptitude A	0.89	4.97	Academic Fluency	0.97	2.60
Math Aptitude B	0.89	4.97	Academic Applications	0.96	3.00
Writing Aptitude A	0.89	4.97	Academic Knowledge	0.95	3.35
Writing Aptitude B	0.90	4.86	Phoneme-Grapheme Knowledge	0.94	3.82
Oral Language	0.90	4.74	Brief Achievement	0.97	2.60
Broad Oral Language	0.92	4.24	Broad Achievement	0.99	1.50

Alternate Forms

The WJ IV ACH Standard Battery is available in three forms (Forms A, B, and C) to allow repeated testing of an individual over time or in different settings. The three forms of the WJ IV ACH Standard tests were carefully constructed to be as equivalent as possible with regard to content coverage, difficulty, and measurement precision. Two studies were conducted to assess the alternate-forms reliability of select WJ IV ACH Standard tests. Table 10 reports median alternate-forms reliability coefficients (i.e., *coefficients of stability*) from the studies.[2]

[2] The complete results of these two studies, as well as more detailed information regarding the equivalence of the WJ IV ACH Standard alternate forms with regard to content coverage, difficulty, and measurement precision, can be found in a separate WJ IV Assessment Service Bulletin—*Woodcock-Johnson IV Assessment Service Bulletin No. 1: WJ IV Tests of Achievement Alternate-Forms Equivalence* (LaForte & McGrew, 2014)—available at http://www.riverpub.com/products/wj-iv/pdf/WJ-IV_Tests-of-Achievement_Form.pdf

Table 10.
Median Alternate-Forms Reliability Coefficients for Select WJ IV ACH Standard Tests

Test	Form Comparison		
	A & B	A & C	B & C
Test 1: Letter-Word Identification	0.93[a]	0.93[a]	0.94[a]
Test 2: Applied Problems	0.91[a]	0.91[a]	0.92[a]
Test 3: Spelling	0.94[a]	0.93[a]	0.94[a]
Test 4: Passage Comprehension	0.91[a]	0.93[a]	0.92[a]
Test 5: Calculation	0.95[a]	0.94[a]	0.95[a]
Test 6: Writing Samples	0.87[a]	0.94[a]	0.92[a]
Test 9: Sentence Reading Fluency	0.85	0.87	0.88
Test 10: Math Facts Fluency	0.95	0.94	0.95
Test 11: Sentence Writing Fluency	0.84	0.89	0.79

[a] In cases where the ratio of the *W*-score standard deviation in the alternate-forms study sample to the *W*-score standard deviation in the norming sample was .80 or less, a correction for attenuation due to range restriction was applied to the correlation. The correction was originally developed by Bryant and Gokhale (1972) and Alexander (1990) and is described by Sackett and Yang (2000). It is particularly suited for cases where unrestricted variances for both correlated variables are known.

Evidence to Support the Use and Interpretation of WJ IV Scores

The WJ IV Technical Manual (McGrew et al., 2014) outlines several propositions for the use and interpretation of the WJ IV scores and provides support for each proposition within a framework consistent with that outlined in the *Standards for Educational and Psychological Testing* (AERA et al., 2014).

Representativeness of Content, Process, and Construct Coverage

The WJ IV includes tests measuring a complex set of unique abilities, defined by CHC theory, constituting cognitive ability, oral language ability, and academic achievement. Evidence to support this proposition, often termed *content validity evidence* or *substantive validity evidence*, for the WJ IV test scores is provided via the specification of test and cluster content according to contemporary CHC theory and research.[3] This aspect of the WJ IV validity argument builds upon the theories contained in the three prior editions of the battery: the WJ (Woodcock & Johnson, 1977), the WJ-R (Woodcock & Johnson, 1989), and the WJ III (Woodcock et al., 2001).

CHC Theory Content Coverage

The WJ IV test design blueprint pushes the design of tests "beyond CHC theory" (McGrew, 2012; Schneider & McGrew, 2012) as CHC theory was defined in the WJ III. In addition to its reliance on contemporary CHC theory as the basis for the overarching test battery design blueprint, the WJ IV plan was influenced by the incorporation of contemporary findings from neurocognitive, neuropsychological, and developmental research.

The distinction between broad and narrow abilities is an important concept in CHC theory. As in the WJ III, most of the WJ IV tests were designed to measure one narrow ability. This CHC-based test design approach, first operationalized in the WJ III, focuses on increasing CHC construct representation and decreasing construct-irrelevant variance in tests (Benson, 1998; McGrew & Flanagan, 1998; Messick, 1995). To increase

[3] Refer to Chapter 1 and Appendix A of the WJ IV Technical Manual (McGrew et al., 2014) for a description of contemporary CHC theory and the tests and clusters contained in the WJ IV.

breadth, clusters were constructed to subsume two or more qualitatively different narrow abilities. The principle of cluster interpretation was adopted to improve the content validity of measures for broad abilities such as reading, fluid reasoning, and general intelligence.

The seven broad CHC factors measured in the WJ IV COG include fluid reasoning (*Gf*), comprehension-knowledge (*Gc*), short-term working memory (*Gwm*), cognitive processing speed (*Gs*), auditory processing (*Ga*), long-term retrieval (*Glr*), and visual processing (*Gv*). Two-test narrow clusters are available for the CHC abilities of quantitative reasoning (RQ), auditory memory span (MS), number facility (N), perceptual speed (P), and lexical knowledge (VL). Cognitive efficiency, which represents the amalgam of processing speed (*Gs*) and short-term working memory (*Gwm*), is represented by two- and four-test clusters.

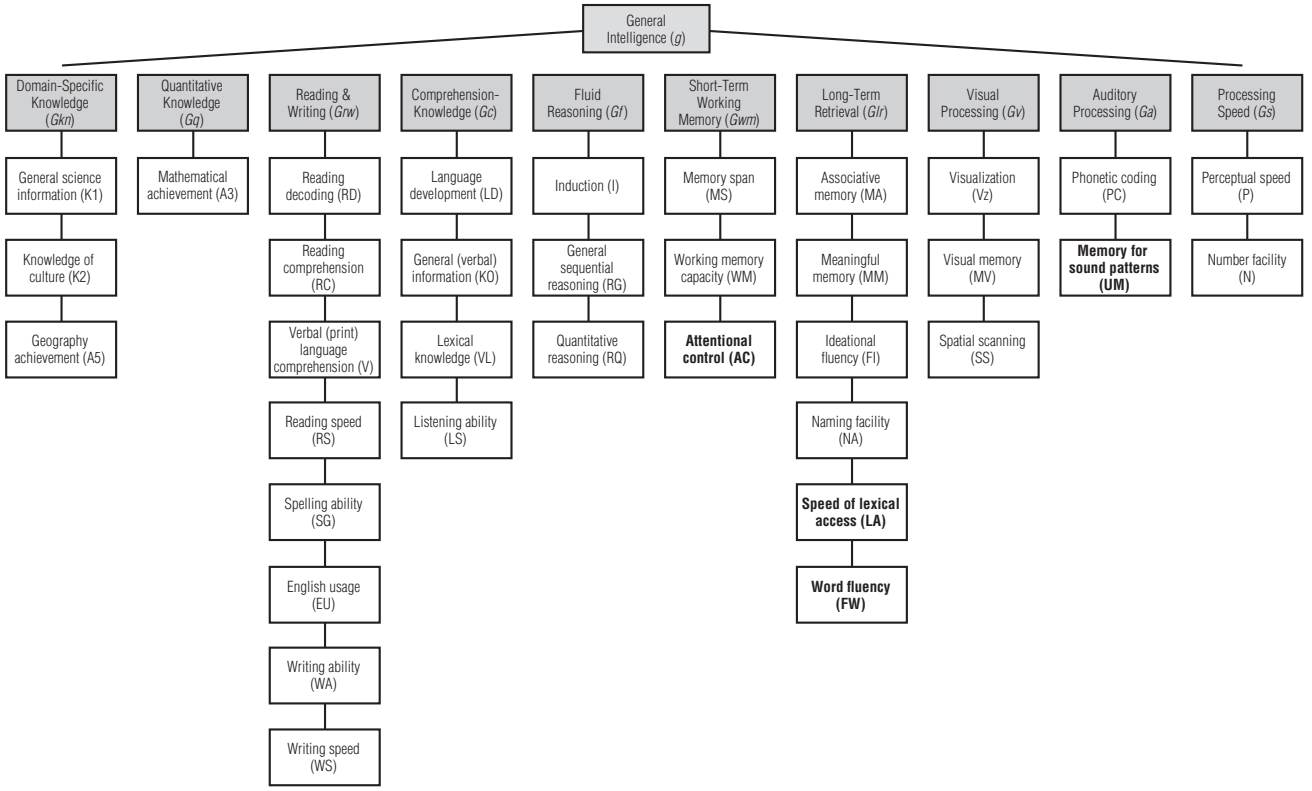
The WJ IV OL includes 12 tests that also measure abilities in the broad CHC domains of comprehension-knowledge (*Gc*), auditory processing (*Ga*), long-term retrieval (*Glr*), and short-term working memory (*Gwm*). The WJ IV OL tests were developed to measure the most important aspects of oral language ability, including oral expression, listening comprehension (LS), phonetic coding (PC), speed of lexical access (LA), vocabulary (VL/LD), and auditory memory span (MS).

The WJ IV ACH contains 20 tests that tap two other identified CHC cognitive abilities—quantitative knowledge (*Gq*) and reading and writing ability (*Grw*). The WJ IV ACH also includes additional measures of comprehension-knowledge (*Gc*), long-term retrieval (*Glr*), and auditory processing (*Ga*). The 20 WJ IV ACH tests were developed to measure the major aspects of academic achievement, including reading, mathematics, written language, and curricular knowledge. The specification of item content in these tests was based primarily on the goal of providing a broad sampling of achievement areas rather than an in-depth assessment of a relatively narrow area.

Test Content, Process, and Constructs

Figure 1 presents a mapping, created by the WJ IV authors, of the contemporary broad and narrow CHC abilities measured by the complete WJ IV. All broad CHC abilities are represented by at least one narrow CHC ability; most are represented by measures of two to seven narrow abilities. For the majority of WJ III tests retained in the WJ IV, independent support for the CHC content classifications has been provided by multiple iterations of cross-battery CHC expert consensus (Flanagan, Ortiz, & Alfonso, 2007, 2013; Flanagan, Ortiz, Alfonso, & Mascolo, 2006; McGrew, 1997; McGrew & Flanagan, 1998).

Figure 1.
Contemporary broad and narrow CHC ability content coverage by WJ IV COG, WJ IV OL, and WJ IV ACH.



Note. Bold font indicates narrow abilities that reflect modified definitions in CHC theory or proposed abilities based on results presented in the WJ IV Technical Manual (see Chapter 1 and Appendix A). The material in this figure is adapted from Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (3rd ed.) (p. 99–144). New York, NY: Guilford Press. Copyright Guilford Press. Reprinted with the permission of the Guilford Press.

Tables 11, 12, and 13 provide further descriptions of the broad and narrow constructs measured by the WJ IV COG, WJ IV OL, and WJ IV ACH tests, respectively, as well as stimulus and response characteristics, task requirements, and inferred cognitive processes.

Table 11.
WJ IV COG Test Content, Process, and Construct Descriptions

Cognitive Test	Primary Broad CHC Ability *Narrow Ability*	Stimuli	Task Requirements	Cognitive Processes	Response
1: Oral Vocabulary A: Synonyms B: Antonyms	Comprehension-Knowledge (*Gc*) *Lexical knowledge* (VL) *Language development* (LD)	Auditory (words)	Listening to a word and providing a synonym; listening to a word and providing an antonym	Semantic activation, access, and matching	Oral (words)
2: Number Series	Fluid Reasoning (*Gf*) *Quantitative reasoning* (RQ) *Induction* (I)	Visual (numeric)	Determining a numerical sequence	Representation and manipulation of points on a mental number line; identifying and applying an underlying rule/ principle to complete a numerical sequence	Oral (numbers)
3: Verbal Attention	Short-Term Working Memory (*Gwm*) *Working memory capacity* (WM) *Attentional control* (AC)	Auditory (words, numbers)	Listening to a series of numbers and animals intermingled and answering a specific question regarding the sequence	Controlled executive function; working memory capacity; recoding of acoustic, verbalized stimuli held in immediate awareness; selective auditory attention; attentional control	Oral (words)
4: Letter-Pattern Matching	Processing Speed (*Gs*) *Perceptual speed* (P)	Visual (letters)	Rapidly locating and circling identical letters or letter patterns	Speeded visual perception and matching; visual discrimination; orthographic processing; divided attention	Motoric (circling)
5: Phonological Processing A: Word Access B: Word Fluency C: Substitution	Auditory Processing (*Ga*) *Phonetic coding* (PC) *Word fluency* (*Glr*-FW) *Speed of lexical access* (*Glr*-LA)	Auditory (words)	Providing a word with a specific phonic element; naming as many words as possible that begin with a specified sound; substituting part of a word to make a new word	Semantic activation, access; speed of lexical access	Oral (words)
6: Story Recall	Long-Term Retrieval (*Glr*) *Meaningful memory* (MM) *Listening ability* (*Gc*-LS)	Auditory (text)	Listening to and recalling details of stories	Construction of propositional representations and recoding	Oral (passages)
7: Visualization A: Spatial Relations B: Block Rotation	Visual Processing (*Gv*) *Visualization* (Vz)	Visual (shapes, designs)	Identifying two-dimensional pieces that form a shape; identifying two three-dimensional rotated block patterns that match a target	Visual feature detection; manipulation (mental rotation) of visual images in space; matching	Oral (letters) or Motoric (pointing)
8: General Information A: Where B: What	Comprehension-Knowledge (*Gc*) *General (verbal) information* (K0)	Auditory (questions)	Identifying where an object is found and what people typically do with an object	Semantic activation and access to declarative generic knowledge	Oral (phrases, sentences)
9: Concept Formation	Fluid Reasoning (*Gf*) *Induction* (I)	Visual (drawings)	Identifying, categorizing, and determining rules	Rule-based categorization; rule switching; induction/ inference	Oral (words)
10: Numbers Reversed	Short-Term Working Memory (*Gwm*) *Working memory capacity* (WM) *Attentional control* (AC)	Auditory (numbers)	Listening to and recalling a sequence of digits in reversed order	Span of apprehension and recoding in working memory; working memory capacity, attentional capacity	Oral (numbers)
11: Number-Pattern Matching	Processing Speed (*Gs*) *Perceptual speed* (P)	Visual (numbers)	Rapidly locating and circling identical numerals from a defined set	Speeded visual perception and matching; visual discrimination; divided attention	Motoric (circling)

Table 11. *(cont.)*
WJ IV COG Test Content, Process, and Construct Descriptions

Cognitive Test	Primary Broad CHC Ability *Narrow Ability*	Stimuli	Task Requirements	Cognitive Processes	Response
12: Nonword Repetition	Auditory Processing (*Ga*) *Phonetic coding* (PC) *Memory for sound patterns* (UM) *Memory span* (*Gwm*-MS)	Auditory (nonsense words)	Listening to a nonsense word and repeating it exactly	Analysis of a sequence of acoustic phonological elements in immediate awareness; efficiency of the phonological loop	Oral (words)
13: Visual-Auditory Learning	Long-Term Retrieval (*Glr*) *Associative memory* (MA)	Visual (rebuses) Auditory (words)	Learning and recalling pictographic representations of words	Paired-associative encoding via directed spotlight attention; storage and retrieval	Oral (sentences)
14: Picture Recognition	Visual Processing (*Gv*) *Visual memory* (MV)	Visual (pictures)	Recognizing a subset of previously presented pictures within a field of similar distracting pictures	Formation of iconic memories and matching of visual stimuli to stored visual representations	Oral (words) or Motoric (pointing)
15: Analysis-Synthesis	Fluid Reasoning (*Gf*) *General sequential reasoning* (RG)	Visual (drawings)	Analyzing puzzles (using symbolic formulations) to determine missing components	Algorithmic reasoning; deduction	Oral (words)
16: Object-Number Sequencing	Short-Term Working Memory (*Gwm*) *Working memory capacity* (WM)	Auditory (words, numbers)	Listening to a series of numbers and words intermingled and recalling in two reordered sequences	Recoding of acoustic, verbalized stimuli held in immediate awareness; working memory capacity	Oral (words, numbers)
17: Pair Cancellation	Processing Speed (*Gs*) *Perceptual speed* (P) *Spatial scanning* (*Gv*-SS) *Attentional control* (*Gwm*-AC)	Visual (drawings)	Rapidly locating and marking a repeated pattern	Executive processing; attentional control; inhibition and interference control; sustained attention	Motoric (circling)
18: Memory for Words	Short-Term Working Memory (*Gwm*) *Memory span* (MS)	Auditory (words)	Listening to and repeating a sequence of unrelated words	Formation of echoic memories and verbalizable span of echoic store	Oral (words)

Table 12.
WJ IV OL Test Content, Process, and Construct Descriptions

Oral Language Test	Primary Broad CHC Ability *Narrow Ability*	Stimuli	Task Requirements	Cognitive Processes	Response
1: Picture Vocabulary (*10: Vocabulario sobre dibujos)*	Comprehension-Knowledge (*Gc*) *Lexical knowledge* (VL) *Language development* (LD)	Visual (pictures)	Identifying objects	Object recognition; lexical access and retrieval	Oral (words)
2: Oral Comprehension (*11: Comprensión oral*)	Comprehension-Knowledge (*Gc*) *Listening ability* (LS)	Auditory (text)	Listening to an oral passage and identifying a missing key word that makes sense	Construction of propositional representations through syntactic and semantic integration of orally presented passages in real time	Oral (words)
3: Segmentation	Auditory Processing (*Ga*) *Phonetic coding* (PC)	Auditory (words)	Listening to a word and breaking it into syllables or phonemes	Analysis of acoustic, phonological elements in immediate awareness	Oral (word parts, phonemes)
4: Rapid Picture Naming	Long-Term Retrieval (*Glr*) *Naming facility* (NA) *Speed of lexical access* (LA)	Visual (pictures)	Recognizing objects, then retrieving and articulating their names rapidly	Speed/fluency of retrieval and oral production of recognized objects; speeded serial naming; rapid object recognition	Oral (words)
5: Sentence Repetition	Short-Term Working Memory (*Gwm*) *Memory span* (MS) Comprehension-Knowledge (*Gc*) *Listening ability* (LS)	Auditory (words, sentences)	Listening to and repeating words, phrases, or sentences in the correct sequence	Formation of echoic memories aided by a semantic, meaning-based code	Oral (words, sentences)
6: Understanding Directions (*12: Comprensión de indicaciones*)	Short-Term Working Memory (*Gwm*) *Working memory capacity* (WM) Comprehension-Knowledge (*Gc*) *Listening ability* (LS)	Visual (pictures) Auditory (text)	Studying a picture, then listening to a sequence of instructions and following the directions by pointing to items in the picture	Construction of a mental structure in immediate awareness and modification of the structure via mapping	Motoric (pointing)
7: Sound Blending	Auditory Processing (*Ga*) *Phonetic coding* (PC)	Auditory (phonemes)	Synthesizing language sounds (phonemes) to say a word	Synthesis of acoustic, phonological elements in immediate awareness; matching the sequence of elements to stored lexical entries; lexical activation and access	Oral (words)
8: Retrieval Fluency	Long-Term Retrieval (*Glr*) *Speed of lexical access* (LA) *Ideational fluency* (FI)	Auditory (directions only)	Naming as many examples as possible in a given category within 1 minute	Recognition, fluent retrieval, and oral production of examples of a semantic category; activation of semantic network; speeded name generation	Oral (words)
9: Sound Awareness A: Rhyming B: Deletion	Auditory Processing (*Ga*) *Phonetic coding* (PC)	Auditory (questions, words)	Providing a rhyming word; removing parts of words to make a new word	Access, retrieval, and application of the rules of English phonology	Oral (words)

Note. Test name in italics signifies the Spanish version of the English test.

Table 13.
WJ IV ACH Test Content, Process, and Construct Descriptions

Achievement Test	Primary Broad CHC Ability *Narrow Ability*	Stimuli	Task Requirements	Cognitive Processes	Response
1: Letter-Word Identification	Reading & Writing Ability (*Grw*) *Reading decoding* (RD)	Visual (text)	Identifying printed letters and words	Feature detection and analysis (for letters) and recognition of visual word forms from a phonological lexicon; access of pronunciations associated with visual word forms	Oral (letter names, words)
2: Applied Problems	Quantitative Knowledge (*Gq*) *Mathematical achievement* (A3) Fluid Reasoning (*Gf*) *Quantitative reasoning* (RQ)	Auditory (questions) Visual (numeric, text)	Performing math calculations in response to orally presented problems	Construction of mental models via language comprehension, application of calculation and/or quantitative reasoning; formation of insight	Oral (numbers, words)
3: Spelling	Reading & Writing Ability (*Grw*) *Spelling ability* (SG)	Auditory (words)	Spelling orally presented words	Access to and application of knowledge of orthography of word forms by mapping whole-word phonology onto whole-word orthography, by translating phonological segments into graphemic units, or by activating spellings of words from the semantic lexicon	Motoric (writing)
4: Passage Comprehension	Reading & Writing Ability (*Grw*) *Reading comprehension* (RC)	Visual (text)	Identifying a missing key word that makes sense in the context of a written passage	Construction of propositional representations; integration of syntactic and semantic properties of printed words and sentences into a representation of the whole passage	Oral (words)
5: Calculation	Quantitative Knowledge (*Gq*) *Mathematical achievement* (A3)	Visual (numeric)	Performing various mathematical calculations	Access to and application of knowledge of numbers and calculation procedures; verbal associations between numbers represented as strings of words	Motoric (writing)
6: Writing Samples	Reading & Writing Ability (*Grw*) *Writing ability* (WA)	Auditory (text) Visual (text)	Writing meaningful sentences for a given purpose	Retrieval of word meanings; application of psycholinguistic rules of case, grammar, and syntax; planning and construction of bridging inferences in immediate awareness (auditory and/or visual buffer)	Motoric (writing)
7: Word Attack	Reading & Writing Ability (*Grw*) *Reading decoding* (RD) Auditory Processing (*Ga*) *Phonetic coding* (PC)	Visual (word)	Reading phonically regular nonwords	Grapheme-to-phoneme translation and accessing pronunciations of pseudowords not contained in the mental lexicon	Oral (words)

Table 13. *(cont.)*
WJ IV ACH Test Content, Process, and Construct Descriptions

Achievement Test	Primary Broad CHC Ability *Narrow Ability*	Stimuli	Task Requirements	Cognitive Processes	Response
8: Oral Reading	Reading & Writing Ability (*Grw*) *Reading comprehension* (RC) *Verbal (print) language comprehension* (V)	Visual (text)	Reading sentences orally with accuracy and fluency	Integration of orthographic, phonological, and semantic processes; articulatory planning and motor execution	Oral (sentences)
9: Sentence Reading Fluency	Reading & Writing Ability (*Grw*) *Reading comprehension* (RC) *Reading speed* (RS) Processing Speed (*Gs*)	Visual (text)	Reading printed statements rapidly and responding true or false (yes or no)	Speeded semantic decision making requiring reading ability and generic knowledge	Motoric (circling)
10: Math Facts Fluency	Quantitative Knowledge (*Gq*) *Mathematical achievement* (A3) Processing Speed (*Gs*) *Number facility* (N)	Visual (numeric)	Adding, subtracting, and multiplying rapidly	Speeded access to and application of digit-symbol arithmetic procedures	Motoric (writing)
11: Sentence Writing Fluency	Reading & Writing Ability (*Grw*) *Writing ability* (WA) *Writing speed* (WS) Processing Speed (*Gs*)	Visual (words with pictures)	Formulating and writing simple sentences rapidly	Speeded formation of constituent sentence structures requiring fluent access to semantic and syntactic knowledge	Motoric (writing)
12: Reading Recall	Reading & Writing Ability (*Grw*) *Reading comprehension* (RC) Long-Term Retrieval (*Glr*) *Meaningful memory* (MM)	Visual (text)	Reading and recalling details of stories	Construction of propositional representations and recoding	Oral (passages)
13: Number Matrices	Fluid Reasoning (*Gf*) *Quantitative reasoning* (RQ)	Visual (numeric)	Determining a two-dimensional numerical pattern	Access to verbal-visual numeric codes; transcoding verbal and/or visual representations of numeric information into analogical representations; determining the relationship between/among numbers on the first part of the structure and mapping (projecting) the structure to complete the analogy	Oral (numbers)
14: Editing	Reading & Writing Ability (*Grw*) *English usage* (EU)	Visual (text)	Identifying and correcting errors in written passages	Access and application of lexical and syntactic information about details of word forms and writing conventions	Oral (sentences)
15: Word Reading Fluency	Reading & Writing Ability (*Grw*) *Reading comprehension* (RC) *Reading speed* (RS) Processing Speed (*Gs*)	Visual (words)	Rapidly reading words and marking the two in each row that are semantically related	Speeded semantic decision making requiring reading ability	Motoric (slash marks)

Table 13. *(cont.)*
WJ IV ACH Test Content, Process, and Construct Descriptions

Achievement Test	Primary Broad CHC Ability *Narrow Ability*	Stimuli	Task Requirements	Cognitive Processes	Response
16: Spelling of Sounds	Reading & Writing Ability (*Grw*) *Spelling ability* (SG) Auditory Processing (*Ga*) *Phonetic coding* (PC)	Auditory (letters, words)	Spelling letter patterns that are regular patterns in written English	Translating spoken elements of nonwords into graphemic units; phonologically mediated mapping of orthography	Motoric (writing)
17: Reading Vocabulary	Reading & Writing Ability (*Grw*) *Reading comprehension* (RC) Comprehension-Knowledge (*Gc*) *Lexical knowledge* (VL)	Visual (words)	Reading words and providing an appropriate synonym or antonym	Recognition of visual word forms; semantic access and activation; semantic matching	Oral (words)
18: Science	Domain-Specific Knowledge (*Gkn*) *General science information* (K1) Comprehension-Knowledge (*Gc*) *General (verbal) information* (K0)	Auditory (questions) Visual (text, pictures)	Responding to questions about science	Implicit, declarative category-specific memory	Oral (words, sentences)
19: Social Studies	Domain-Specific Knowledge (*Gkn*) *Knowledge of culture* (K2) *Geography achievement* (A5) Comprehension-Knowledge (*Gc*) *General (verbal) information* (K0)	Auditory (questions) Visual (text, pictures)	Responding to questions about social studies	Implicit, declarative category-specific memory	Oral (words, sentences)
20: Humanities	Domain-Specific Knowledge (*Gkn*) *Knowledge of culture* (K2) Comprehension-Knowledge (*Gc*) *General (verbal) information* (K0)	Auditory (questions) Visual (text, pictures)	Responding to questions about humanities	Implicit, declarative category-specific memory	Oral (words, sentences)

Developmental Patterns of WJ IV Ability Clusters

The WJ IV tests and clusters display average score changes consistent with the developmental growth and decline of cognitive abilities and achievement across the life span. Divergent growth curves provide evidence for the existence of distinct, unique abilities (Carroll, 1993). Figures 2 through 5 present examples of growth curves, or "difference curves," from ages 6 to 90 for several WJ IV COG, WJ IV OL, and WJ IV ACH clusters. The difference curves illustrate that the unique abilities measured by the WJ IV follow different developmental courses or trajectories over the age span from childhood to geriatric levels. The examples were constructed using age 6 years, 0 months (6-0) as a starting point and subtracting the norm-based REF *W* score for age 6-0 for each cluster from all other REF *W*s for that cluster through age 90. This procedure produced difference curves all starting with an assigned common origin of zero.

Figure 2.
Plot of WJ IV COG GIA, seven CHC factor clusters, and the Gf-Gc *Composite* W*-score difference curves by age.*

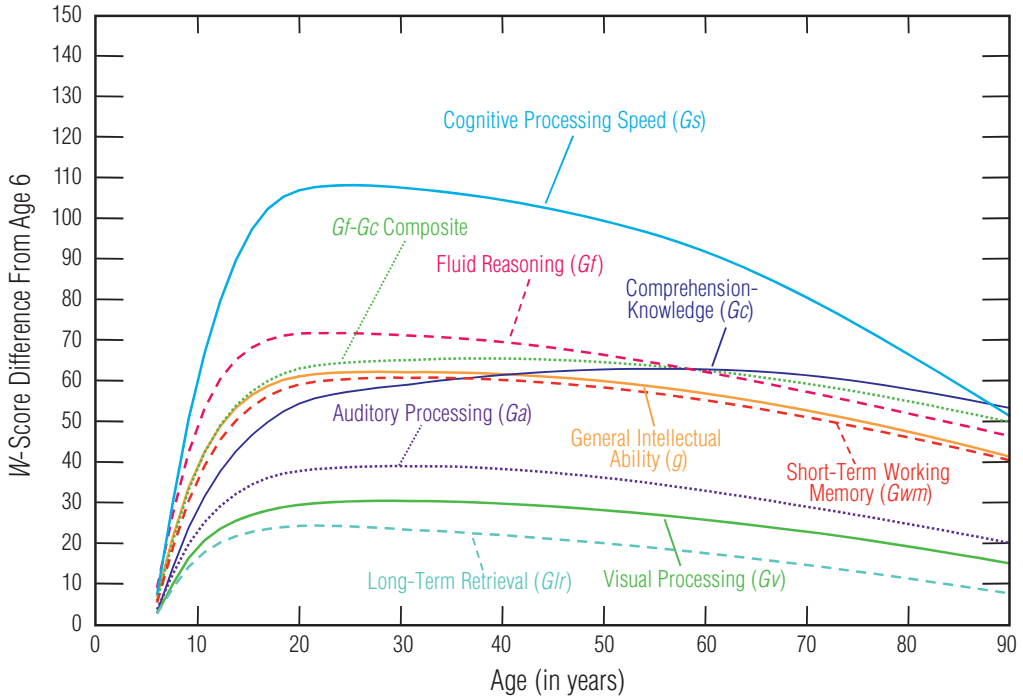


Figure 3.
Plot of WJ IV COG GIA, three narrow cognitive ability, and three other clinical cluster W*-score difference curves by age.*

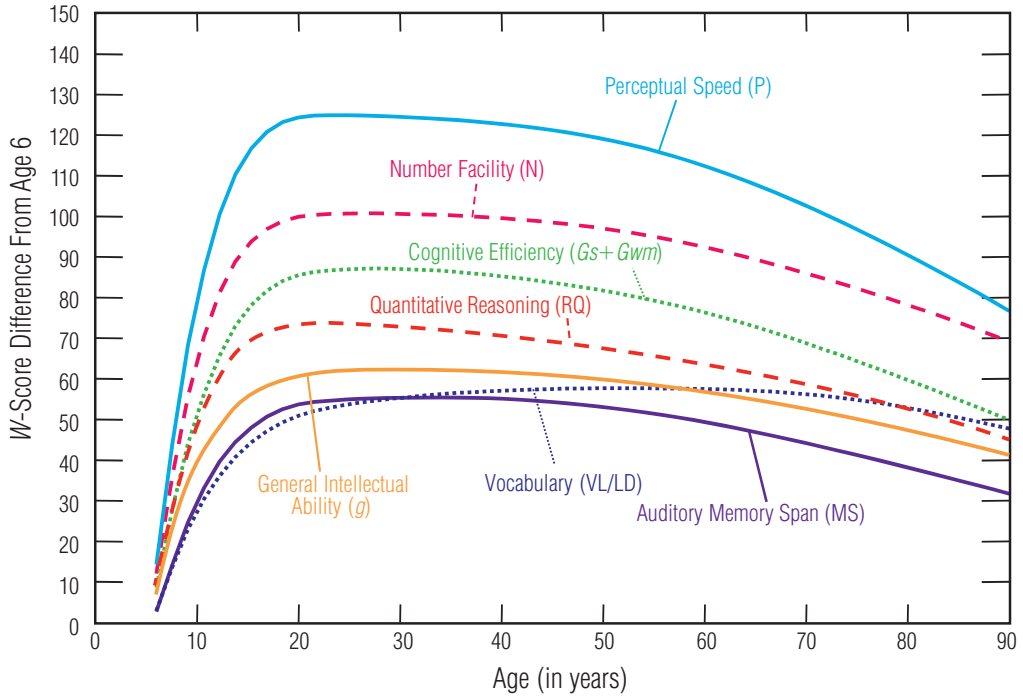


Figure 4.
Plot of five WJ IV OL cluster W-score difference curves by age.

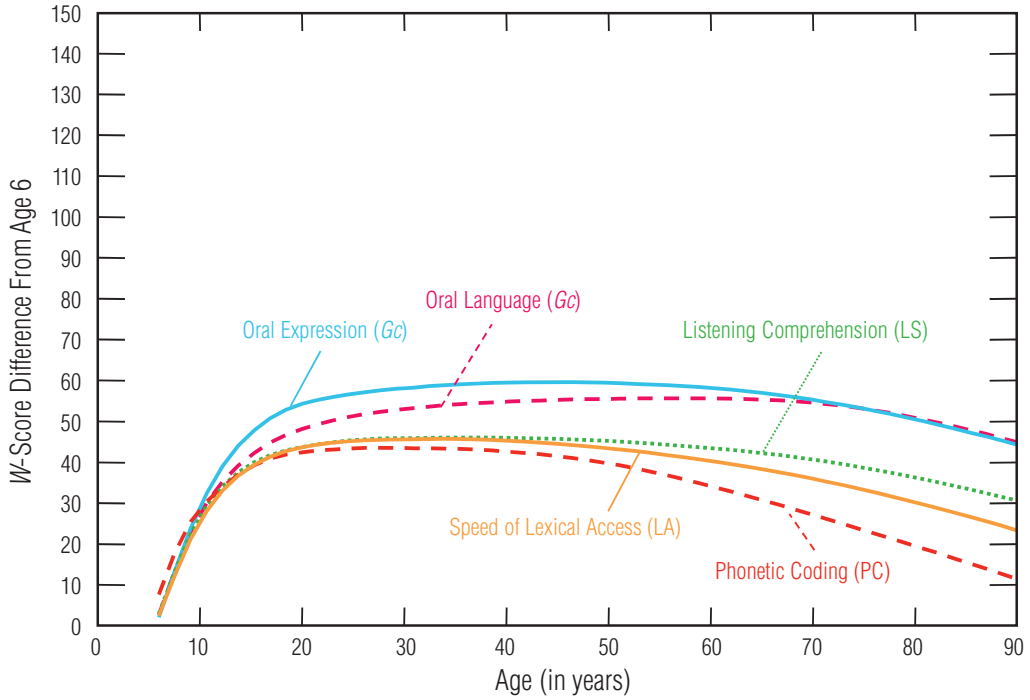


Figure 5.
Plot of 13 WJ IV ACH cluster W-score difference curves by age.

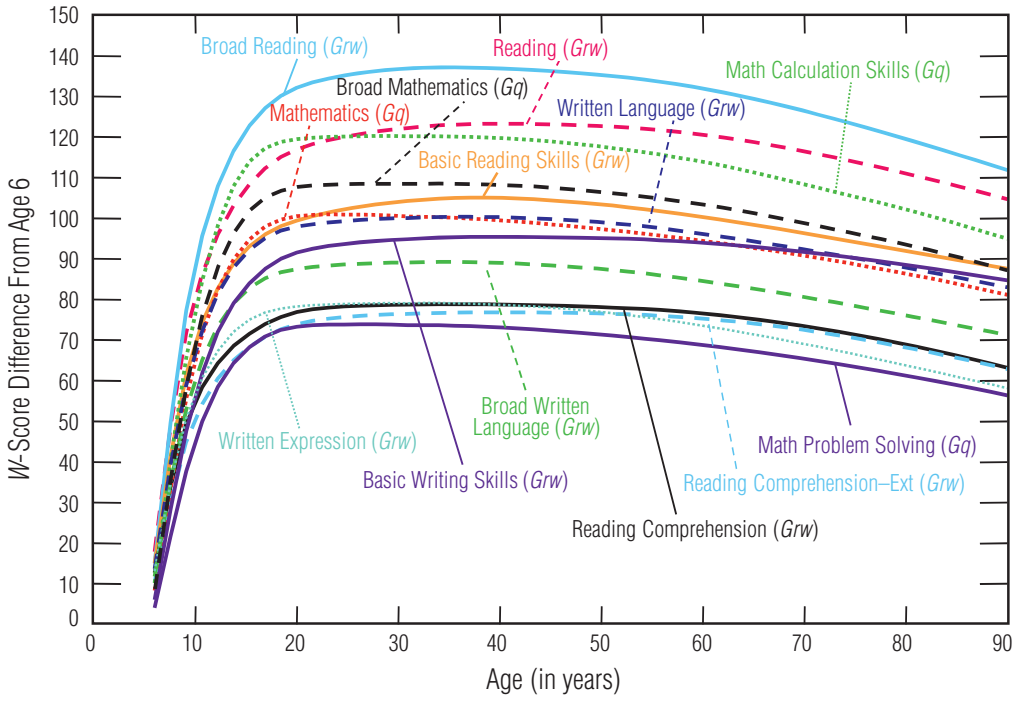


Figure 2 presents difference curves for the GIA (g), seven CHC cognitive factor clusters, and the *Gf-Gc* Composite. The patterns of growth and decline of the seven WJ IV CHC cognitive factor clusters differ markedly, providing evidence to support the existence of distinct abilities. Figure 3 demonstrates the existence of distinct patterns of growth and decline among the abilities measured by the WJ IV COG narrow ability factors and clinical clusters of Auditory Memory Span (MS), Vocabulary (VL/LD),

Quantitative Reasoning (RQ), Cognitive Efficiency (*Gs*+*Gwm*), Number Facility (N), and Perceptual Speed (P). The GIA (*g*) curve is included for comparison purposes.

Five WJ IV OL cluster growth curves are presented in Figure 4. Given that the WJ IV OL clusters contain tests from the broad CHC domains of *Gc*, *Ga*, and *Glr*, the similarity of the WJ IV OL cluster difference curves to the corresponding CHC domain difference curves in Figure 2 supports the validity of these WJ IV OL cluster scores for measuring key language-related cognitive abilities.

Growth curves for the 13 broad and narrow WJ IV ACH clusters are presented in Figure 5. A number of points regarding achievement cluster measures are apparent in Figure 5. First, a majority of the WJ IV ACH clusters show rapid acceleration of growth from age 6 through approximately 15 years. Second, the majority of achievement levels peak at a much higher point relative to their origin (in this case, 6 years) than the cognitive abilities do. Third, most achievement levels do not demonstrate as much absolute decline across the age span as the cognitive abilities do; the achievement skills are generally maintained at higher levels into the older age ranges. These three features distinguish the achievement cluster curves from most cognitive growth curves.

The existence of unique developmental patterns for most of the WJ IV broad and narrow abilities, across and within CHC domains, is one form of evidence that, combined with information about the test's content, structure, and relationship to other variables, supports the validity of the WJ IV scores for measuring an individual's cognitive abilities, oral language abilities, and academic achievement.

Internal Structure of the WJ IV

The primary source of validity evidence relevant to the internal structure of educational and psychological tests is the extent to which relationships among test scores conform to the relationships implied by the underlying theoretical construct (AERA et al., 2014). Two forms of internal structure evidence are presented for the WJ IV. First, the pattern of intercorrelations among the WJ IV test and cluster scores is described. Next, exploratory and confirmatory multivariate statistical methods are used to analyze the relations between the WJ IV tests.

Test and Cluster Intercorrelations

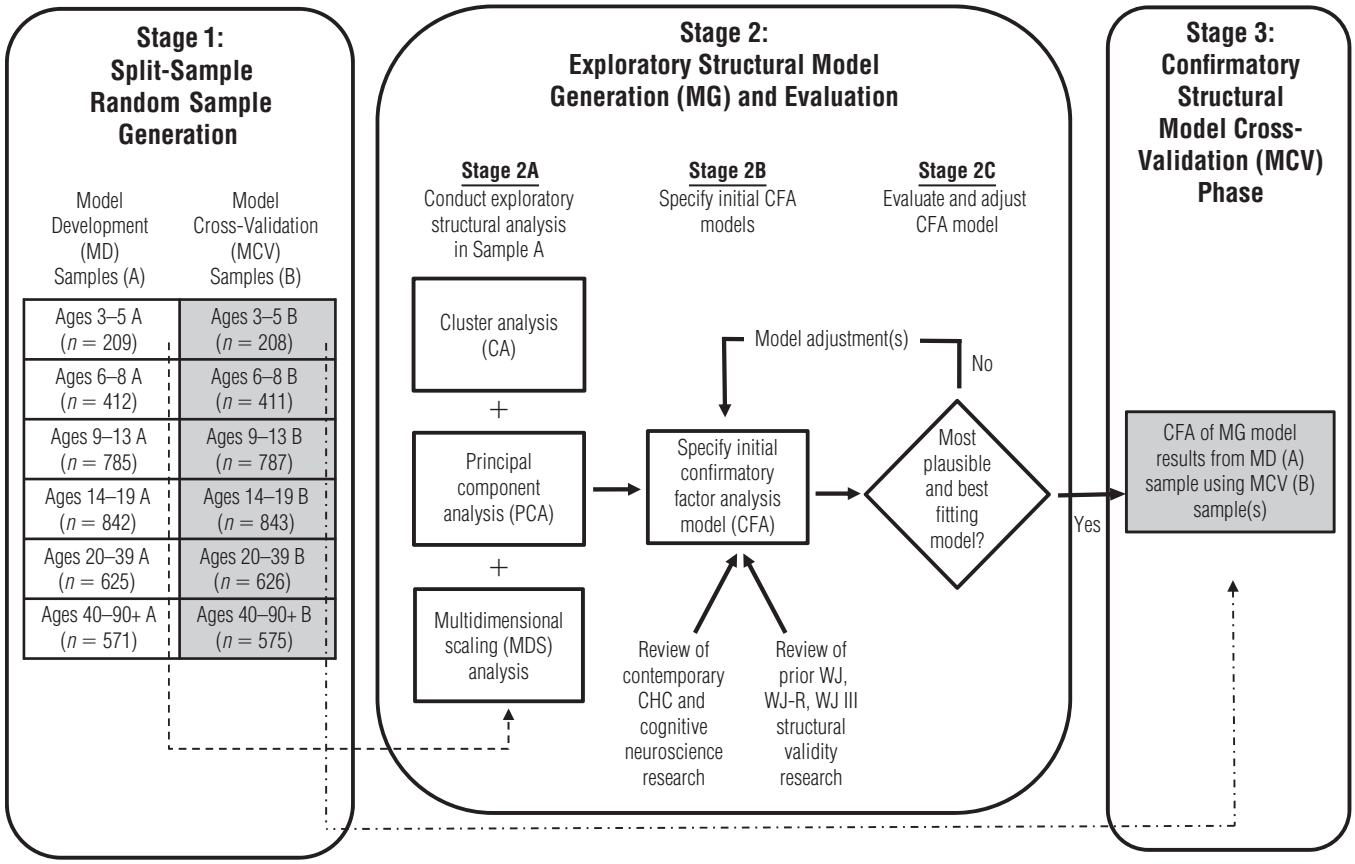
The direction and magnitude of correlations among test and cluster scores can provide evidence that the scores conform to theoretical expectations about the underlying constructs (AERA et al., 2014; Campbell & Fiske, 1959). The test and cluster intercorrelations for the WJ IV provide empirical support for several inferences about the relations between the WJ IV scores.[4] First, correlations are generally higher among related CHC domain tests or clusters than among unrelated tests or clusters. Second, the range of broad CHC cognitive cluster intercorrelations (typically .30 to .60) is lower than those reported among the primary achievement clusters, providing evidence that the WJ IV COG clusters measure distinct cognitive abilities. Third, within the achievement clusters, correlations are consistently higher between clusters from the same achievement domain and lower between clusters from different domains.

[4] Complete correlation matrices for all tests and clusters are reported in Appendices E and F, respectively, of the WJ IV Technical Manual (McGrew et al., 2014) for six broad age group samples.

Three-Stage Structural Validity Analysis

The nature and number of test changes in the WJ IV made it likely that a strict confirmatory CHC-based analysis of the WJ IV—similar to that performed on the WJ III—might fail to identify new structural dimensions and relations. Instead, a systematic exploratory, model generation, and cross-validation structural validity strategy was applied to the WJ IV norming data. The WJ IV split-sample, multiple-stage, exploratory-confirmatory approach is the most thorough scientific approach to the examination of the structural validity of any contemporary battery of cognitive, oral language, and achievement tests. This three-stage process is portrayed in Figure 6. A summary of the process is provided here; however, readers are encouraged to consult Chapter 5 of the WJ IV Technical Manual (McGrew et al., 2014) for further details.

Figure 6.
Three-stage structural validity procedures for the WJ IV.



As illustrated in Figure 6, the WJ IV norming sample was divided into six age-differentiated groups. Each sample was randomly split into model development (MD; sample A) and model cross-validation (MCV; sample B) samples of approximately equal size (see Stage 1 in Figure 6). Each of the six MD samples was analyzed with three different exploratory multivariate methods—cluster analysis (CA), principal components analysis (PCA), and multidimensional scaling (MDS) analysis (see Stage 2A in Figure 6). The use of three methodological lenses allows for the detailed exploration of the relations among the complete collection of WJ IV tests. These analyses were first conducted using

the age 9 through 13 sample as an exemplar. The results from this age group were then used as an approximate starting model (Stage 2B) for all other age group samples. The next step was the specification of the initial model-generating (MG) confirmatory factor analysis (CFA) models based on the integration of the CA, PCA, and MDS results from Stage 2A. A comprehensive review of contemporary CHC and neuroscience research as well as structural validity research on all three prior editions of the Woodcock-Johnson tests was integrated with the exploratory results from Stage 2A to specify the initial WJ IV MG CFA models (see Stage 2B in Figure 6).

After evaluating these models (see Stage 2C in Figure 6), two were found to be most plausible. The broad CHC factor top-down model was specified to best represent the broad CHC constructs outlined in contemporary CHC theory. The broad + narrow CHC factor bottom-up model focused on specifying and evaluating plausible narrow and broad CHC factors. In both models, all model parameters for the exemplar age group MD sample were positive, significant ($p < .05$), and meaningful. In Stage 3, the two models were taken "as is" and cross-validated with the exemplar age group MCV sample (see Figure 6). The WJ IV CFA models were evaluated for overall statistical model fit and for size, statistical significance, and interpretability of all model parameter estimates (Brown, 2006). The broad CHC factor top-down model is the preferred model per the parsimony principle (also known as Occam's razor), which states that "given two models with similar fit to the data, the simpler model is preferred" (Kline, 2011, p. 102).

Table 14 presents all test and factor loading parameters from the CFA maximum-likelihood (ML) results for the broad CHC factor top-down model for the MD and MCV samples across all age groups. The large number of significant model parameters in the MD and MCV samples supports the generalizability and stability of the structural validity of this model for the WJ IV. The high loadings of the broad CHC ability factors (median loadings from .79 to .95, except .59 on *Gs*) on the general intelligence factor (*g*) indicate that the respective WJ IV clusters represent broad abilities that are influenced to a significant degree by general intelligence. The pattern of CHC *g*-factor loadings is generally consistent with the extant research (Carroll, 1993). Finally, the highest median *g* loading for the *Glr* factor (.95) is noteworthy. Research (Reynolds, Keith, Flanagan, & Alfonso, 2013) showed that a factor defined by associative memory tests was the second-highest-loading factor on *g*.

Table 14.

WJ IV Broad CHC 9-Factor (Top-Down; Model 2) CFA ML Model Results in Six Age Groups

Latent Factors/ Tests	Age Group												Median
	3–5		6–8		9–13		14–19		20–39		40–90+		
	MD (*n* = 209)	MCV (*n* = 209)	MD (*n* = 412)	MCV (*n* = 411)	MD (*n* = 785)	MCV (*n* = 787)	MD (*n* = 842)	MCV (*n* = 843)	MD (*n* = 625)	MCV (*n* = 626)	MD (*n* = 571)	MCV (*n* = 574)	
g													
Glr	*1.00*	0.98	0.99	0.99	0.92	0.95	0.89	0.95	0.95	0.93	*1.00*	0.97	0.95
Gf	0.87	0.80	0.94	0.94	0.91	0.95	0.94	0.93	0.93	0.94	0.94	0.93	0.94
Gq	0.81	0.87	0.84	0.87	0.88	0.91	0.91	0.89	0.87	0.88	0.89	0.90	0.89
Gc	0.79	0.83	0.76	0.73	0.79	0.78	0.84	0.84	0.87	0.84	0.88	0.87	0.84
Gv	0.85	0.81	0.81	0.81	0.72	0.74	0.81	0.86	0.84	0.83	0.87	0.85	0.82
Ga	0.88	0.88	0.81	0.79	0.56	0.63	0.72	0.76	0.81	0.81	0.87	0.88	0.80
Grw	0.80	0.83	0.75	0.78	0.78	0.81	0.80	0.79	0.80	0.78	0.81	0.82	0.80
Gwm	0.73	0.80	0.76	0.79	0.76	0.75	0.78	0.77	0.81	0.81	0.85	0.88	0.79
Gs	0.75	0.83	0.54	0.63	0.52	0.63	0.62	0.56	0.42	0.58	0.59	0.60	0.59
Gc													
ORLVOC	0.77	0.81	0.83	0.85	0.86	0.84	0.87	0.88	0.88	0.89	0.88	0.88	0.87
SOC	0.74	0.79	0.76	0.82	0.80	0.80	0.84	0.84	0.85	0.83	0.86	0.83	0.83
PICVOC	0.78	0.77	0.77	0.78	0.82	0.81	0.83	0.84	0.85	0.83	0.85	0.79	0.82
HUM	0.76	0.76	0.75	0.74	0.78	0.76	0.79	0.82	0.84	0.84	0.85	0.86	0.81
ORLCMP	0.77	0.73	0.77	0.72	0.76	0.74	0.79	0.79	0.81	0.80	0.83	0.83	0.79
GENINF	0.67	0.61	0.65	0.69	0.77	0.74	0.78	0.77	0.81	0.82	0.82	0.81	0.78
SCI	0.72	0.75	0.75	0.72	0.74	0.72	0.78	0.79	0.80	0.76	0.82	0.80	0.77
VRBANL	0.51	0.35	0.73	0.70	0.55	0.52	0.51	0.57	0.63	0.59	0.59	0.60	0.59
RDGVOC			0.49	0.51	0.51	0.50	0.56	0.52	0.52	0.56	0.56	0.53	0.52
PSGCMP			0.21	0.20	0.30	0.27	0.31	0.32	0.42	0.41	0.38	0.35	0.31
PHNPRO	—	—	0.28	0.31	0.30	0.26	0.36	0.28	0.20	0.25	—	—	0.27
SENREP	0.23	0.25	0.45	0.28	0.24	0.19	0.28	0.14	0.25	0.33	0.29	0.19	0.26
STYREC			0.19	0.22	—	—	—	—	—	—	—	—	—
UNDDIR	0.46	0.34	0.34	0.16	—	—	—	—	—	—	—	—	—
Gf													
NUMSER			0.76	0.81	0.77	0.82	0.80	0.80	0.76	0.72	0.82	0.78	0.79
NUMMAT			0.66	0.72	0.73	0.72	0.75	0.75	0.77	0.77	0.78	0.80	0.75
CONFRM	0.63	0.60	0.65	0.67	0.61	0.64	0.68	0.69	0.74	0.72	0.71	0.70	0.69
ANLSYN			0.60	0.62	0.58	0.63	0.61	0.63	0.67	0.67	0.69	0.70	0.63
VRBANL	0.32	0.44	—	—	0.28	0.29	0.36	0.29	0.26	0.31	0.31	0.30	0.30
Gwm													
VRBATN	0.73	0.74	0.77	0.73	0.77	0.74	0.77	0.76	0.77	0.74	0.81	0.79	0.77
OBJNUM			0.77	0.74	0.68	0.70	0.74	0.74	0.79	0.78	0.81	0.77	0.75
MEMWRD	0.67	0.68	0.69	0.63	0.54	0.46	0.59	0.58	0.70	0.74	0.72	0.73	0.66
UNDDIR	0.29	0.37	0.41	0.56	0.64	0.64	0.66	0.65	0.69	0.65	0.68	0.69	0.65
NWDREP	0.59	0.67	0.51	0.50	0.50	0.62	0.43	0.58	0.60	0.62	0.64	0.66	0.59
SENREP	0.44	0.42	0.27	0.44	0.51	0.51	0.45	0.58	0.47	0.43	0.51	0.58	0.49
NUMREV	0.44	0.35	0.49	0.47	0.36	0.29	0.36	0.36	0.50	0.48	0.69	0.67	0.48
LETPAT			—	—	—	—	—	—	0.20	0.12	0.23	0.13	—

Table 14. *(cont.)*
WJ IV Broad CHC 9-Factor (Top-Down; Model 2) CFA ML Model Results in Six Age Groups

Latent Factors/ Tests	Age Group 3–5 MD (n = 209)	3–5 MCV (n = 209)	6–8 MD (n = 412)	6–8 MCV (n = 411)	9–13 MD (n = 785)	9–13 MCV (n = 787)	14–19 MD (n = 842)	14–19 MCV (n = 843)	20–39 MD (n = 625)	20–39 MCV (n = 626)	40–90+ MD (n = 571)	40–90+ MCV (n = 574)	Median
Gs													
NUMPAT	0.70	0.65	0.83	0.85	0.78	0.80	0.79	0.73	0.71	0.79	0.81	0.78	0.79
LETPAT			0.78	0.78	0.73	0.80	0.75	0.76	0.68	0.71	0.66	0.70	0.74
WRDFLU			0.69	0.68	0.69	0.70	0.72	0.64	0.72	0.74	0.71	0.74	0.70
PAIRCN			0.63	0.65	0.58	0.61	0.57	0.60	0.44	0.54	0.49	0.59	0.58
MTHFLU			0.53	0.60	0.52	0.54	0.50	0.48	0.52	0.42	0.51	0.54	0.52
SNRDFL			0.45	0.44	0.54	0.49	0.51	0.50	0.51	0.55	0.48	0.46	0.50
SNWRFL			0.43	0.47	0.49	0.50	0.46	0.49	0.51	0.51	0.50	0.41	0.49
RPCNAM	0.68	0.57	—	—	0.30	0.33	0.36	0.30	0.31	0.40	0.29	0.28	0.31
RETFLU	0.42	0.69	—	—	0.17	0.20	0.28	0.22	0.18	0.30	0.25	0.31	0.23
CALC			0.28	0.36	0.21	0.23	0.18	0.16	0.17	0.13	0.13	0.16	0.18
Ga													
SEGMNT	0.68	0.68	0.75	0.80	0.73	0.74	0.74	0.73	0.78	0.74	0.73	0.79	0.74
SNDBLN	0.62	0.68	0.63	0.54	0.62	0.56	0.65	0.66	0.69	0.70	0.70	0.73	0.65
PHNPRO	0.79	0.81	0.58	0.62	0.60	0.63	0.56	0.58	0.68	0.64	0.88	0.88	0.62
SNDAWR	0.77	0.85	0.60	0.50	0.51	0.49	0.53	0.56	0.65	0.62	0.66	0.74	0.58
SPLSND			0.47	0.47	0.48	0.48	0.56	0.56	0.53	0.48	0.54	0.60	0.51
WRDATK			0.24	0.27	0.23	0.22	0.27	0.29	0.25	0.22	0.24	0.23	0.24
NWDREP			0.18	0.20	0.19	0.07	0.22	0.11	—	—	—	—	0.18
MEMWRD	—	—	—	—	0.25	0.29	0.14	0.16	—	—	—	—	—
Glr													
STYREC	0.50	0.60	0.45	0.35	0.51	0.58	0.57	0.62	0.56	0.55	0.61	0.62	0.57
VAL	0.58	0.54	0.51	0.47	0.42	0.44	0.51	0.48	0.59	0.60	0.54	0.61	0.51
MEMNAM	0.33	0.27	0.38	0.29	0.41	0.38	0.38	0.42	0.47	0.46	0.48	0.53	0.41
RETFLU	—	—	0.48	0.50	0.44	0.37	0.35	0.40	0.44	0.33	0.42	0.37	0.41
RPCNAM	—	—	0.45	0.40	0.27	0.22	0.16	0.20	0.25	0.16	0.30	0.27	0.26
WRTSMP			0.16	0.22	0.23	0.18	0.28	0.31	0.23	0.13	0.23	0.18	0.22
RDGREC			0.19	0.24	0.23	0.27	0.17	0.16	—	—	—	—	0.21
Gv													
VISUAL	0.58	0.54	0.77	0.70	0.63	0.74	0.65	0.70	0.75	0.75	0.74	0.73	0.74
VISCLO	0.58	0.69	0.55	0.48	0.59	0.55	0.61	0.59	0.62	0.61	0.65	0.60	0.60
PICREC	0.65	0.62	0.64	0.57	0.49	0.52	0.34	0.40	0.47	0.45	0.50	0.49	0.49
PAIRCN			—	—	0.20	0.16	0.24	0.18	0.35	0.31	0.28	0.14	0.22
Grw													
LWIDNT	0.88	0.90	0.93	0.93	0.88	0.91	0.90	0.90	0.92	0.92	0.92	0.93	0.92
SPELL	0.80	0.85	0.87	0.86	0.87	0.88	0.88	0.87	0.86	0.86	0.89	0.90	0.87
EDIT			0.79	0.81	0.81	0.84	0.81	0.83	0.81	0.82	0.85	0.85	0.81
ORLRDG			0.82	0.85	0.76	0.79	0.79	0.78	0.80	0.84	0.85	0.85	0.81
WRDATK			0.64	0.62	0.61	0.65	0.59	0.57	0.60	0.62	0.62	0.64	0.62
PSGCMP	0.75	0.77	0.70	0.69	0.54	0.61	0.59	0.56	0.47	0.50	0.54	0.56	0.56
RDGREC			0.60	0.51	0.42	0.45	0.45	0.48	0.60	0.58	0.69	0.67	0.55
SNRDFL			0.57	0.51	0.46	0.50	0.46	0.47	0.52	0.45	0.53	0.53	0.50
WRTSMP			0.59	0.54	0.41	0.54	0.45	0.42	0.45	0.56	0.51	0.60	0.53
SNWRFL			0.55	0.48	0.45	0.45	0.45	0.42	0.47	0.42	0.45	0.52	0.45
RDGVOC			0.47	0.47	0.43	0.45	0.40	0.45	0.42	0.38	0.41	0.45	0.44
SPLSND			0.45	0.44	0.42	0.42	0.35	0.33	0.33	0.39	0.36	0.31	0.37
SNDAWR			0.28	0.37	0.33	0.39	0.30	0.25	0.16	0.24	0.22	0.12	0.27
WRDFLU			0.29	0.21	0.27	0.22	0.17	0.24	0.28	0.20	0.26	0.21	0.23

Table 14. *(cont.)*
WJ IV Broad CHC 9-Factor (Top-Down; Model 2) CFA ML Model Results in Six Age Groups

Latent Factors/ Tests	Age Group 3–5 MD (*n* = 209)	3–5 MCV (*n* = 209)	6–8 MD (*n* = 412)	6–8 MCV (*n* = 411)	9–13 MD (*n* = 785)	9–13 MCV (*n* = 787)	14–19 MD (*n* = 842)	14–19 MCV (*n* = 843)	20–39 MD (*n* = 625)	20–39 MCV (*n* = 626)	40–90+ MD (*n* = 571)	40–90+ MCV (*n* = 574)	Median
Gq													
APPROB	0.84	0.84	0.90	0.88	0.90	0.89	0.92	0.92	0.94	0.92	0.93	0.92	0.92
NUMSEN	0.92	0.92	0.85	0.85	0.82	0.83	0.89	0.88	0.89	0.89	0.90	0.88	0.88
CALC			0.68	0.60	0.70	0.69	0.77	0.76	0.78	0.76	0.80	0.79	0.76
MTHFLU			0.41	0.30	0.40	0.36	0.38	0.39	0.42	0.44	0.40	0.37	0.39
NUMREV	0.34	0.31	0.19	0.21	0.26	0.39	0.32	0.33	0.22	0.24	—	—	0.25

Note. MD = model development sample; MCV = model cross-validation sample. Factor parameter estimates are maximum-likelihood (ML) estimates. Underlined font = parameter estimates were not significant in the MCV sample ($p < .05$). Latent factor loadings on *g* and test loadings on broad CHC factors were sorted by median factor parameter loadings for ages 6 through 90+ samples. Median values are only reported for tests that had significant loadings in the majority of samples. Italic font = Heywood cases constrained/fixed to 1.0. Dashes for tests under factors indicate that a test/factor loading was not specified in that sample. Gray shading indicates tests that were not administered to examinees ages 3 through 5.

Although the broad CHC factor top-down model was the more parsimonious and plausible structural model for the WJ IV battery, the more complex broad + narrow CHC factor bottom-up model offers potential important insights that warrant future research. Table 15 presents all test- and factor-loading parameters for the broad + narrow CHC factor bottom-up model for the MD and MCV samples across all age groups.

Table 15.
WJ IV Broad + Narrow CHC 13-Factor (Bottom-Up; Model 3) CFA ML Model Results in Five Age Groups

Latent Factors/ Tests	Age Group 6–8 MD (*n* = 412)	6–8 MCV (*n* = 411)	9–13 MD (*n* = 785)	9–13 MCV (*n* = 787)	14–19 MD (*n* = 842)	14–19 MCV (*n* = 843)	20–39 MD (*n* = 625)	20–39 MCV (*n* = 626)	40–90+ MD (*n* = 571)	40–90+ MCV (*n* = 574)	Median
g											
Gf	*1.00*	*1.00*	0.96	*1.00*	0.98	0.99	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*
Gq	0.86	0.89	0.84	0.91	0.91	0.88	0.86	0.87	0.90	0.90	0.89
Glr	0.82	0.83	0.96	0.85	0.79	0.86	0.83	0.85	0.93	0.94	0.85
Gc	0.71	0.66	0.77	0.76	0.82	0.83	0.86	0.83	0.88	0.86	0.83
Gv	0.78	0.74	0.69	0.74	0.81	0.86	0.83	0.83	0.84	0.85	0.82
Grw	0.77	0.79	0.80	0.82	0.80	0.79	0.81	0.78	0.82	0.82	0.80
Ga	0.80	0.76	0.57	0.61	0.73	0.77	0.80	0.79	0.81	0.85	0.78
Gwm	0.71	0.75	0.74	0.71	0.74	0.76	0.76	0.78	0.78	0.83	0.76
Gs	0.52	0.62	0.49	0.64	0.61	0.56	0.41	0.57	0.58	0.59	0.57
Gc											
SOC	0.78	0.83	0.81	0.81	0.85	0.84	0.86	0.84	0.87	0.84	0.84
PICVOC	0.79	0.80	0.79	0.83	0.84	0.84	0.86	0.83	0.85	0.80	0.83
HUM	0.75	0.76	0.80	0.77	0.80	0.83	0.85	0.85	0.86	0.86	0.82
ORLCMP	0.76	0.71	0.79	0.73	0.78	0.79	0.81	0.79	0.83	0.83	0.79
SCI	0.76	0.73	0.78	0.73	0.79	0.80	0.81	0.77	0.83	0.81	0.79
GENINF	0.65	0.69	0.73	0.72	0.77	0.77	0.80	0.81	0.81	0.80	0.77
ORLVOC	0.72	0.69	0.75	0.75	0.68	0.78	0.67	0.76	0.76	0.80	0.75
RDGVOC	0.46	0.49	0.48	0.47	0.53	0.50	0.49	0.54	0.54	0.51	0.50
VRBANL	0.44	0.40	0.49	0.47	0.43	0.50	0.47	0.48	0.48	0.54	0.48
SENREP	0.44	0.29	0.25	0.23	0.30	0.16	0.31	0.36	0.36	0.23	0.30
UNDDIR	0.34	0.16	0.20	0.24	0.16	0.09	0.21	0.14	0.34	0.09	0.18
STYREC	0.25	0.25	—	—	—	—	—	—	—	—	—

Table 15. *(cont.)*
WJ IV Broad + Narrow CHC 13-Factor (Bottom-Up; Model 3) CFA ML Model Results in Five Age Groups

Latent Factors/ Tests	Age Group 6–8 MD (*n* = 412)	6–8 MCV (*n* = 411)	9–13 MD (*n* = 785)	9–13 MCV (*n* = 787)	14–19 MD (*n* = 842)	14–19 MCV (*n* = 843)	20–39 MD (*n* = 625)	20–39 MCV (*n* = 626)	40–90+ MD (*n* = 571)	40–90+ MCV (*n* = 574)	Median
LA	0.47	0.47	0.43	0.44	0.43	0.49	0.39	0.33	0.20	0.26	0.43
RETFLU	0.70	0.64	0.73	0.73	0.70	0.68	0.74	0.69	0.74	0.75	0.71
RPCNAM	0.45	0.36	0.42	0.44	0.37	0.36	0.40	0.34	0.47	0.43	0.41
PHNPRO	0.41	0.50	0.39	0.43	0.46	0.37	0.26	0.45	0.36	0.29	0.40
Gf											
***Gf*-Verbal**	0.89	0.84	0.75	0.78	0.88	0.88	0.89	0.92	0.88	0.87	0.88
CONFRM	0.71	0.77	0.77	0.79	0.75	0.76	0.78	0.75	0.77	0.77	0.77
ANLSYN	0.63	0.67	0.47	0.54	0.37	0.51	0.45	0.61	0.50	0.45	0.50
VRBANL	0.41	0.47	0.43	0.43	0.48	0.41	0.45	0.44	0.45	0.40	0.43
PSGCMP	0.18	0.24	0.15	0.13	0.27	0.29	0.35	0.34	0.26	0.27	0.27
ORLVOC	0.15	0.23	0.18	0.13	0.23	0.12	0.24	0.16	0.14	0.10	0.15
***Gf*-Quantitative**	0.87	0.93	0.86	0.88	0.88	0.85	0.86	0.82	0.86	0.88	0.87
NUMSER	0.86	0.86	0.85	0.89	0.86	0.87	0.81	0.81	0.88	0.82	0.86
NUMMAT	0.69	0.71	0.77	0.74	0.79	0.80	0.82	0.85	0.83	0.85	0.80
ANLSYN	—	—	0.19	0.19	0.26	0.16	0.26	0.09	0.22	0.29	0.20
APPROB	—	—	0.30	0.15	0.17	0.17	0.20	0.16	0.20	0.08	0.17
Gwm											
OBJNUM	0.78	0.75	0.70	0.72	0.76	0.75	0.81	0.80	0.84	0.79	0.77
VRBATN	0.77	0.72	0.79	0.74	0.75	0.76	0.77	0.74	0.81	0.78	0.76
MEMWRD	0.69	0.64	0.49	0.53	0.72	0.70	0.72	0.75	0.74	0.74	0.71
UNDDIR	0.42	0.55	0.55	0.46	0.52	0.57	0.52	0.52	0.40	0.62	0.52
NWDREP	0.49	0.48	0.44	0.59	0.43	0.58	0.49	0.49	0.63	0.66	0.49
NUMREV	0.47	0.47	0.31	0.33	0.42	0.40	0.51	0.50	0.56	0.42	0.44
SENREP	0.28	0.44	0.52	0.48	0.44	0.57	0.41	0.40	0.44	0.56	0.44
LETPAT	—	—	—	—	—	—	0.23	0.15	0.26	0.17	—
LA[a]	0.48	0.52	0.42	0.41	0.47	0.40	0.43	0.53	0.70	0.56	0.48
Gs											
NUMPAT	0.83	0.85	0.85	0.80	0.79	0.73	0.70	0.79	0.80	0.77	0.80
LETPAT	0.78	0.79	0.80	0.80	0.76	0.76	0.66	0.71	0.65	0.68	0.76
WRDSPD	0.67	0.66	0.63	0.69	0.71	0.64	0.71	0.73	0.72	0.73	0.70
PAIRCN	0.66	0.67	0.58	0.60	0.58	0.59	0.44	0.54	0.41	0.51	0.58
MTHSPD	0.53	0.60	0.56	0.54	0.50	0.49	0.53	0.43	0.51	0.53	0.53
SNRDSP	0.45	0.43	0.53	0.49	0.51	0.51	0.53	0.55	0.50	0.49	0.51
SNWRSP	0.41	0.45	0.49	0.51	0.47	0.51	0.53	0.52	0.53	0.46	0.50
RPCNAM	0.25	0.25	0.23	0.26	0.25	0.23	0.28	0.30	0.23	0.22	0.25
CALC	0.28	0.36	0.29	0.23	0.18	0.16	0.19	0.12	0.14	0.16	0.18
Ga											
SEGMNT	0.75	0.81	0.71	0.74	0.74	0.73	0.78	0.76	0.74	0.80	0.75
SNDBLN	0.63	0.55	0.65	0.56	0.66	0.66	0.69	0.71	0.72	0.74	0.66
SNDAWR	0.61	0.49	0.51	0.50	0.54	0.57	0.66	0.60	0.63	0.73	0.58
PHNPRO	0.52	0.49	0.62	0.55	0.51	0.54	0.67	0.53	0.62	0.69	0.54
SPLSND	0.51	0.49	0.48	0.51	0.59	0.58	0.54	0.50	0.55	0.61	0.53
WRDATK	0.26	0.29	0.24	0.23	0.27	0.29	0.25	0.23	0.24	0.23	0.24
NWDREP	0.20	0.23	0.24	0.10	0.22	0.11	0.14	0.16	—	—	0.18
MEMWRD	—	—	0.26	0.22	—	—	—	—	—	—	—

Table 15. *(cont.)*
WJ IV Broad + Narrow CHC 13-Factor (Bottom-Up; Model 3) CFA ML Model Results in Five Age Groups

Latent Factors/ Tests	Age Group 6–8 MD (*n* = 412)	6–8 MCV (*n* = 411)	9–13 MD (*n* = 785)	9–13 MCV (*n* = 787)	14–19 MD (*n* = 842)	14–19 MCV (*n* = 843)	20–39 MD (*n* = 625)	20–39 MCV (*n* = 626)	40–90+ MD (*n* = 571)	40–90+ MCV (*n* = 574)	Median
Glr											
STYREC	0.49	0.43	0.49	0.65	0.64	0.69	0.66	0.60	0.66	0.64	0.64
WRTSMP	0.30	0.34	0.11	0.26	0.37	0.39	0.35	0.19	0.27	0.20	0.28
RDGREC	0.20	0.32	0.07	0.25	0.19	0.20	0.20	0.09	—	—	0.20
MA	0.79	0.68	0.73	0.69	0.68	0.70	0.79	0.83	0.79	0.83	0.76
VAL	0.71	0.71	0.58	0.70	0.80	0.73	0.81	0.79	0.74	0.74	0.74
MEMNAM	0.49	0.51	0.61	0.51	0.50	0.54	0.54	0.54	0.60	0.63	0.54
VISCLO	0.43	0.32	—	—	—	—	—	—	—	—	—
Gv											
VISUAL	0.80	0.75	0.67	0.75	0.65	0.71	0.75	0.76	0.72	0.71	0.74
VISCLO	0.25	0.28	0.61	0.54	0.61	0.58	0.62	0.60	0.71	0.65	0.60
PICREC	0.63	0.59	0.50	0.52	0.35	0.41	0.47	0.45	0.47	0.47	0.47
PAIRCN	0.00	0.00	0.14	0.17	0.24	0.17	0.34	0.31	0.42	0.28	0.21
Grw											
LWIDNT	0.93	0.93	0.88	0.91	0.90	0.91	0.92	0.92	0.92	0.93	0.92
SPELL	0.87	0.86	0.90	0.87	0.87	0.87	0.86	0.86	0.89	0.90	0.87
EDIT	0.79	0.81	0.87	0.84	0.81	0.83	0.81	0.82	0.85	0.85	0.83
ORLRDG	0.82	0.85	0.77	0.79	0.79	0.78	0.80	0.84	0.85	0.85	0.81
PSGCMP	0.71	0.66	0.70	0.72	0.62	0.60	0.54	0.55	0.65	0.63	0.64
WRDATK	0.62	0.61	0.56	0.64	0.57	0.56	0.59	0.61	0.63	0.65	0.61
RDGREC	0.62	0.49	0.55	0.49	0.46	0.46	0.46	0.52	0.68	0.66	0.51
SNRDSP	0.58	0.52	0.45	0.49	0.46	0.46	0.51	0.45	0.53	0.51	0.50
WRTSMP	0.52	0.49	0.51	0.50	0.42	0.38	0.40	0.53	0.49	0.58	0.50
RDGVOC	0.51	0.51	0.44	0.48	0.42	0.47	0.45	0.40	0.43	0.47	0.46
SNWRSP	0.56	0.51	0.42	0.44	0.44	0.41	0.45	0.41	0.44	0.49	0.44
SPLSND	0.41	0.43	0.37	0.39	0.32	0.31	0.32	0.38	0.37	0.31	0.37
SNDAWR	0.27	0.38	0.34	0.38	0.28	0.23	0.15	0.26	0.27	0.15	0.27
WRDSPD	0.31	0.24	0.29	0.22	0.18	0.24	0.28	0.21	0.26	0.22	0.24
Gq											
NUMSEN	0.85	0.85	0.89	0.84	0.89	0.89	0.89	0.89	0.90	0.88	0.89
APPROB	0.90	0.88	0.62	0.75	0.78	0.78	0.78	0.80	0.77	0.86	0.78
CALC	0.68	0.60	0.65	0.69	0.77	0.76	0.78	0.77	0.80	0.80	0.77
MTHSPD	0.41	0.31	0.39	0.36	0.38	0.39	0.42	0.45	0.41	0.38	0.39
NUMREV	0.21	0.22	0.34	0.32	0.29	0.31	0.22	0.24	0.18	0.32	0.27

Note. MD = model development sample; MCV = model cross-validation sample. Underlined factor codes (e.g., LA) are first-order factors. Italic font parameters = Heywood cases constrained/fixed to 1.0. Factor parameter estimates are maximum-likelihood (ML) estimates. Underlined font parameters = estimates that were not significant in the MCV sample ($p < .05$). Median values were only reported for tests that had significant loadings in the majority of samples. Dashes for tests under factors indicate that a test/factor loading was not specified in that sample. Latent factor *g* loadings and broad CHC factor test loadings were sorted by median factor parameter loadings for ages 6 through 90+ samples.

[a] Compare these *Gwm*-LA loadings with the *Gc*-LA loadings. Note the similar low to mid loadings.

An examination of Table 15 reveals that the pattern, size, and significance of the test indicators that load directly on the nine broad CHC factors are similar to those reported for the broad CHC factor top-down model. The broad CHC factors loading on the general intelligence (*g*) factor, however, differ slightly from those reported for the broad CHC factor top-down model. The *Gf* factor, which includes a broad array of test indicators that load on two separate, lower-order *Gf* factors (*Gf*-Verbal and *Gf*-Quantitative), has *g* loadings at or near unity (1.0). This is consistent with a

number of research studies that have suggested that *Gf*, when properly measured, may be identical to *g* (Gustafsson, 1984; Keith, 2005; Kvist & Gustafsson, 2008; Reynolds & Keith, 2007; Schneider & McGrew, 2012). Additionally, the presence of four possible narrow factors, as suggested by the exploratory CA, PCA, and MDS analyses, provides tentative validity evidence for the Speed of Lexical Access (LA)[5], *Gf*-Verbal, *Gf*-Quantitative, and Associative Memory (*Glr*-MA) factors.

Relationship of WJ IV Scores to Other Measures

A variety of studies were conducted to examine relations of the WJ IV scores with other measures of cognitive abilities, oral language abilities, and academic achievement. The WJ IV COG scores were examined in five studies that included the following external measures: the *Wechsler Intelligence Scale for Children®–Fourth Edition* (WISC®-IV) (Wechsler, 2003), the *Wechsler Adult Intelligence Scale®–Fourth Edition* (WAIS®-IV) (Wechsler, 2008), the *Wechsler Preschool and Primary Scale of Intelligence™–Third Edition* (WPPSI™-III) (Wechsler, 2002), the *Kaufman Assessment Battery for Children–Second Edition* (KABC-II) (Kaufman & Kaufman 2004a), the *Stanford-Binet Intelligence Scales, Fifth Edition* (SB5) (Roid, 2003), and the *Differential Abilities Scales–Second Edition* (DAS-II) (Elliott, 2007). Each of these external measures is an individually administered assessment of intelligence and cognitive abilities. Table 16 presents correlations for the WJ IV COG GIA (*g*), Brief Intellectual Ability (BIA), and *Gf-Gc* Composite cluster scores with the composite measures of general intelligence (*g*) from the external measures.[6] The .72 to .86 correlations for the WJ IV GIA cluster with the general intelligence total scores from the other intelligence batteries support the conclusion that the WJ IV GIA is a strong and valid measure of the complex set of abilities that constitute general intelligence. The magnitude of the correlations between the briefer WJ IV BIA and *Gf-Gc* Composite clusters and the general intelligence scores from the other batteries support the validity of the BIA cluster as a valid screening measure of general intelligence and the intended use of the *Gf-Gc* Composite cluster as a valid indicator of general intelligence when evaluating a person's pattern of cognitive, oral language, and academic strengths and weaknesses.

Table 16.
Correlations for Select WJ IV COG Measures and Other Measures of Cognitive Abilities

Other Measure	*N*	WJ IV COG General Intellectual Ability (GIA)	WJ IV COG Brief Intellectual Ability (BIA)	WJ IV COG *Gf-Gc* Composite
Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV)[a]	174	0.86	0.83	0.83
Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV)[a]	177	0.84	0.74	0.78
Kaufman Assessment Battery for Children–Second Edition (KABC-II)[b]	50	0.72	0.67	0.57
Stanford-Binet Intelligence Scales, Fifth Edition (SB5)[a]	50	0.80	0.79	0.82

[a]The measure reported is the Full-Scale IQ (*g*) score.

[b]The measure reported is the Mental Processing Index score.

[5]The Speed of Lexical Access factor is interpreted as a cognitively complex measure of efficiency and quickness by which individuals are able to retrieve words. For a more thorough description of the interpretation of this factor, readers are directed to Chapter 5 of the WJ IV Technical Manual (McGrew et al., 2014).

[6]Table 16 does not contain correlations for the WPPSI-III or the DAS-II, because the participants in those studies were not administered all of the WJ IV COG tests required to obtain the GIA, BIA, or *Gf-Gc* Composite cluster scores. Readers interested in details about those studies should consult Chapter 5 of the WJ IV Technical Manual (McGrew et al., 2014), where the complete results from all studies are reported.

The WJ IV OL scores were examined in four studies that included the following external measures: the *Clinical Evaluation of Language Fundamentals®–Fourth Edition* (CELF®-4) (Semel, Wiig, & Secord, 2003), the *Peabody Picture Vocabulary Test–Fourth Edition* (PPVT™-4) (Dunn & Dunn, 2007), the *Comprehensive Assessment of Spoken Language* (CASL) (Carrow-Woolfolk, 1999), and the *Oral and Written Language Scales: Listening Comprehension/Oral Expression* (OWLS™) (Carrow-Woolfolk, 1995). The CELF-4, CASL, and OWLS are individually administered multidimensional batteries of different aspects of oral language ability. The PPVT-4 is an individually administered measure of expressive vocabulary and word retrieval. Table 17 presents correlations for the WJ IV OL Oral Language, Oral Expression, Listening Comprehension, and Speed of Lexical Access cluster scores with the external composite measures of oral language. The magnitude of the correlations between the WJ IV Oral Language, Listening Comprehension, and Oral Expression clusters with select composite scores from the other language batteries supports the validity of these three WJ IV OL clusters as measures of aspects of oral language. As explained in the WJ IV Technical Manual (McGrew et al., 2014), the relative magnitude of these correlations differs as a function of the degree of common oral language abilities measured by the WJ IV OL clusters and the other oral language ability composites. The noticeably lower correlations (.14 to .57) for the WJ IV Speed of Lexical Access cluster indicates that this new cluster measures an aspect of language competence not well represented in the other oral language batteries.

Table 17.
Correlations for Select WJ IV OL Cluster Scores and Other Measures

Other Measure	*N*	WJ IV OL Oral Language Cluster	WJ IV OL Oral Expression Cluster	WJ IV OL Listening Comprehension Cluster	WJ IV OL Speed of Lexical Access Cluster
Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4), Ages 5 Through 8[a]	50	0.63	0.74	0.64	0.31
Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4), Ages 10 Through 18[a]	56	0.75	0.83	0.76	0.42
Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4), Ages 5 Through 8	50	0.74	0.70	0.69	0.43
Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4), Ages 10 Through 18	56	0.76	0.62	0.55	0.14
Comprehensive Assessment of Spoken Language (CASL), Ages 3 Through 6[b]	50	0.60	0.48	0.58	0.48
Comprehensive Assessment of Spoken Language (CASL), Ages 7 Through 17[b]	50	0.85	0.72	0.76	0.57
Oral and Written Language Scales: Listening Comprehension/Oral Expression (OWLS), Ages 3 Through 6[c]	50	0.60	0.46	0.57	0.24
Oral and Written Language Scales: Listening Comprehension/Oral Expression (OWLS), Ages 7 Through 17[c]	50	0.68	0.62	0.64	0.41

[a]The measure reported is the Core Language Composite score.

[b]The measure reported is the Core Composite score.

[c]The measure reported is the Oral Composite score.

Five studies examined the relationships between WJ IV ACH scores and scores from the following external achievement measures: the *Kaufman Test of Educational Achievement–Second Edition* (KTEA™-II) (Kaufman & Kaufman, 2004b), the *Wechsler Individual Achievement Test®–Third Edition* (WIAT®-III) (Wechsler, 2009), and the *Oral and Written Language Scales–Written Expression* (OWLS-WE) (Carrow-Woolfolk, 1996). The KTEA-II and WIAT-III are individually administered multidimensional batteries of oral language and reading, math, and writing achievement. The OWLS-WE is an individually administered measure of written expression. Table 18 presents correlations for select WJ IV ACH cluster scores with other measures of achievement. In most cases, the WJ IV reading, math, and written language cluster scores have the highest correlations with measures of the same achievement domain composites from the other batteries, providing support for the WJ IV ACH cluster scores as measures of domain-level achievement. Additionally, the high correlations between the WJ IV Brief Achievement and Broad Achievement clusters with the total and comprehensive achievement composites from the other batteries (.85 to .93) support the use of these WJ IV ACH clusters as measures of global achievement. The magnitude and pattern of the correlations between the WJ IV OL and WJ IV ACH cluster scores and the other achievement battery composite scores provide concurrent validity evidence for the WJ IV OL and WJ IV ACH clusters.

Table 18.
Correlations for Select WJ IV ACH and OL Cluster Scores and Other Measures of Achievement

Other Measure	*N*	WJ IV OL and ACH Measures[a]								
		OL	RE	MA	WL	AS	AF	AA	BFA	BDA
***Kaufman Test of Educational Achievement–Second Edition* (KTEA-II), Ages 8 Through 12**										
Oral Language	49	0.64	0.63	0.72	0.54	0.65	0.60	0.68	0.65	0.68
Reading	49	0.73	0.94	0.77	0.85	0.94	0.81	0.87	0.94	0.94
Math	49	0.64	0.70	0.94	0.66	0.78	0.73	0.84	0.78	0.83
Written Language	47	0.69	0.80	0.68	0.81	0.86	0.75	0.74	0.85	0.84
Comprehensive Achievement	47	0.75	0.83	0.88	0.77	0.87	0.81	0.87	0.88	0.91
***Kaufman Test of Educational Achievement–Second Edition* (KTEA-II), Ages 13 Through 18**										
Oral Language	49	0.74	0.70	0.64	0.71	0.66	0.50	0.75	0.70	0.67
Reading	49	0.74	0.83	0.73	0.78	0.83	0.64	0.80	0.88	0.79
Math	50	0.52	0.65	0.87	0.58	0.79	0.68	0.70	0.73	0.79
Written Language	50	0.58	0.78	0.71	0.87	0.87	0.67	0.76	0.87	0.81
Comprehensive Achievement	48	0.72	0.83	0.86	0.82	0.90	0.72	0.85	0.91	0.88

Table 18. *(cont.)*
Correlations for Select WJ IV ACH and OL Cluster Scores and Other Measures of Achievement

Other Measure	*N*	WJ IV OL and ACH Measures[a]								
		OL	RE	MA	WL	AS	AF	AA	BFA	BDA
Wechsler Individual Achievement Test–Third Edition **(WIAT–III), Grades 1 Through 8**										
Oral Language	51	0.82	0.56	0.51	0.38	0.48	0.42	0.54	0.53	0.50
Total Reading	51	0.65	0.93	0.67	0.84	0.88	0.71	0.83	0.89	0.86
Mathematics	51	0.65	0.75	0.92	0.74	0.83	0.68	0.87	0.84	0.84
Written Expression	48	0.51	0.77	0.61	0.78	0.83	0.63	0.69	0.81	0.77
Total Achievement	48	0.76	0.92	0.81	0.83	0.92	0.72	0.89	0.93	0.90
Wechsler Individual Achievement Test–Third Edition **(WIAT–III), Grades 9 Through 12**										
Oral Language	49	0.79	0.61	0.47	0.49	0.46	0.53	0.70	0.52	0.63
Total Reading	48	0.30	0.78	0.55	0.66	0.80	0.58	0.55	0.83	0.78
Mathematics	49	0.28	0.48	0.84	0.55	0.66	0.46	0.71	0.70	0.68
Written Expression	44	0.13	0.60	0.44	0.73	0.76	0.63	0.44	0.71	0.72
Total Achievement	43	0.39	0.73	0.69	0.70	0.83	0.63	0.67	0.85	0.85
Oral and Written Language Scales–Written Expression **(OWLS-WE), Ages 7 Through 17**										
Written Expression	51	—	—	—	0.75	—	—	—	—	—

[a]OL = Oral Language; RE = Reading; MA = Mathematics; WL = Written Language; AS = Academic Skills; AF = Academic Fluency; AA = Academic Applications; BFA = Brief Achievement; BDA = Broad Achievement

Performance of Clinical Samples on WJ IV Measures

The relationship between WJ IV scores and clinical group designation (e.g., individuals with learning disabilities or individuals with intellectual disabilities) provides a form of test-criterion validity evidence. Select WJ IV tests were administered to individuals within the following nine clinical groups: gifted, intellectual disabilities (ID)/mental retardation (MR), learning disabilities (LD; divided into separate groups for reading, math, and writing disabilities), language delay, attention deficit hyperactivity disorder (ADHD), head injury, and autism spectrum disorders (ASD). The comprehensiveness of the WJ IV battery made it impossible to administer all tests and clusters to all clinical groups. Instead, a diagnostic-group targeted approach to test selection was used. The patterns of mean scores for the individuals in each of the clinical groups were generally consistent with expectations. For example, the gifted and ID/MR groups displayed large differences on all tests and clusters administered, with the WJ IV test and cluster standard scores for the gifted group typically above 115 and the test and cluster standard scores for the ID/MR group typically in the 50 to 60 range. The three LD groups displayed mean WJ IV COG and WJ IV OL test and cluster standard scores that were in the 80 to 89 range.

Complete results and interpretation of the WJ IV clinical validity studies, including a description of the inclusion criteria for each study, are presented in Chapter 5 of the WJ IV Technical Manual (McGrew et al., 2014).

Summary

The procedures used to develop and validate the WJ IV have produced a diagnostic system that can be used with confidence in a variety of settings. Throughout the design and development of the WJ IV, test standards as outlined in the *Standards for Educational and Psychological Testing* (AERA et al., 2014) were followed. Special efforts were made to provide all of the relevant types of validity evidence and to provide fair, unbiased measures of an individual's cognitive abilities, oral language abilities, and academic achievement. The WJ IV Technical Manual was designed to provide test users with a comprehensive resource for evaluating the validity of the scores and interpretations from the WJ IV battery for measuring an individual's level of functioning. Interested users should consult the WJ IV Technical Manual (McGrew et al., 2014) for more in-depth details about the technical characteristics of the test.

References

Alexander, R. A. (1990). Correction formulas for correlations restricted by selection on an unmeasured variable. *Journal of Educational Measurement, 27*, 187–189.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.

Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice, 17*(1), 10–22.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Bryant, N. D., & Gokhale, S. (1972). Correcting correlations for restrictions in range due to selection on an unmeasured variable. *Educational and Psychological Measurement, 32*, 305–310.

Campbell, T., & Fiske, D. (1959). Convergent and discriminant validation by multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytical studies*. New York, NY: Cambridge University Press.

Carroll, J. B. (1998). Human cognitive abilities: A critique. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 5–24). Mahwah, NJ: Lawrence Erlbaum.

Carrow-Woolfolk, E. (1995). *Oral and Written Language Scales: Listening Comprehension/ Oral Expression*. Torrance, CA: Western Psychological Services.

Carrow-Woolfolk, E. (1996). *Oral and Written Language Scales: Written Expression*. Torrance, CA: Western Psychological Services.

Carrow-Woolfolk, E. (1999). *Comprehensive Assessment of Spoken Language*. Torrance, CA: Western Psychological Services.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin, 38*, 592.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40*, 153–193.

Cattell, R. B. (1950). *Personality: A systematic theoretical and factoral study*. New York, NY: McGraw-Hill.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* (4th ed.). San Antonio, TX: Pearson.

Efron, B., & Tibshirani, R. (1993). *An introduction to bootstrap*. New York, NY: Chapman and Hall.

Elliott, C. D. (2007). *Differential Ability Scales* (2nd ed.) San Antonio, TX: Pearson.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). *Essentials of cross-battery assessment* (2nd ed.). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. T. (2006). *The achievement test desk reference: A guide to learning disability identification* (2nd ed.). Hoboken, NJ: Wiley.

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods, 11*(4), 323–343.

Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8*, 179–203.

Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate psychology* (2nd ed., pp. 645–865). New York, NY: Academic Press.

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock, *WJ-R technical manual* (pp. 197–232). Rolling Meadows, IL: Riverside.

Kaufman, A. S., & Kaufman, N. L. (2004a). *Kaufman Assessment Battery for Children* (2nd ed.). San Antonio, TX: Pearson.

Kaufman, A. S., & Kaufman, N. L. (2004b). *Kaufman Test of Educational Achievement* (2nd ed.). San Antonio, TX: Pearson.

Keith, T. Z. (2005). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 581–614). New York, NY: Guilford Press.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.

Kvist, A. V., & Gustafsson, J. E. (2008). The relation of fluid intelligence and the general factor as a function of cultural background: A test of Cattell's investment theory. *Intelligence, 26*, 422–436.

LaForte, E. M., & McGrew, K. S. (2014). *WJ IV Tests of Achievement Alternate-Forms Equivalence* (Woodcock-Johnson IV Assessment Service Bulletin No. 1). Rolling Meadows, IL: Riverside.

McArdle, J. A. (1994). Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research, 29*(4), 409–454.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive *Gf-Gc* framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–180). New York, NY: Guilford Press.

McGrew, K. S. (2012, September). *Implications of 20 years of CHC cognitive-achievement research: Back-to-the-future and beyond CHC*. Paper presented at the Richard Woodcock Institute, Tufts University, Medford, MA.

McGrew, K. S., Dailey, D., & Schrank, F. A. (2007). *Woodcock-Johnson III/Woodcock-Johnson III Normative Update score differences: What the user can expect and why* (Woodcock-Johnson III Assessment Service Bulletin No. 9). Rolling Meadows, IL: Riverside.

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR):* Gf-Gc *cross-battery assessment*. Boston, MA: Allyn & Bacon.

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). Technical Manual. *Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside.

McGrew, K. S., & Wendling, B. (2010). CHC cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools, 47*(7), 651–675.

McGrew, K. S., Werder, J. K., & Woodcock, R. W. (1991). Technical Manual. *Woodcock-Johnson Psycho-Educational Battery–Revised*. Rolling Meadows, IL: Riverside.

McGrew, K. S., & Woodcock, R. W. (2001). Technical Manual. *Woodcock-Johnson III*. Rolling Meadows, IL: Riverside.

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. New York, NY: Guilford Press.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses in performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741–749.

Mosier, C. I. (1943). On the reliability of a weighted composite. *Psychometrika, 8*, 161–168.

Reynolds, M. R., & Keith, T. Z. (2007). Spearman's law of diminishing returns in hierarchical models of intelligence for children and adolescents. *Intelligence, 35*(1), 267–281.

Reynolds, M. R., Keith, T. Z., Flanagan, D. P., & Alfonso, V. C. (2013). A cross-battery, reference variable, confirmatory factor analytic investigation of the CHC taxonomy. *Journal of School Psychology, 51*, 535–555.

Rhemtulla, M., & Little, T. D. (2012). Tools of the trade: Planned missing data designs for research in cognitive development. *Journal of Cognition and Development, 13*(4), 425–438.

Roid, G. H. (2003). *Stanford Binet Intelligence Scales* (5th ed.). Austin, TX: PRO-ED.

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*, 112–118.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York, NY: Chapman and Hall.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). New York, NY: Guilford Press.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014a). *Woodcock-Johnson IV Tests of Achievement*. Rolling Meadows, IL: Riverside.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014b). *Woodcock-Johnson IV Tests of Oral Language*. Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014a). *Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014b). *Woodcock-Johnson IV Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside.

Semel, W., Wiig, W. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals* (4th ed.). San Antonio, TX: Pearson.

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence* (3rd ed.). San Antonio, TX: Pearson.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children* (4th ed.). San Antonio, TX: Pearson.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). San Antonio, TX: Pearson.

Wechsler, D. (2009). *Wechsler Individual Achievement Test* (3rd ed.). San Antonio, TX: Pearson.

Wolf, A. (2006). *Shorter tests through the use of planned missing data in sampling designs*. (Unpublished doctoral dissertation). Jena Thüringer Universitäts- und Landesbibliothek, Thuringia, Germany. Retrieved from http://www.db-thueringen.de/servlets/DerivateServlet/Derivate-11463/Dissertation.pdf

Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock-Johnson Psycho-Educational Battery*. Rolling Meadows, IL: Riverside.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery–Revised*. Rolling Meadows, IL: Riverside.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Rolling Meadows, IL: Riverside.

Houghton
Mifflin
Harcourt™

Riverside

3800 Golf Road, Suite 200
Rolling Meadows, IL 60008

800.323.9540
www.wj-iv.com