WJ V Technical Abstract

Erica M. LaForte, PhD

David Dailey

Kevin S. McGrew, PhD

This is a technical abstract for the Woodcock-Johnson® V (WJ V™; McGrew, Mather, LaForte, & Wendling, 2025), a comprehensive assessment system for measuring general intellectual ability (g), specific cognitive abilities, oral language abilities, and academic achievement from age 4 through 90+. It describes the updates, organization, and technical aspects of the WJ V, including reliability information and evidence to support the validity of the WJ V test and cluster score interpretations. While this document provides a high-level summary of these topics, readers should consult the Woodcock-Johnson V Technical Manual *(LaForte et al., 2025) for more comprehensive documentation.*

For technical information, please call 800.323.9540 or visit our website at https://www.riversideinsights.com/woodcock_johnson_v

Reference Citation

- To cite this document, use:

LaForte, E. M., Dailey, D., & McGrew, K. S. (2025). *WJ V Technical Abstract*. Riverside Assessments, LLC.

WJ V Technical Abstract

The *Woodcock-Johnson® V* (*WJ V™;* McGrew, Mather, LaForte, & Wendling, 2025) consists of two distinct batteries, the *Woodcock-Johnson V Tests of Cognitive Abilities* (WJ V COG; McGrew, Mather, & LaForte, 2025) and the *Woodcock-Johnson V Tests of Achievement* (WJ V ACH; Mather, McGrew, LaForte, & Wendling, 2025a). A collection of additional clinical and diagnostic tests in the *Woodcock-Johnson V Virtual Test Library* (WJ V VTL; Mather, McGrew, LaForte, & Wendling, 2025b) can be used alone or in combination with tests from the WJ V COG or WJ V ACH. Together, these components form a comprehensive system for measuring general intellectual ability (*g*),[1] 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 5,837 individuals ranging in age from 3 to 90+ years.

This document provides a high-level summary of the WJ V content updates, battery organization, and technical details. Much of it is abstracted from the expansive WJ V Technical Manual (LaForte et al., 2025) and assembled in such a way that assessment professionals who are evaluating the WJ V can gain a basic understanding of the battery's content, organization, and technical quality. Professionals who desire information beyond the summary presented in this document should consult the WJ V Technical Manual, which is downloadable from the Resources tab of *Riverside Score®*.

Overview of the WJ V

The WJ V is a theoretical, structural, interpretive, and digital revision of the *Woodcock-Johnson IV* (WJ IV™; Schrank, McGrew, & Mather, 2014). The WJ V provides measures of general intelligence; broad and narrow cognitive abilities as defined by the contemporary Cattell-Horn-Carroll theories of cognitive abilities (CHC; Schneider & McGrew, 2018); reading, mathematics, and writing achievement; oral language; and other cognitive and linguistic abilities related to academic achievement. The revision incorporates recent ideas from cognitive, neurocognitive, and developmental psychology; reading-, writing-, and math-related research; and user feedback to provide administration and interpretive options that meet contemporary assessment needs. In addition, the revision marks the move to a digital testing platform.

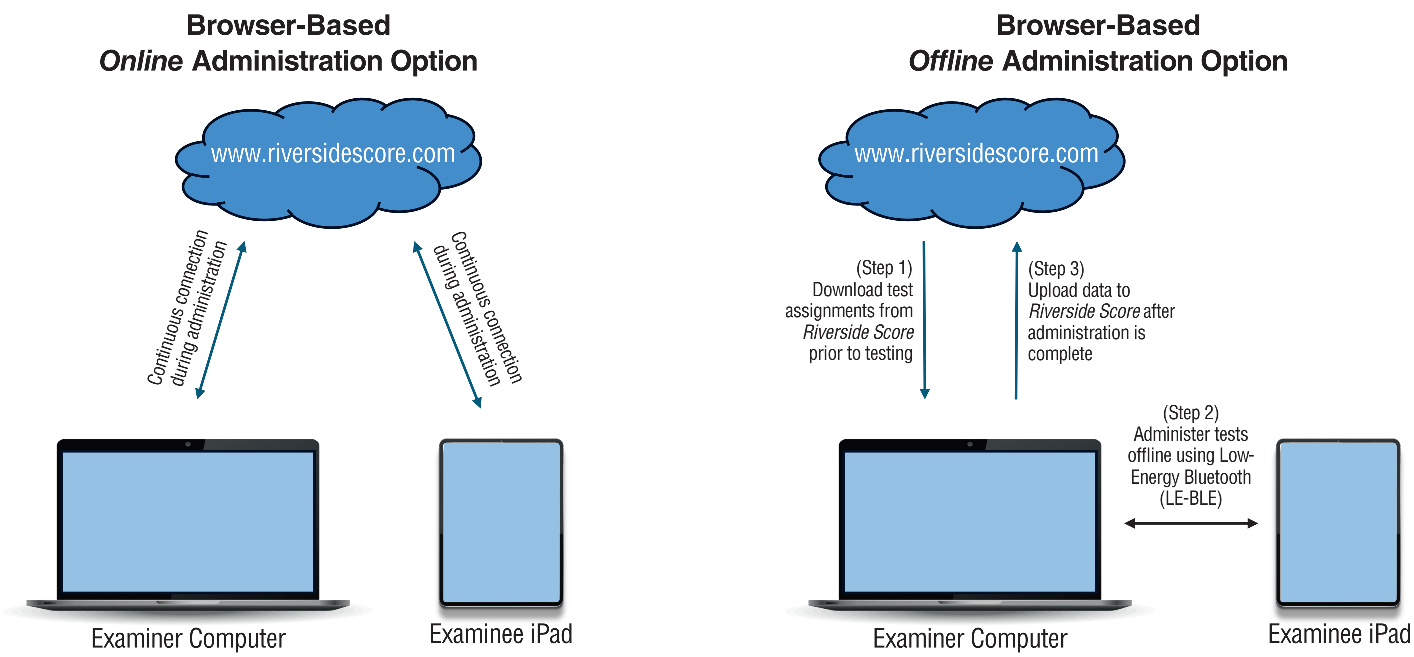
The WJ V maintains the traditional Woodcock-Johnson focus on quality while advancing (a) the comprehensive measurement of academic achievement and academic-related cognitive and linguistic abilities and (b) the measurement of cognitive abilities per CHC theory from its initial articulation in the *Woodcock-Johnson Psycho-Educational Battery–Revised* (WJ-R®; Woodcock & Johnson, 1989) and subsequent refinements in the *Woodcock-Johnson III* (WJ III®; Woodcock et al., 2001), *Woodcock-Johnson III Normative Update* (WJ III NU®; Woodcock et al., 2001, 2007), and *Woodcock-Johnson IV*. Throughout the development of the WJ V, the test authors and the *Riverside Insights®* team were guided by seven high-level revision goals that ensured both continuity and innovation. According to these goals, the WJ V should (a) be a comprehensive battery for measuring cognitive abilities, academic skills, and oral language abilities; (b) provide options for comparing an individual's performance within and across abilities; (c) reflect the most recent conceptualization of CHC theories; (d) provide robust options for measuring academic achievement in the eight IDEA-specific learning disability (SLD) areas; (e) provide increased

[1] Throughout this document, g refers to psychometric g and not to biological brain-based g (LaForte et al., 2025; McGrew, 2023).

flexibility for selective testing; (f) retain the focus on psychometric quality associated with the previous editions of Woodcock-Johnson batteries; and (g) be digitally delivered.

Except for tests that require written responses from the examinee, the WJ V is a fully digital product; most of the tests are administered and scored via a web-based digital interface. A laptop computer and iPad® replace the traditional Test Book easel and Test Record components (see Figure 1). Examiners have two options for administration: online or offline. In the online option, the examiner's laptop and the examinee's iPad are connected via web browser to a shared testing session so no Bluetooth connection is required between the two devices. In the offline administration option, the examiner downloads the test assignment from *Riverside Score* when an internet connection is available. Administration is conducted via a Low-Energy Bluetooth (LE-BLE) connection between the examiner's laptop and the examinee's iPad, and the examiner's browser captures all test data. After the administration is complete, the test data is uploaded from the examiner's laptop to *Riverside Score* when an internet connection becomes available. From a practical standpoint, the transition from paper-and-pencil to digital administration represents perhaps the most significant change in the battery's history.

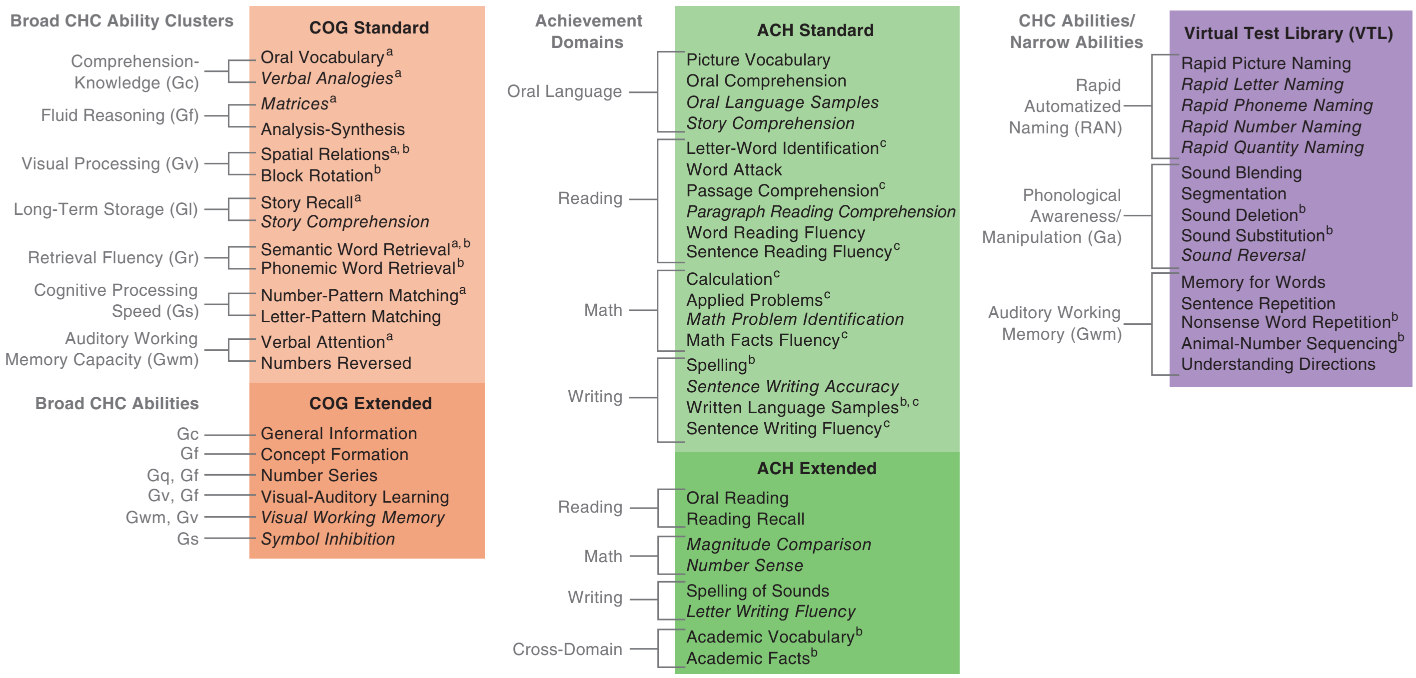
Figure 1.
Online and Offline Administration Options for the WJ V



Organization of the WJ V

Figure 2 contains a diagram showing the conceptual organization of the WJ V. Historically in the WJ, *batteries* referred to distinct physical units (i.e., Test Books, Test Records, etc.) containing different sets of tests. The digital WJ V on the *Riverside Score* platform maintains this battery structure, with tests labeled on the platform as belonging to the Cognitive Battery (Standard or Extended), the Achievement Battery (Standard or Extended), or the Virtual Test Library. There is no separate oral language battery in the WJ V; most of the tests that were in the WJ IV Tests of Oral Language have been moved to either the WJ V ACH or the WJ V VTL. In Figure 2, the tests within the COG, ACH, and VTL have been organized in a way that facilitates a high-level inspection of the WJ V structure.

Figure 2.
Conceptual Organization of the WJ V



Note. New tests are in italic font.
[a] Test is included in the General Intellectual Ability (GIA) cluster.
[b] Test is renamed or revised in WJ V.
[c] Test is included in the Broad Achievement cluster.

In the COG battery, the 14 Standard tests combine into pairs as shown to form 7 broad CHC ability clusters; the tests marked with a superscript *a* ([a]) also contribute to the General Intellectual Ability (GIA) cluster. The 6 tests in the COG Extended provide additional single-test measures of several broad CHC abilities. The ACH battery contains 4 oral language tests, 8 reading tests, 6 math tests, 6 writing tests, and 2 cross-domain achievement tests. The tests marked with a superscript *b* ([b]) contribute to the Broad Achievement cluster. Although not apparent in Figure 2, tests within the ACH battery combine in other ways to form various clusters measuring basic skills, fluency, applications, and knowledge both within and across academic domains, as well as oral language skills. The Virtual Test Library (VTL) contains 5 tests measuring different aspects of rapid automatized naming (RAN), several tests measuring aspects of phonological processing (Ga), and several other tests measuring different aspects of auditory working memory (Gwm). The VTL can be used alone, or in combination with the COG and/or ACH batteries, to gain additional diagnostic utility from the WJ V.

Tables 1 and 2 are the Selective Testing Tables for the WJ V COG + VTL and the WJ V ACH, respectively. One cluster, Phonemic Retrieval Fluency, requires one test each from the COG and VTL; one test, Story Comprehension, contributes to clusters in both the COG and ACH batteries. There are 13 tests in the WJ V that do not contribute to any clusters but may provide practitioners with diagnostically useful information for specific referral questions.

Table 1.
Selective Testing Table for the WJ V COG

			General Intelligence Clusters			CHC Broad Ability Clusters								CHC Narrow Ability and Clinical Clusters						
	CHC Broad Ability	Test	General Intellectual Ability (GIA)	Brief Intellectual Ability (BIA)	Gf-Gc Composite	Comprehension-Knowledge (Gc)	Fluid Reasoning (Gf)	Auditory Working Memory Capacity (Gwm)	Cognitive Processing Speed (Gs)	Retrieval Fluency (Gr)	Long-Term Storage (Gl)	Visual Processing (Gv)	Phonological Awareness (Ga)	Phonological Manipulation (Ga)	Auditory Memory Span (Gwm)	Cognitive Efficiency (CE)	Phonemic Retrieval Fluency (Gr)	RAN–Reading (Gs, Gr)	RAN–Math (Gs, Gr)	Single Tests
COG Standard Tests	Gc	Oral Vocabulary	■	■	■	■														
	Gf	Matrices	■	■	■		■													
	Gv	Spatial Relations	■									■								
	Gl	Story Recall	■								■									
	Gr	Semantic Word Retrieval	■							■										
	Gwm	Verbal Attention	■	■				■								■				
	Gs	Number-Pattern Matching	■						■							■				
	Gc, Gf	Verbal Analogies	■		■	■														
	Gf	Analysis-Synthesis			■		■													
	Gv	Block Rotation										■								
	Gl	Story Comprehension									■									
	Gr	Phonemic Word Retrieval								■							■			
	Gwm	Numbers Reversed						■												
	Gs	Letter-Pattern Matching							■											
COG Extended Tests	Gc	General Information																		■
	Gf	Concept Formation																		■
	Gq, Gf	Number Series																		■
	Gv, Gf	Visual-Auditory Learning																		■
	Gwm, Gv	Visual Working Memory																		■
	Gs	Symbol Inhibition																		■
Virtual Test Library	Ga, Gwm	Nonsense Word Repetition																		■
	Gs, Gr	Rapid Picture Naming																■		
	Gwm	Animal-Number Sequencing																		■
	Ga	Sound Reversal																		■
	Gs, Gr	Rapid Letter Naming																■		
	Gwm, Gf	Understanding Directions																		■
	Ga	Sound Blending											■							
	Gr, Ga	Rapid Phoneme Naming															■	■		
	Gwm	Memory for Words													■					
	Ga	Segmentation											■							
	Gs, Gr	Rapid Number Naming																	■	
	Gwm	Sentence Repetition													■					
	Ga	Sound Deletion												■						
	Gs, Gv	Rapid Quantity Naming																	■	
	Ga	Sound Substitution												■						

Note. ■ Indicates tests that are required to create the cluster listed.

Table 2.
Selective Testing Table for the WJ V ACH

			Oral Language Clusters				Reading Clusters				Math Clusters				Writing Clusters				Cross-Domain Achievement Clusters					
	CHC Broad Ability	Test	Oral Language (Gc)	Listening Comprehension (Gc)	Oral Expression (Gc, Gl)	Vocabulary (Gc)	Brief Reading (Grw)	Basic Reading Skills (Grw)	Reading Fluency (Grw)	Reading Comprehension (Grw)	Brief Math (Gq)	Math Calculation Skills (Gq)	Number Concepts (Gq)	Math Problem Solving (Gq)	Brief Writing (Grw)	Basic Writing Skills (Grw)	Spelling Skills (Grw, Ga)	Written Expression (Grw)	Broad Achievement (Grw, Gq, Gs)	Academic Skills/Brief Achievement (Grw, Gq)	Academic Fluency (Gs)	Academic Applications (Grw, Gq)	Academic Knowledge (Gc)	Phoneme-Grapheme Knowledge (Grw, Ga)
ACH Standard Tests	Gc	Picture Vocabulary	■		■	■																		
	Grw	Letter-Word Identification					■	■											■	■				
	Gq	Calculation									■	■							■	■				
	Grw	Spelling														■	■		■	■				
	Gc	Oral Comprehension	■	■																				
	Grw, Ga	Word Attack						■																■
	Gs, Gq	Math Facts Fluency										■							■		■			
	Grw	Sentence Writing Accuracy													■	■								
	Grw, Gc	Passage Comprehension					■			■									■			■		
	Gq	Applied Problems									■			■					■			■		
	Gs, Grw	Sentence Reading Fluency							■										■		■			
	Grw	Written Language Samples													■			■	■			■		
	Gl, Gc	Oral Language Samples	■		■																			
	Gs, Grw	Sentence Writing Fluency																■	■		■			
	Grw, Gc	Paragraph Reading Comprehension								■														
	Gl	Story Comprehension	■	■																				
	Gs, Grw	Word Reading Fluency							■															
	Gq, Gf	Math Problem Identification												■										
ACH Extended Tests	Gs, Gq	Magnitude Comparison											■											
	Gq	Number Sense											■											
	Ga, Grw	Spelling of Sounds															■							■
	Grw	Oral Reading																						
	Grw, Gl	Reading Recall																						
	Gc	Academic Vocabulary				■																	■	
	Gc	Academic Facts																					■	
	Gs	Letter Writing Fluency																						

Note. ■ Indicates tests that are required to create the cluster listed.

Changes from WJ IV to WJ V

Routine content updates are made during the test development process for every new edition of the Woodcock-Johnson. These updates may include dropping out-of-date items, adding contemporary items to a test's item pool, revising test artwork, changing administration procedures, and clarifying test or item instructions. These minor updates typically have little impact on the psychometric constructs measured by the tests. More significant updates, such as dropping, renaming, or moving tests and adding new tests, are also made to align the battery with recent developments in educational theory, research, and practice. Several of these changes to the WJ V are evident in Figures 2 through 4 and are described in this section.

Changes to the COG Battery

The WJ V test design blueprint is based on the most current articulation of CHC theories by Schneider and McGrew (2018), which is a blend of Carroll's (1993) and Horn's (1991) theoretical models. The following changes highlight the reflection of contemporary CHC theories in the WJ V.

- **Removing auditory processing (Ga) from the COG battery.** The Ga tests and clusters were moved to the Virtual Test Library (VTL) in the WJ V. Auditory processing was one of the least robust broad cognitive ability domains identified in Carroll's (1993) seminal extant factor analysis research and today remains one of the least investigated broad CHC ability domains (Schneider & McGrew, 2018). The removal of Ga measures from the WJ V GIA cluster and from the COG battery was supported by both the lack of an interdisciplinary framework for understanding Ga abilities and WJ IV user feedback noting deflated GIA scores for examinees with low auditory processing abilities. However, the measure of auditory processing has not been lost in the WJ V. On the contrary, the addition of a new Ga test, Sound Reversal, and the inclusion of two new Ga clusters in the VTL, Phonological Awareness and Phonological Manipulation, provides even more robust measurement of Ga in the WJ V.
- **Splitting the Glr cluster into separate Gl and Gr clusters.** Based on findings from factor-analytic research published over the past 15 to 20 years and following the recommendation of Schneider & McGrew (2018), the Long-Term Retrieval (Glr) cluster from the WJ IV was split into two separate clusters in the WJ V COG battery: Long-Term Storage (Gl),[2] which measures an individual's efficiency in storing new information in long-term memory, and Retrieval Fluency (Gr), which measures the speed at which an individual can load information from long-term memory into working memory stores for further cognitive processing (Schneider & McGrew, 2018). Measuring Gl and Gr broad abilities with separate cluster scores allows practitioners to assess an examinee's ability to store new information in long-term memory (Gl) separately from assessing their ability to load information from long-term memory into working memory structures for further processing.
- **Updating the composition of the Fluid Reasoning (Gf) cluster.** The composition of the Fluid Reasoning broad ability cluster is completely new in the WJ V. The two new tests comprising the Gf cluster, Matrices and Verbal Analogies, replace the WJ IV Number Series and Concept Formation tests as the primary measures of fluid reasoning and provide more robust measurement of *figural-visual* and *auditory-linguistic* Gf abilities, respectively. The updated Gf cluster composition does not include a controlled learning measure, consistent with Schneider and McGrew's (2018) recommendation that fluid reasoning be measured, at least partially, "in the wild" (i.e., with tasks that do not provide external examiner- or digital-administered scaffolding or feedback).
- **Updating the content of the Visual Processing (Gv) cluster.** The WJ IV Picture Recognition test was dropped, and the WJ IV Visualization subtests Spatial Relations and Block Rotation have been expanded into full-length versions. These two tests now comprise the WJ V Visual Processing cluster. This change was made in response to recent research suggesting strong links between Gv abilities and the STEM fields of science, technology, engineering, and mathematics (Hegarty, 2010). Although many studies looking at CHC-achievement relations have failed to establish a clear link between Gv and STEM success, Schneider and McGrew (2018) suggested that this may be due to the fact that Gv tests in the major intelligence batteries tend to measure "threshold" abilities and not complex visual thought. To address this concern, new items have been added to both

[2] *Long-term retrieval* is the term used for the WJ V Gl cluster, which differs from Schneider and McGrew's (2018) Gl term of *learning efficiency*. The reason for this change is described in Appendix A of the WJ V Technical Manual (LaForte et al., 2025).

Spatial Relations and Block Rotation to increase the spatial manipulation and processing demands of these tests.

- **Updating the Cognitive Efficiency (CE) cluster.** Cognitive efficiency represents the amalgam of processing speed (Gs) and working memory capacity (Gwm). The WJ IV Cognitive Efficiency (CE) cluster included the Letter-Pattern Matching and Numbers Reversed tests; adding Verbal Attention and Number-Pattern Matching yielded an extended version of the cluster. The WJ V Cognitive Efficiency cluster contains only the two latter tests.
- **Updating the General Intellectual Ability (GIA) cluster.** The seven-test WJ IV GIA has been replaced with an eight-test version in the WJ V. The following are some of the changes reflected in the WJ V GIA.
  - **Eliminating differential weighting of the GIA tests across age groups.** Re-analysis of the WJ IV norming data revealed that differentially weighted GIA scores correlate almost perfectly with equally weighted GIA scores and therefore, the added computational complexity of differential weighting was deemed unnecessary. Thus, the tests in the WJ V GIA are equally weighted across all ages.
  - **Updating the battery composition to reflect current CHC theory.** Following the removal of the auditory processing (Ga) tests from the WJ V COG battery, the GIA cluster no longer includes a Ga test. In addition, two tests, Story Recall and Semantic Word Retrieval, now reflect the separate Gl and Gr CHC broad abilities in the GIA cluster.
  - **Replacing Number Series with the new Matrices test as the primary measure of Fluid Reasoning.** The WJ IV GIA cluster included the Number Series test as the primary measure of fluid reasoning. However, WJ IV user feedback, as well as published and unpublished research, suggested that including Number Series in the WJ IV GIA may have confounded the measurement of general intelligence for individuals with deficits in mathematics skills. In addition to being well-supported measures of fluid reasoning, matrices tests may be less influenced by cultural and linguistic factors than other types of Gf tests (McCallum, 2017), measuring "nonverbal" intelligence and higher levels of cognition (Jensen, 1998; Pahor, et al., 2019; Raven, 2000).
  - **Including a full-length Spatial Relations test as the primary Gv measure.** In the WJ IV GIA, the visualization (Gv) broad CHC ability was measured with a two-part Visualization test that included shortened Spatial Relations and Block Rotation subtests. Spatial Relations showed higher overall g loadings than Block Rotation in the structural analysis of the WJ III, WJ IV, and WJ V norming data and was thus chosen to be the primary Gv measure in the WJ V. New, more complex items were added to the full-length WJ V Spatial Relations test to increase the demands for complex manipulation and processing of visual-spatial stimuli in working memory.
  - **Including Semantic Word Retrieval as the primary Gr measure.** Semantic Word Retrieval, which appeared in the WJ IV Tests of Oral Language (WJ IV OL; Schrank, Mather, & McGrew, 2014) as Retrieval Fluency, was moved into the WJ V COG and assumes the position as the Gr measure in the GIA.
  - **Adding an eighth test, Verbal Analogies, as a mixed measure of Gf and Gc.** In the WJ V norming data, the Verbal Analogies test showed the highest psychometric *g* loading of any test (median = 0.72) across all age groups. Although the eight tests are equally weighted in the GIA cluster, the addition of this mixed Gf/Gc test inherently places heavier emphasis on the Gf and Gc broad abilities in the GIA. This is consistent with research by several prominent intelligence scholars, who have identified that problem solving (Gf) and verbal abilities (Gc) are the hallmarks, or the "king and queen" of intelligence (Schrank, McGrew, & Mather, 2015).

- **Adding new tests that measure the executive functions of working memory, processing speed, and inhibitory control.** The adaptation of the WJ V into a fully digital format provided an opportunity to add new tests that would have been difficult or impossible to administer in a paper format. Additionally, the WJ V author team wanted to enhance the capabilities of the WJ V for measuring *executive functions*, which refer to a set of cognitive processes that regulate an individual's thoughts and behaviors (Miyake et al., 2000; Miyake & Friedman, 2012). In the past 15 to 20 years, there has been an increase in research on executive functions and their relationship to academic achievement (c.f., Best et al., 2011; Duncan et al., 2007; Titz & Karbach, 2014), learning disabilities (c.f., Booth, et al., 2010; Smith-Spark et al., 2016), and conditions such as attention deficit/hyperactivity disorder (ADHD; c.f., Berlin et al., 2004; Diamond, 2005) and autism spectrum disorder (ASD; c.f., Demetriou et al., 2019). To this end, two new tests were added to the WJ V to measure the working memory, attentional control, and inhibitory control aspects of executive functioning:
  - **Visual Working Memory.** This test, a mixed measure of Gwm and Gv, assesses an examinee's ability to recall a visual-spatial pattern of stimuli "in the context of processing." The structural analysis of the WJ V norming data provides support for the bifurcation of Gwm into separate auditory and visual capacity constructs. Designed as a robust replacement for the WJ IV Picture Recognition test, this new test expands the measurement of Gwm in the WJ V to include visual, as well as auditory (i.e., verbal), working memory capacity. Visual Working Memory may be a useful Gwm measure for examinees with auditory processing difficulties. Additionally, recent research has suggested that auditory (i.e., verbal) working memory abilities and visual working memory abilities differentially predict reading and math achievement for children of different ages (c.f., Van de Weijer-Bergsma et al., 2015).
  - **Symbol Inhibition.** The new Symbol Inhibition test, designed to replace the WJ IV Pair Cancellation test, is a measure of Gs that taps into the executive functions of sustained attentional control, working memory, and inhibitory control. Although further research is needed to evaluate its convergent validity with other performance and behavioral measures of executive functioning, it shows early potential as a measure of executive functioning for both typical and clinical groups (LaForte, in press). In a small clinical sample of children with ADHD who were administered tests from the WJ V battery during the norming study, the Symbol Inhibition score was among the lowest of the group's mean test scores.
- **Shifting several tests to the VTL.** As noted earlier, all Ga clusters have been moved to the WJ V VTL. The following tests were also moved from the WJ IV COG to the WJ V VTL:
  - Nonword Repetition, which was renamed Nonsense Word Repetition to more accurately describe the measured construct;
  - Object-Number Sequencing, which was renamed Animal-Number Sequencing; and
  - Memory for Words.

Table 3 contains a description of each WJ V COG test and its broad CHC classification.

Table 3.
WJ V COG Tests, Broad CHC Classifications, and Tasks

COG Standard Tests	Broad CHC Classification	Description of Test Task
Oral Vocabulary	Gc	This test is comprised of two subtests: Synonyms and Antonyms. The examinee hears a word presented from an audio recording (which is also viewable on the iPad screen) and says an appropriate synonym or antonym.
Matrices	Gf	The examinee must deduce a rule and select, from among four options, the one that best completes the pattern in a figural matrix. Early items contain 2 x 2 matrices; later items contain 3 x 3 matrices.
Spatial Relations[a]	Gv	Using visual-mental rotation processes, the examinee must determine which two or three 2-dimensional puzzle pieces (from among six options) go together to form the shape in the key.
Story Recall	Gl	The examinee listens to short stories from an audio recording and then retells the stories with as much detail as possible.
Semantic Word Retrieval	Gr	The examinee has 1 minute to say as many words as possible that fit into a semantic category. There are three trials, each with a different semantic category.
Verbal Attention	Gwm	The examinee hears a series of intermingled animals and digits presented from an audio recording. Then the examinee must answer a specific question about the sequence; for example, "Tell me the animal that came before five."
Number-Pattern Matching	Gs	The examinee has 3 minutes to tap pairs of identical 1- to 3-digit numbers among rows of six numbers.
Verbal Analogies	Gc, Gf	The examinee sees three words of a verbal analogy (e.g., *A* is to *B* as *C* is to . . .) and hears the examiner read the analogy orally. The examinee then says a word to complete the analogy.
Analysis-Synthesis	Gf	During the training phases of this controlled learning test, the examinee learns to use a key containing colored squares to solve puzzles. The examinee then uses deductive reasoning to solve each puzzle and name the missing color(s). With the exception of the last several items, the examiner provides immediate feedback for correct and incorrect answers.
Block Rotation[a]	Gv	Using visual-mental rotation processes, the examinee must determine which two (from among five options) 3-dimensional block figures match the figure in the key.
Story Comprehension[b]	Gl	The examinee listens to short stories from an audio recording and then answers story-specific comprehension questions read orally by the examiner.
Phonemic Word Retrieval[c]	Gr	The examinee has 1 minute to say as many words as possible that begin with a specific sound. There are three trials, each with a different beginning sound.
Numbers Reversed	Gwm	The examinee hears a sequence of numbers from an audio recording and then says the numbers in reverse order.
Letter-Pattern Matching	Gs	The examinee has 3 minutes to tap pairs of identical nonword combinations of one to four letters among rows of letters or letter combinations.

Table 3. *(cont.)*
WJ V COG Tests, Broad CHC Classifications, and Tasks

COG Extended Tests	Broad CHC Classification	Description of Test Task
General Information	Gc	This test is comprised of two subtests: Where and What. The examinee answers "Where would you find . . .?" and "What would you do with . . .?" questions read orally by the examiner.
Concept Formation	Gf	During the training phases of this controlled-learning task, the examinee learns rules for solving puzzles that require grouping of pictures on shape, size, color, and quantity. The examinee then uses inductive reasoning to state the rule that explains how one or more pictures is/are different from the other pictures in each puzzle. With the exception of the last several items, the examiner provides immediate feedback for correct and incorrect answers.
Number Series	Gq, Gf	The examinee sees a series of numbers with one number missing and must determine the underlying rule to provide the missing number.
Visual-Auditory Learning	Gv, Gf	During the training phases, the examinee learns relationships between words and pictures (rebuses). The examinee must then read "sentences" formed by the rebuses. Sentences increase in difficulty as new rebuses are presented in each training phase; the examiner provides oral feedback and error correction. This controlled learning task mirrors the early reading process.
Visual Working Memory	Gwm, Gv	After briefly viewing a pattern of dots on the screen (from 1 to 9 dots presented inside randomly displayed patterns of squares), the examinee completes a simple visual distractor task, while concurrently retaining the dot patterns in active memory. Then the examinee must recall the location of the dots from the first screen (in randomly displayed patterns of 2 to 23 empty boxes) immediately after the visual distractor task.
Symbol Inhibition	Gs	The examinee has 1 minute to quickly tap successive colored shapes in a row of shapes but must not tap (i.e., inhibit) the shape(s) that are identical to the shape(s) in the key at the top of the screen. The task becomes more complex as additional shapes are added to the key.

Note. Italic font denotes *new* tests. For pragmatic reasons, some tests in this table may contain CHC ability classifications that do not appear in Table 1. This table is intended to contain more comprehensive and detailed information for test interpretation, whereas Table 1 is intended to be a user-friendly selective testing tool. Some of the differences in the broad CHC classifications listed in the two tables reflect differences in the factor loadings of the tests in three equally plausible CHC structural models, which are evaluated and discussed in detail in the WJ V Technical Manual (LaForte et al., 2025).
[a] In the WJ IV, this test was a subtest of Visualization.
[b] Included in both the WJ V COG and ACH batteries.
[c] In the WJ IV, this test was a subtest of Phonological Processing.

Changes to the ACH Battery

Recent research, together with WJ IV user feedback, informed the updates to the WJ V ACH battery to align with the stated goals of the WJ V. The ACH updates include the following.

- **Adding the oral language tests and clusters.** Prior to the WJ IV, the oral language tests and clusters resided in the achievement battery. The WJ IV included a stand-alone Oral Language battery that housed not only the oral language measures, but also several measures of auditory processing (Ga) and auditory working memory (Gwm). In the WJ V, the oral language measures have been moved back to the ACH battery.
- **Adding several new tests to better assess the IDEA SLD areas.** Several new tests were added to the WJ V ACH to provide users with more comprehensive assessment of oral language and achievement areas for evaluation of specific learning disabilities (SLDs) under the Individuals with Disabilities Education Act (IDEA, 2004). These include:

    - **Story Comprehension.** This new test replaces the WJ IV Understanding Directions test in the Listening Comprehension cluster. In Story Comprehension, the examinee hears a story from an audio recording and orally answers factual and inferential questions about the story. Together with Oral Comprehension, a cloze listening comprehension task, the WJ V cluster provides more ecologically valid measurement of listening comprehension.
    - **Oral Language Samples.** This new test provides an oral-language parallel to the Written Language Samples (formerly Writing Samples) test, allowing examiners to make comparisons between an examinee's oral and written language expression skills. It replaces Sentence Repetition in the WJ V Oral Expression cluster. In Oral Language Samples, the examinee provides one-sentence responses to orally administered prompts, sometimes accompanied by pictures. Together with Picture Vocabulary, which assesses single-word vocabulary, this WJ V cluster provides more comprehensive measurement of oral expression.
    - **Paragraph Reading Comprehension.** This new test provides a reading parallel to the oral-language Story Comprehension test, allowing examiners to make comparisons between an examinee's listening and reading comprehension skills. This test replaces Reading Recall and, together with the cloze-format Passage Comprehension test, contributes to the WJ V Reading Comprehension cluster. In Paragraph Reading Comprehension, the examinee reads short passages and then orally answers factual and inferential questions about the passage. With the addition of this new test, the WJ V Reading Comprehension cluster better reflects the tasks a student will encounter in the classroom.
    - **Math Problem Identification.** This new test replaces Number Matrices and, together with Applied Problems, contributes to the WJ V Math Problem Solving cluster. In Math Problem Identification, the examinee sees a math problem that is unsolvable and must say what is wrong, or why the problem cannot be solved. This test was developed to better mimic the "messy," real-world applications of math concepts, which require an individual to assess whether sufficient information exists to solve a problem. The addition of Math Problem Identification provides a more ecologically valid measurement of math problem solving skills in the WJ V.
    - **Sentence Writing Accuracy.** This new test assesses writing mechanics, including spelling, punctuation, and capitalization, as well as the examinee's ability to accurately write dictated sentences. This test was added in response to user criticism that by failing to penalize examinees for errors in writing mechanics, the WJ IV produced inflated writing scores. Developed as an ecologically valid measure to better represent classroom writing tasks, Sentence Writing Accuracy combines with Written Language Samples in the new Brief Writing cluster and with Spelling (which replaces the WJ IV Editing test) in the Basic Writing Skills cluster.
- **Adding two new tests to measure quantitative knowledge (Gq).** The new Number Sense and Magnitude Comparison tests comprise a new cluster, Number Concepts, in the WJ V ACH. The Number Sense test measures an examinee's ability to understand number relationships and to compare, judge, and estimate size, quantity, and position. Research suggests that the skills measured by the Number Sense test are strong predictors of later reading and math achievement (Duncan et al., 2007). Magnitude Comparison is a timed test that measures the examinee's ability to assess numerical magnitude. Meta-analytic research also suggests that the skills measured by Magnitude Comparison are predictive of later math achievement (Siegler & Braithwaite, 2017).
- **Adding a test to screen for *emerging* writing difficulties.** The new Letter Writing Fluency test was designed as a screening measure for writing difficulties in children ages

4 to 9. Poor automaticity of letter writing has been found to be related to dysgraphia and difficulty with written language composition (Berninger, 2020; Kent & Wanzek, 2016; Ray et al., 2022). The WJ V Letter Writing Fluency test includes two 30-second tasks: printing the letters of the alphabet from memory in any order and then copying a model of the alphabet in order. This test does not contribute to any WJ V clusters.

- **Adding new clusters or updating cluster compositions to better reflect the measurement of the academic domains of reading, math, and writing.** New WJ V ACH clusters and existing clusters with content updates or minor name changes include the following:
    - **Brief Reading, Brief Math, and Brief Writing.** The word *brief* was added to the WJ V Reading, Math, and Written Language (now Writing) academic-domain clusters in the WJ V. The WJ V authors felt that it was important for those who use and interpret WJ V ACH scores to understand that these two-test clusters do not cover the full breadth of relevant skills necessary for academic success within each domain. For example, although the WJ V Brief Reading cluster includes measures of word reading and reading comprehension, nonsense word decoding and fluency are also important aspects of reading that are not measured by the Brief Reading cluster. The Brief Reading and Brief Math clusters are identical in content to their WJ IV counterparts: Brief Reading contains the Letter-Word Identification and Passage Comprehension tests, and Brief Math contains the Calculation and Applied Problems tests. The WJ V Brief Writing cluster composition, however, is slightly different than its WJ IV counterpart; it includes the Written Language Samples (formerly Writing Samples) test but replaces the WJ IV Editing test with Sentence Writing Accuracy, which measures the spelling, capitalization, and punctuation aspects of written language production.
    - **Reading Fluency.** In the WJ IV, this cluster was comprised of the Sentence Reading Fluency and Oral Reading tests. To better measure reading automaticity, Oral Reading has been replaced with Word Reading Fluency in the WJ V Reading Fluency cluster.
    - **Spelling Skills.** Within the WJ IV, spelling was a strong predictor of writing skills (Parkin, 2021). The new two-test Spelling Skills cluster includes the Spelling and Spelling of Sounds tests, allowing examiners more robust measurement of both real- and nonsense-word spelling.
- **Updating the cross-domain academic knowledge tests.** In the WJ IV, academic knowledge was measured via a cluster that contained Science, Social Studies, and Humanities tests. In the WJ V, out-of-date items from these domain-specific tests were eliminated, and the item pools were combined and reconfigured into separate Academic Facts and Academic Vocabulary tests. Together, these two tests now form the Academic Knowledge cluster. New items were written for both tests' item pools to cover 21st-century science, technology, engineering, and math (STEM) knowledge and vocabulary. In these two tests, the examinee answers factual or vocabulary questions, read orally by the examiner, about topics in language arts, mathematics, science, social studies, and humanities. Organizing the measurement of academic knowledge into separate cross-domain *facts* and *vocabulary* tests in the WJ V had an additional benefit: It allowed the creation of a more robust WJ V Vocabulary (Gc) cluster that comprises the Academic Vocabulary and Picture Vocabulary tests. Where the former measures vocabulary acquired through formal schooling, the latter is a traditional measure of vocabulary that is acquired primarily through experience or acculturation.

Table 4 contains a description of each WJ V ACH test and its broad CHC classification.

Table 4.
WJ V ACH Tests, Broad CHC Classifications, and Tasks

ACH Standard Tests	Broad CHC Classification	Description of Test Task
Picture Vocabulary	Gc	The examinee orally names pictured objects.
Letter-Word Identification	Grw	On the earliest items, the examiner says a letter and the examinee taps (from among six options) the letter on the screen. On later items, the examinee reads printed letters and words aloud.
Calculation	Gq	Working in a Response Booklet, the examinee solves math problems that are presented in a traditional format, including problems of addition, subtraction, multiplication, division, and more complex mathematical operations.
Spelling	Grw	Working in a Response Booklet, the examinee writes words that the examiner dictates orally.
Oral Comprehension	Gc	In this cloze test, the examinee listens to an audio recording of a short passage that contains a missing word and then says a word that makes sense in the context of the passage.
Word Attack	Grw, Ga	On the early items, the examinee must produce the sounds for single letters and letter combinations. On later items, the examinee reads printed, phonically regular nonsense words (i.e., pseudowords) aloud.
Math Facts Fluency	Gs, Gq	Working in a Response Booklet, the examinee has 3 minutes to solve simple arithmetic problems (addition, subtraction, and multiplication).
Sentence Writing Accuracy	Grw	Working in a Response Booklet, the examinee writes sentences that are dictated from an audio recording. The examinee's responses are scored based on dictation accuracy, spelling, capitalization, and punctuation.
Passage Comprehension	Grw, Gc	In this cloze test, the examinee reads a short passage that contains a missing word and then says a word that makes sense in the context of the passage. Passage difficulty varies by length, the presence or absence of pictorial stimuli, and the complexity of vocabulary and syntax.
Applied Problems	Gq	The examinee answers math problems, some of which are accompanied by visual stimuli, that are read orally by the examiner.
Sentence Reading Fluency	Gs, Grw	The examinee has 3 minutes to read simple sentences and then indicate whether each sentence is true by tapping *Yes* or *No*.
Written Language Samples	Grw	Working in a Response Booklet, the examinee writes single-word, two-word, or full-sentence responses to prompts read orally by the examiner. The examinee is not penalized for errors in basic writing skills such as spelling or punctuation.
Oral Language Samples	Gl, Gc	The examinee listens to prompts that are read orally by the examiner, some of which are accompanied by words and pictures. On early items, the examinee must provide a word that finishes the sentence. On later items, the examinee must provide a complete sentence that satisfies the prompt requirements.
Sentence Writing Fluency	Gs, Grw	Working in a Response Booklet, the examinee has 5 minutes to write simple sentences about pictures using all three stimulus words provided for each picture.
Paragraph Reading Comprehension	Grw, Gc	The examinee silently reads short stories and then answers story-specific comprehension questions read orally by the examiner.
Story Comprehension[a]	Gl	The examinee listens to short stories from an audio recording and then answers story-specific comprehension questions read orally by the examiner.
Word Reading Fluency	Gs, Grw	The examinee has 3 minutes to tap pairs of semantically related words among rows of six words.

Table 4. *(cont.)*
WJ V ACH Tests, Broad CHC Classifications, and Tasks

ACH Standard Tests	Broad CHC Classification	Description of Test Task
Math Problem Identification	Gq, Gf	The examinee hears unsolvable math problems read orally by the examiner, some of which are accompanied by visual stimuli. The examinee must then describe what is missing or what is wrong with each math problem that makes it unsolvable.
ACH Extended Tests	**Broad CHC Classification**	**Description of Test Task**
Magnitude Comparison	Gs, Gq	The examinee has 2 minutes to tap the value in each pair of numerical values that is greater. Values are expressed as groups of shapes, numbers (single digits that progress to larger numbers), fractions, or decimals.
Number Sense	Gq	The examinee provides oral answers to math-related questions, some of which are accompanied by visual stimuli, that are read orally by the examiner. The questions require the examinee to demonstrate understanding of number relationships and math vocabulary, as well as the abilities to compare, judge, and estimate size, quantity, position, and volume.
Spelling of Sounds	Ga, Grw	Working in a Response Booklet, the examinee writes nonsense words that are dictated from an audio recording.
Oral Reading	Grw	The examinee orally reads sentences that increase in difficulty.
Reading Recall	Grw, Gl	The examinee silently reads short stories and then retells the stories with as much detail as possible.
Academic Vocabulary[b]	Gc, Gkn	The examinee answers vocabulary questions about academic topics such as literacy, mathematics, science, and humanities that are read orally by the examiner.
Academic Facts[b]	Gc, Gkn	The examinee answers factual questions about academic topics such as literacy, mathematics, science, and humanities that are read orally by the examiner.
Letter Writing Fluency[c]	Gs, Gps	Working in a Response Booklet, the examinee has 30 seconds to print the letters of the alphabet from memory in any order and then 30 seconds to copy a model of the alphabet, in order.

Note. Italic font denotes *new* tests. For pragmatic reasons, some tests in this table may contain CHC ability classifications that do not appear in Table 2. This table is intended to contain more comprehensive and detailed information for test interpretation, whereas Table 2 is intended to be a user-friendly selective testing tool. Some of the differences in the broad CHC classifications listed in the two tables reflect differences in the factor loadings of the tests in three equally plausible CHC structural models, which are evaluated and discussed in detail in the WJ V Technical Manual (LaForte et al., 2025).
[a] Included in both the COG and ACH batteries.
[b] This test contains items from the WJ IV Science, Social Studies, and Humanities tests.
[c] Norms available for ages 4 through 9 only.

Creation of the Virtual Test Library

The transition from paper-and-pencil to digital administration in the WJ V allows greater flexibility for selective testing, enabling tailored combinations of tests and clusters for diagnostic and instructional purposes. While several tests in the Virtual Test Library (VTL) are carried over from the WJ IV OL, the VTL is not intended as a direct replacement for the WJ IV OL. Instead, it serves as a flexible library of tests that can be used independently or in conjunction with the WJ IV COG and/or WJ ACH batteries to evaluate rapid automatized naming (RAN) and narrow aspects of the auditory processing (Ga), auditory working memory (Gwm), and retrieval fluency (Gr) broad CHC abilities. The notable features of the WJ V VTL, including tests and clusters that are new in the WJ V, are described below.

- **In-depth measurement of Rapid Automatized Naming (RAN).** The extant research literature has demonstrated a link between rapid automatized naming (RAN) and reading and math achievement and disorders (e.g., see Araújo et al., 2015; Kirby et al., 2010; Koponen et al., 2017; Norton & Wolf, 2012; Schneider & McGrew, 2018; Swanson et al., 2003). The author team developed four new rapid automatized naming (RAN) tests for the

WJ V VTL: Rapid Letter Naming, Rapid Phoneme Naming, Rapid Number Naming, and Rapid Quantity Naming. These tests are all mixed measures of the CHC broad abilities of processing speed (Gs) and retrieval fluency (Gr). Together with the Rapid Picture Naming test carried over from the WJ IV, these tests provide practitioners with measures that are sensitive to the early detection of reading and math disorders. These tests can be combined into a RAN–Reading cluster containing Rapid Phoneme Naming, Rapid Picture Naming, and Rapid Letter Naming, and a RAN–Math cluster containing Rapid Number Naming and Rapid Quantity Naming.

- **Measurement of CHC broad and narrow auditory processing (Ga) abilities.** As noted earlier, the measurement of Ga has moved from the WJ IV COG to the WJ V VTL and has undergone significant changes. The WJ IV Auditory Processing (Ga) cluster comprised the three-part Phonological Processing test and Nonword Repetition in the COG battery. From a statistical perspective, the WJ IV Phonological Processing test appeared to be a good measure of psychometric *g*; however, Schneider (2016) questioned whether the cognitive complexity of the Phonological Processing test actually represented more complex cognitive processing or whether, instead, it only appeared more cognitively complex in statistical analyses because of the mixed format of the three subtests. This criticism was echoed by post-WJ IV unpublished and published analyses (McGrew, 2023) of WJ IV tests and subtests. In the WJ V, the CHC broad ability Ga is measured by two clusters: Phonological Awareness, which includes Sound Blending and Segmentation; and Phonological Manipulation, which includes Sound Deletion and Sound Substitution. The Sound Deletion test is new in the WJ V; the Sound Substitution test is a full-length version of the WJ V COG Phonological Processing: Substitution (Ga) subtest.
- **Measurement of the auditory short-term storage (Gwm-Wa) narrow CHC ability.** The Auditory Memory Span cluster from the WJ IV OL battery remains intact in the WJ V and includes the Memory for Words and Sentence Repetition (formerly Memory for Sentences) tests. It has been moved into the WJ V VTL.
- **Measurement of phonemic retrieval fluency ability (Gr-FP), a proposed new narrow CHC ability.** Combining the WJ V COG Phonemic Word Retrieval test with the WJ V VTL Rapid Phoneme Naming Test yields the new Phonemic Retrieval Fluency (Gr-FP) cluster. Phonemic Word Retrieval appeared in the WJ IV COG as a subtest of the three-part Phonological Processing test.

Table 5 contains a description of each WJ V VTL test and its broad CHC classification.

Table 5.
WJ V VTL Tests, Broad CHC Classifications, and Tasks

VTL Tests	Broad CHC Classification	Description of Test Task
Nonsense Word Repetition	Ga, Gwm	The examinee hears a nonsense word presented from an audio recording and must say the word exactly as presented.
Rapid Picture Naming	Gs, Gr	The examinee has 1 minute to quickly name pictures of common objects presented in successive rows.
Animal-Number Sequencing	Gwm	The examinee hears an intermingled list of animals and digits presented from an audio recording and must then say the words, first naming all the animals in order and then naming all the numbers in order.
Sound Reversal	Ga	The examinee hears a simple word and must say the sounds in the word backward to form a different word.
Rapid Letter Naming	Gs, Gr	The examinee has 1 minute to quickly name single letters presented in successive rows.
Understanding Directions	Gwm, Gf	The examinee studies a detailed picture scene for 10 seconds and then follows prompts from an audio recording to tap elements of the picture in a prescribed order.

Table 5. *(cont.)*
WJ V VTL Tests, Broad CHC Classifications, and Tasks

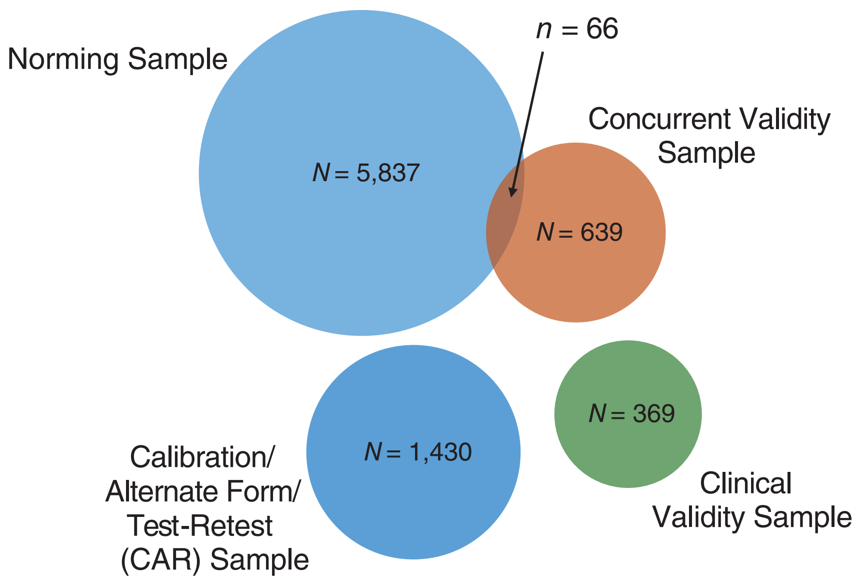
VTL Tests	Broad CHC Classification	Description of Test Task
Sound Blending	Ga	The examinee hears a series of syllables or phonemes presented from an audio recording and must blend the sounds together to say the whole word.
Rapid Phoneme Naming	Gs, Gr, Ga	The examinee has 1 minute to quickly pronounce phonemes (e.g., /m/) for single letters (i.e., graphemes) presented in successive rows.
Memory for Words	Gwm	The examinee hears a list of unrelated words presented from an audio recording and must say them in the same order.
Segmentation	Ga	The examinee hears a word and then says the word in parts; parts range from compound words to syllables to individual speech sounds (phonemes).
Rapid Number Naming	Gs, Gr	The examinee has 1 minute to quickly name single-digit numbers presented in successive rows.
Sentence Repetition	Gwm	The examinee hears a list of words, phrases, and sentences presented from an audio recording and must say each exactly as it was presented.
Sound Deletion[a]	Ga	The examinee must delete a word part or phoneme from a word presented from an audio recording and then say the new word.
Rapid Quantity Naming	Gs, Gr, Gv	The examinee has 1 minute to quickly say the number of dots in each pattern of 1 to 9 dots that are presented in successive rows.
Sound Substitution[a]	Ga	The examinee must replace part of a word with a new part presented from an audio recording and then say the new word.

Note. Italic font denotes *new* tests. For pragmatic reasons, some tests in this table may contain CHC ability classifications that do not appear in Table 1. This table is intended to contain more comprehensive and detailed information for test interpretation, whereas Table 1 is intended to be a user-friendly selective testing tool. Some of the differences in the broad CHC classifications listed in the two tables reflect differences in the factor loadings of the tests in three equally plausible CHC structural models, which are evaluated and discussed in detail in the WJ V Technical Manual (LaForte et al., 2025).
[a] In the WJ IV, this test was a subtest of Sound Awareness.

Research Samples

The data collection effort for the WJ V comprised four separate studies conducted concurrently between February 2022 and January 2024. The total number of research cases collected across the four studies was 8,209. Of these, 5,837 were norming cases, 1,430 were calibration/alternate-form/test-retest (CAR) cases, 639 were concurrent validity cases (66 of which overlapped with the norming study), and 369 were clinical validity cases. Figure 3 shows the four study samples and their respective *N* counts by data collection organization.

Figure 3.
WJ V Research Study Samples



Norming Study

Given the size of the WJ V battery (60 tests), it was neither practical nor feasible for norming study participants to be administered all the tests in the norming battery. Instead, a multiple matrix sampling (MMS) design was employed, where each examinee was administered a subset of the tests in the norming study. MMS methods are statistically sound procedures for gathering data within practical constraints imposed by examinee availability, time, and battery size (Jewsbury, et al., 2024; Mislevy et al., 1992; von Davier et al., 2009). In the WJ V norming study, each examinee was administered approximately one third of the tests in the battery for a total testing time of approximately 3 hours. Then, using the scores obtained from the administered tests, plausible scores for the unadministered tests were generated (imputed) for each examinee.[3]

Adequate selection and measurement of a norming sample is one of the more difficult, yet crucial, tasks in the development of a test. The *Standards for Educational and Psychological Testing* (AERA et al., 2014) state, "The validity of norm-referenced interpretations depends in part on the appropriateness of the reference group to which test scores are compared" (p. 97). The sampling plan for the WJ V norming study called for a total of 6,000 examinees across 24 sampling age groups: 1-year groups from ages 3 through 19, 10-year groups from ages 20 through 79, and a 20-year group from ages 80 and up, with a target *n* of 250 examinees per age group. The sampling plan was designed to be representative, within practical limits, of the U.S. population from ages 36 months to 90 years and older. The norming study sampling plan was stratified to control for census region, sex, ethnicity, race, and parent education level (for children) or examinee education level (for adults).

[3] A detailed discussion of the MMS and imputation methods utilized in the WJ V research design is beyond the scope of this document; however, interested readers may consult pages 139–150 of the WJ V Technical Manual (LaForte et al., 2025) for a description of the simulation study that was conducted to determine optimal sampling and imputation parameters to apply to the WJ V norming study.

During the 19-month period from February 2022 through August 2023, WJ V norming data were collected from 5,837 individuals:

- 562 children ages 3 through 5 who were not enrolled in K–12 school,
- 3,106 students ages 4 through 19 who were enrolled in kindergarten through 12th grade, and
- 2,169 individuals ages 17 and older who were no longer enrolled in high school.

Table 6 displays the distribution of the target and obtained WJ V norming sample by age and by grade. Although the obtained norming sample fell slightly short of 250 examinees for some age groups, the average number of cases per age group from ages 3 through 79 was 245. (The significantly lower *n* of 190 examinees for the 80+ age group is not unexpected given traditional challenges with recruiting older adults combined with the heightened health risks, mid- and post-pandemic, for social contact within this age group.) The higher density of examinees at ages 3 through 19 reflects the need to collect more concentrated data for ages when the abilities measured by the WJ V undergo the greatest rate of growth.

Table 6.
Distribution of the WJ V Norming Sample by Age and Grade

Age	*n*	Grade	*n*
2	2	Kindergarten	196
3	219	1	241
4	242	2	238
5	213	3	248
6	224	4	234
7	235	5	252
8	249	6	239
9	242	7	266
10	248	8	255
11	249	9	248
12	251	10	231
13	253	11	239
14	259	12	219
15	244		
16	233		
17	236		
18	242		
19	233		
20 to 29	298		
30 to 39	314		
40 to 49	285		
50 to 59	242		
60 to 69	224		
70 to 79	210		
80+	190		
Total *N*	5,837	**Total *N***	3,106

Note. Although the lowest age for WJ V norms is 48 months (4 years, 0 months), 3-year-old children were included in the norming study to provide information about the trajectory of the norm curves at age 4. Among the youngest children were two examinees who had not yet turned 3.

Table 7 contains the sampling variables and their distributions in the U.S. population (per the 2020 U.S. Census) and in the WJ V norming sample for preschool, K–12, and adult examinees. Following the U.S. Office of Management and Budget's (OMB) Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity (1997), Hispanic examinees of any race were counted as one category; each of the race alone, Not Hispanic groups of examinees were counted as separate categories; and the Other/Mixed, Not Hispanic examinees were counted as one category.

Table 7.
Distribution of Sampling Variables in the U.S. Population and in the WJ V Preschool, K–12, and Adult Norming Samples

	Preschool[a] Sample (n = 562)			K–12 Sample (n = 3,106)			Adult[b] Sample (n = 2,169)		
Sampling Variable	**Target Percentage**	**Number Obtained**	**Percentage in Norm Sample**	**Target Percentage**	**Number Obtained**	**Percentage in Norm Sample**	**Target Percentage**	**Number Obtained**	**Percentage in Norm Sample**
Census Region									
Northeast	15.6	103	18.3	15.7	423	13.6	17.5	352	16.2
Midwest	20.3	159	28.3	20.7	534	17.2	20.7	403	18.6
South	38.2	181	32.2	38.9	1,328	42.8	37.7	887	40.9
West	25.9	119	21.2	24.7	821	26.4	24.1	527	24.3
Sex									
Male	51.1	264	47.0	51.1	1,526	49.1	48.9	907	41.8
Female	48.9	297	52.8	48.9	1,558	50.2	51.1	1,248	57.5
Prefer not to say/Prefer to use another description	—	1	0.1	—	22	0.7	—	14	0.6
Race/Ethnicity[d]									
White, Not Hispanic	49.3	300	53.4	50.4	1,565	50.4	62.9	1,432	66.0
Black, Not Hispanic	14.0	84	14.9	13.6	452	14.6	12.1	270	12.4
AIANAT, Not Hispanic	0.8	—	—	0.8	10	0.3	0.7	17	0.8
Asian, Not Hispanic	5.3	9	1.6	5.2	91	2.9	6.0	88	4.1
NHPI, Not Hispanic	0.2	1	0.2	0.2	8	0.3	0.2	2	0.1
Other/Mixed, Not Hispanic	5.1	33	5.9	4.2	175	5.6	1.6	38	1.8
Hispanic, Any Race	25.3	131	23.3	25.5	786	25.3	16.5	306	14.1
Not Reported	—	4	0.7	—	19	0.6	—	16	0.7
Education[c, d]									
Less Than High School Diploma	10.2	37	6.6	11.6	240	7.7	9.8	96	4.4
High School Diploma or Equivalent	24.1	103	18.3	23.1	588	18.9	28.6	591	27.2
Some College or Vocational Training	27.7	162	28.8	29.0	932	30.0	28.0	744	34.3
4-Year College/University Degree	24.3	158	28.1	23.2	840	27.0	21.5	442	20.4
Advanced College Degree	13.7	102	18.1	13.1	482	15.5	12.1	296	13.6
Other/Not Reported	—	—	—	—	24	0.8	—	—	—

Notes. Target percentages were the 2017 National Population Projections obtained from *Projected Population by Single Year of Age, Sex, Race and Hispanic Origin for the United States: 2016 to 2060* (U.S. Census Bureau, 2018). AIANAT = American Indian or Alaska Native, NHPI = Native Hawaiian or Other Pacific Islander. Percentages may not sum to 100 due to rounding.

[a] *Preschool* is defined as a child who is age 6 or younger and is not yet enrolled in K–12 school.

[b] *Adult* is defined as an individual who is age 17 or older and is no longer enrolled in K–12 school.

[c] *Education* is the parent education level for Preschool and K–12 examinees and the examinee's own education level for adults.

[d] Although overall adult population and sample proportions are reported in this table, this variable was further stratified by age group for adults in the sampling plan. See the WJ V Technical Manual (LaForte et al., 2025) for population and obtained sample proportions by sampling age group.

Calibration/Alternate-Form/Test-Retest (CAR) Study

As the name implies, the WJ V CAR study had three purposes: (a) to collect item data for Rasch calibrating the item pools for the second forms of the nonspeeded tests; (b) to collect alternate-form reliability data for the second forms of the speeded/fluency tests; and (c) to collect test-retest data for the Semantic Word Retrieval, Phonemic Word Retrieval, and Letter Writing Fluency tests (which do not have unique second forms in the WJ V).

Examinees were recruited for the CAR study from among the overall WJ V participant survey respondents. If an examinee did not meet the demographic requirements for an open sampling cell in the norming study but met the age requirements for an open CAR study cell, they were assigned to the CAR study. The sampling plan for the CAR study included 70 examinees at each year of age for ages 3 and 4, 59 examinees at each year for ages 5 to 19, and 59 examinees at each 10 years of age from ages 20 to 80+. Table 8 contains the target and obtained *n*s for the study at each age. With the exception of a smaller obtained *n* for ages 80+, the sampling targets were generally met. Although the research design for the CAR study did not require the study sample to be matched to the U.S. Census proportions for race and ethnicity, the sample characteristics were closely monitored throughout the data collection phase of the study to ensure that the sample contained diversity with respect to race, ethnicity, geographic region, and education level.

Table 8.
Distribution of the WJ V CAR Study Sample by Age and Grade

Age	Target *n*	Obtained *n*
3	70	71
4	70	67
5	59	63
6	59	60
7	59	59
8	59	59
9	59	60
10	59	59
11	59	59
12	59	62
13	59	59
14	59	58
15	59	60
16	59	59
17	59	59
18	59	59
19	59	56
20 to 29	59	67
30 to 39	59	61
40 to 49	59	60
50 to 59	59	59
60 to 69	59	59
70 to 79	59	58
80+	59	37
Total *N*	**1,438**	**1,430**

Note. Although the lowest age for WJ V norms is 48 months (4y0m), 3-year-old children were included in the CAR study to ensure adequate data were collected on the easiest items on each test.

Table 9 contains the demographic characteristics for the CAR children and adult samples. Across all ages, the sample was racially and ethnically diverse (60% White, 20% Black, 7% other or two or more races, 12% Hispanic) with a relatively high overall education level (< 1% less than high school, 11% high school, and 88% greater than high school). The overall sample was heavily female (56%); however, females were only oversampled among the adult group.

Table 9. *Demographic Characteristics of the WJ V CAR Study Sample*

Sampling Variable	Children (n = 945)		Adults (n = 485)	
	Number Obtained	Percentage of CAR Study Sample	Number Obtained	Percentage of CAR Study Sample
Census Region				
Northeast	87	9.2	54	11.1
Midwest	352	37.2	111	22.9
South	241	25.5	197	40.6
West	265	28.0	123	25.4
Sex				
Male	469	49.6	149	30.7
Female	468	49.5	334	68.9
Prefer not to say/Prefer to use another description	8	0.8	2	0.4
Race/Ethnicity				
White, Not Hispanic	549	58.1	310	63.9
Black, Not Hispanic	182	19.3	99	20.4
AIANAT, Not Hispanic	1	0.1	3	0.6
Asian, Not Hispanic	23	2.4	10	2.1
NHPI, Not Hispanic	0	0.0	0	0.0
Other/Mixed, Not Hispanic	53	5.6	13	2.7
Hispanic, Any Race	131	13.9	44	9.1
Not Reported	6	0.6	6	1.2
Education[a]				
Less Than High School Diploma	7	0.7	5	1.0
High School Diploma or Equivalent	49	5.2	102	21.0
Greater Than High School Diploma	886	93.8	378	77.9
Not Reported	3	0.3	—	—

Notes. The Children group includes examinees ages 3 to 5 who are not yet enrolled in school and examinees of any age who are enrolled in Grades K–12. The Adults group includes examinees who are age 14 or older and are no longer enrolled in Grades K–12. AIANAT = American Indian or Alaska Native; NHPI = Native Hawaiian or Other Pacific Islander. Percentages may not sum to 100 due to rounding.

[a] *Education* is the parent education level for children and the examinee's own education level for adults.

Concurrent Validity Study

The purpose of the concurrent validity study was to obtain correlations between the WJ V cluster scores and composite scores from other commercially available batteries that purport to measure similar constructs. The study comprised 12 separate samples of examinees who each took a set of tests from the WJ V and a set of subtests from one or more other cognitive, achievement, or oral language test batteries, including:

- *Wechsler Adult Intelligence Scale®–Fourth Edition* (WAIS®-IV; Wechsler, 2008);
- *Wechsler Intelligence Scale for Children®–Fifth Edition* (WISC®-V; Wechsler, 2014);
- *Wechsler Preschool and Primary Scale of Intelligence™–Fourth Edition* (WPPSI™-IV; Wechsler, 2012);

- *Kaufman Assessment Battery for Children–Second Edition Normative Update* (KABC™-II NU; Kaufman & Kaufman, 2004; Kaufman et al., 2018);
- *Reynolds Intellectual Assessment Scales™–Second Edition* (RIAS-2™, Reynolds & Kamphaus, 2015);
- *Comprehensive Test of Phonological Processing, Second Edition* (CTOPP-2; Wagner et al., 2013);
- *The Rapid Automatized Naming and Rapid Alternating Stimulus Tests* (RAN/RAS; Wolf & Denckla, 2005);
- *Mini-Mental State Examination–Second Edition*™ (MMSE-2®; Folstein et al., 2010);
- *Dementia Rating Scale–Second Edition*® (DRS-2™; Jurica et al., 2001);
- *Kaufman Tests of Educational Achievement, Third Edition* (KTEA™-3; Kaufman & Kaufman, 2014); and
- *Wechsler Individual Achievement Test*®*, Fourth Edition* (WIAT®-4; NCS Pearson, 2020).

As in the norming study, it was not practical to administer all WJ V tests to the examinees in the concurrent validity study; instead, examinees were administered only the WJ V tests that comprised the relevant cluster scores within each study sample. These tests were always administered first. Then, the subtests from the other battery (or batteries) were administered within a time period ranging from the same day to no longer than 21 days later. Total testing times for the concurrent validity study samples, including both the WJ V tests and the other-battery tests, ranged from 3 hours, 20 minutes to 5 hours, 50 minutes.

The sampling targets for the concurrent validity study included 100 examinees each for the WAIS-IV and WISC-V and 50 examinees each for the other external batteries. Unlike the WJ V norming study, the sampling plan for the concurrent validity study did not include specific targets for geographic region, race, or ethnicity. Generally, the obtained counts for all concurrent validity samples were within $n = \pm 3$ of targets, except for the RAN/RAS sample, which fell short of the target ($n = 41$). The demographic characteristics of the sample were continuously monitored throughout the study. Although the overall sample does not align exactly with U.S. Census proportions, it contains diversity with respect to race (70% White, 12% Black, and 18% other or multiple races), ethnicity (19% Hispanic), and education level (or parent education level, for children; 6% less than high school, 22% high school, and 72% greater than high school). Table 10 presents the sample size and demographic information for all 12 concurrent validity studies, including age characteristics (i.e., range, mean, and standard deviation) in years and percentages for sex, race, ethnicity, and adult examinee or reference parent education level.

Table 10.
Demographic Characteristics of the Concurrent Validity Study Samples

	External Cognitive Measures								External Achievement Measures			
	WAIS-IV	WISC-V	WPPSI-IV	KABC-II NU	RIAS-2	CTOPP-2	RAN/RAS[b]	MMSE-2 and DRS-2	KTEA -3 Age 8–12	KTEA-3 Age 13–18	WIAT-4 Grades 1–8	WIAT-4 Grades 9–12
Chapter 6 Table Number	6-19	6-20	6-21	6-22	6-23	6-24	6-25	6-26	6-27	6-28	6-29	6-30
N	100	97	47	49	51	47	41	50	49	49	51	49
Age (Years)												
Range	16–80	6–17	4–7	7–18	6–16	7–23	7–21	65–89	8–13	13–19	6–14	13–18
Mean	37.4	11.1	5.4	12.6	11.4	13.5	12.5	71.1	10.0	15.5	10.3	15.8
SD	21.5	3.1	1.1	3.5	3.2	4.3	3.6	5.5	1.5	1.8	2.4	1.3

Table 10. *(cont.)* *Demographic Characteristics of the Concurrent Validity Study Samples*

	External Cognitive Measures								External Achievement Measures			
	WAIS-IV	WISC-V	WPPSI-IV	KABC-II NU	RIAS-2	CTOPP-2	RAN/RAS[b]	MMSE-2 and DRS-2	KTEA -3 Age 8–12	KTEA-3 Age 13–18	WIAT-4 Grades 1–8	WIAT-4 Grades 9–12
Chapter 6 Table Number	6-19	6-20	6-21	6-22	6-23	6-24	6-25	6-26	6-27	6-28	6-29	6-30
Sex												
Male	33.3	51.0	42.6	61.2	51.0	41.3	45.0	26.0	51.0	37.5	37.3	52.1
Female	66.7	49.0	57.4	38.8	49.0	58.7	55.0	74.0	49.0	62.5	62.7	47.9
Race												
White	90.0	45.4	74.5	69.4	72.5	68.1	65.9	94.0	59.2	53.1	72.5	83.7
Black	4.0	26.8	8.5	18.4	13.7	10.6	9.8	—	12.2	16.3	5.9	4.1
AIANAT	—	3.1	2.1	—	—	—	—	—	—	—	—	2.0
Asian	—	2.1	—	2.0	—	4.3	4.9	—	4.1	4.1	5.9	6.1
Other/Mixed	6.0	22.7	14.9	10.2	13.7	17.0	19.5	6.0	24.5	26.5	15.7	4.1
Ethnicity												
Not Hispanic	88.0	68.8	89.1	53.1	82.4	87.2	85.4	96.0	72.3	81.6	86.0	87.8
Hispanic	12.0	31.3	10.9	46.9	17.6	12.8	14.6	4.0	27.7	18.4	14.0	12.2
Education[a]												
< HS Graduate	1.0	10.3	8.5	8.2	2.0	4.3	4.9	—	4.1	8.2	5.9	4.1
HS Graduate	27.0	19.6	19.1	12.2	23.5	14.9	12.2	26.0	14.3	42.9	19.6	16.3
> HS	71.0	70.1	72.3	77.6	74.5	78.7	80.5	74.0	81.6	46.9	74.5	79.6
Not Reported	1.0	—	—	2.0	—	2.1	2.4	—	—	2.0	—	—

Note. AIANAT = American Indian or Alaska Native.
[a] Education is *parent* education level for children and *examinee* education level for adults.
[b] The RAN/RAS study sample is a subset of the CTOPP-2 study sample.

Clinical Validity Studies

According to the *Standards for Educational and Psychological Testing* (AERA et al., 2014), "Categorical variables, including group membership variables, become relevant when the theory underlying a proposed test use suggests that group differences should be present or absent if a proposed test score interpretation is to be supported" (p. 16). To investigate the relationship between WJ V scores and group membership status, selected tests were administered to individuals with the following eight clinical diagnoses: gifted, intellectual disability (ID), specific learning disabilities (SLD; reading, writing, and math), language impairment, attention-deficit/hyperactivity disorder (ADHD), and autism spectrum disorder (ASD).

The WJ V clinical validity study participants were drawn from a variety of educational and clinical settings. The study inclusion criteria are summarized in Table 11. Although study participants had to meet the minimum criteria for inclusion, given the variety of educational and clinical settings from which study participants were drawn, the criteria used for each participant's original diagnosis or classification likely varies within each study group. Furthermore, the person providing the inclusion information was the examinee's parent or guardian—not the school special education staff. Because the clinical validity study participants were not randomly selected from the populations of each respective diagnostic group, the sample results presented in this section should not be considered precise statistical representations of each diagnostic group.

Table 11.
Inclusion Criteria for WJ V Clinical Validity Study Groups

Clinical Validity Group	Age Range	Inclusion Criteria
Gifted[a]	6–12 years	• Currently participating in high ability/gifted and talented school curriculum • Currently receiving gifted services
Intellectual Disability (ID)	6–12 years	• Documented intellectual disability • Intellectual disability must be the primary diagnosis • Currently receiving special education services
Specific Learning Disability (SLD)–Reading	6–18 years	• Documented learning disability in reading • SLD in reading must be the primary diagnosis or eligibility category. • Currently receiving special education services under SLD category
Specific Learning Disability (SLD)–Writing	6–18 years	• Documented learning disability in writing • SLD in writing must be the primary diagnosis or eligibility category. • Currently receiving special education services under SLD category
Specific Learning Disability (SLD)–Math	6–18 years	• Documented learning disability in math • SLD in math must be the primary diagnosis or eligibility category. • Currently receiving special education services under SLD category
Language Impairment	6–18 years	• Documented language impairment • Language impairment must be the primary diagnosis. • Currently receiving special education services
Attention-Deficit/ Hyperactivity Disorder (ADHD)	7–12 years	• Documented ADHD[b] • ADHD must be the primary diagnosis. • Currently receiving special education services and/or 504 plan accommodations
Autism Spectrum Disorder (ASD)	6–18 years	• Documented autism spectrum disorder[c] • ASD must be the primary diagnosis. • Currently receiving special education services

[a] For the Gifted study, examiners documented the specific qualifying criteria for the examinee's school on the Parent Consent form.
[b] For the Attention-Deficit/Hyperactivity Disorder (ADHD) study, a diagnosis of attention deficit disorder (ADD) was not sufficient for inclusion. The examinee had to have a diagnosis of ADHD.
[c] The Autism Spectrum Disorder (ASD) group excluded individuals with a diagnosis of Level 1 ASD.

Table 12 presents sample size and demographic information for all clinical validity groups. In addition to sample size, the age characteristics (range in years, mean, and *SD*) and percentages for sex, race, ethnicity, and parent education are summarized.

Table 12.
Demographic Characteristics of the WJ IV Clinical Validity Study Groups

	Gifted	ID	SLD–Reading	SLD–Writing	SLD–Math	Language Impairment	ADHD	ASD
N	85	20	89	15	37	24	50	49
Age (Years)								
Range	6–12	6–12	6–17	7–17	7–18	6–17	7–12	6–16
Mean	9.2	8.9	11.8	11.7	12.9	10.1	9.7	10.7
SD	1.6	1.9	2.8	3.4	3.4	2.9	1.6	2.9

Table 12. *(cont.)*
Demographic Characteristics of the WJ IV Clinical Validity Study Groups

	Gifted	ID	SLD–Reading	SLD–Writing	SLD–Math	Language Impairment	ADHD	ASD
Sex								
Male	54.1	50.0	40.4	53.3	35.1	58.3	66.0	73.5
Female	45.9	50.0	57.3	46.7	59.5	37.5	34.0	26.5
Prefer not to say	—	—	—	—	2.7	—	—	—
Prefer to use another description	—	—	2.2	—	2.7	4.2	—	—
Race								
White	80.0	70.0	75.3	66.7	64.9	87.5	78.0	73.5
Black	10.6	15.0	10.1	13.3	13.5	8.3	14.0	12.2
AIANAT	—	—	—	—	—	—	—	—
Asian	3.5	—	2.2	—	—	—	—	2.0
Other/Mixed	5.9	15.0	7.9	13.3	10.8	4.2	8.0	4.1
Prefer to describe as	—	—	2.2	—	5.4	—	—	6.1
Prefer not to say	—	—	2.2	6.7	5.4	—	—	2.0
Ethnicity								
Not Hispanic	88.2	75.0	78.7	73.3	64.9	91.7	82.0	75.5
Hispanic	10.6	25.0	19.1	26.7	32.4	8.3	18.0	22.4
Prefer not to say	1.2	—	2.2	—	2.7	—	—	2.0
Parent Education								
< HS Graduate	2.4	10.0	9.0	—	13.5	4.2	4.0	6.1
HS Graduate	—	40.0	16.9	26.7	48.6	25.0	12.0	28.6
> HS	97.6	45.0	74.2	73.3	37.8	70.8	82.0	65.3
No information provided	—	5.0	—	—	—	—	2.0	—

Note. ID = intellectual disability, SLD = specific learning disability, ADHD = attention-deficit/hyperactivity disorder, ASD = Autism spectrum disorder, AIANAT = American Indian or Alaska Native

Norming Procedures

Using the complete *W*-score matrix obtained from the MMS and imputation processes for the 5,837 norming study participants, cluster *W* scores were computed as the arithmetic mean of the contributing test *W* scores. Data from the norming study participants were then 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, or adult). For each demographic variable, if an examinee belonged to a category of the variable that was overrepresented in the WJ V 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 V norming study sample, the examinee's partial weight for that variable was greater than 1.00. 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.

Bootstrap resampling procedures (Efron & Tibshirani, 1993), first implemented and described for the *Woodcock-Johnson III Normative Update* (McGrew, Dailey, & Schrank, 2007; Woodcock et al., 2001, 2007), were used to construct the WJ V norms. In the bootstrap resampling procedures, examinee weights were used as sampling probabilities. 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 norms. When compared to more traditional norm development procedures, such as those used in most other individually administered cognitive, language, and achievement batteries, the bootstrap-based procedures used to construct the WJ V norms produce better estimates of an examinee's ability in relation to peers.

The calculation of derived scores requires the establishment of the "normative" (median) score for each test or measure for individuals at each specific age (for age norms) or grade (for grade norms) where normative interpretations are intended. In the WJ family of tests, this normative score is called the Reference-*W* (REF *W*) score. When plotted as a function of chronological age (or grade), the REF *W* scores assume the characteristics of developmental growth curves. The test and cluster REF *W* curves are visual-graphic representations of the average performance of norming study participants at every age (or grade) for the effective use of the specific measure. The REF *W* curves serve as the foundation for the age- and grade-equivalent scores, relative proficiency index (RPI), and instructional range interpretation features in the WJ V. In addition, when the standard deviations (*SD*s) of the scores at each age are plotted as a function of age or grade, the resultant curves represent the *SD* values that, when combined with the REF *W* values, provide the foundation for the calculation of all other norm-referenced score metrics (e.g., standard scores and percentile ranks).

The published norm tables for WJ V tests and clusters extend from age 4 years, 0 months up to age 100 years, 0 months,[4] although norming data were collected from examinees as young as age 3 years, 0 months to inform the trajectory of the norm curves at age 4 years. The development of the age and grade norms for tests and clusters is described in the following sections.

Age-Based Norms for Tests

For each test, age-based norms were constructed in an 8-step process. This process is described here.

Step 1: Extract examinee information and W-Ability scores from the complete data matrix. For each test score in the imputed *W*-Ability score data matrix, the examinee identifier (ID), examinee age in months, overall examinee weight, and *W*-Ability score were extracted. The *W*-Ability scores generally increase rapidly with age during childhood, then stabilize from middle adulthood on, forming a characteristic "growth curve."

Step 2: Bootstrap resample the W-Ability score data set. Using the examinee weights as sampling probabilities, the examinee *W*-Ability score data set was bootstrap-resampled to produce 251 data sets, each similar in size to the original data set. The resampling procedure was designed to guarantee that every examinee appeared at least once among the resampled data sets, no matter the magnitude of the examinee's weight.

Step 3: Create overlapping blocks of examinees. Each resampled data set was sorted by examinee age in months, then divided into blocks of 250 examinee records each; block starting indices were offset by 50 examinees so that examinees 1 to 250 comprised the first block, examinees 51 to 300 comprised the second block, and so on.

[4] The only exception to this age range is for the Letter Writing Fluency test, which has norms available from age 4 years, 0 months to age 9 years, 11 months.

Step 4: Calculate summary statistics for each block of examinee scores. For each block of 250 examinees, the following summary statistics were calculated:

- the median chronological age in months,
- the median *W*-Ability score within the block (REF *W*),
- the scaled difference between the 10th and 50th percentile values (SDLo), and
- the scaled difference between the 90th and 50th percentile values (SDUp).

The calculation of the SDLo and SDUp allow the modeling of ability within the block as a pair of half-normal distributions.

Step 5: Create "candidate" norm tables. For each test, the REF *W* values and median ages from each of the approximately 110 examinee blocks within each of the 251 bootstrap resamples were plotted and connected with lines, resulting in a "thread cloud" of data points. A piecewise linear fit was made to the data points from all threads, and a smoothing spline was applied to that piecewise linear fit. This smoothing spline became the initial "candidate" for the REF *W* column of the norm table. This plotting, fitting, and smoothing process was then repeated for both the SDLo and SDUp statistics to generate the initial candidates for those columns of the norm tables.

Step 6: Generate and evaluate "candidate" standard scores. After initial candidates were generated for all three columns of the norm table (REF *W*, SDLo, and SDUp) for each test, these values were used to calculate *W*-Difference scores and standard scores for all examinees in the norming sample on all tests. The WJ V Research and author team then inspected the candidate REF *W*, SDLo, and SDUp tables and the resulting scaled scores to determine if they met the following objectives:

- The REF *W* curve reflects the desired properties of a growth curve for the particular ability measured by the test (i.e., it conforms to theoretical expectations about the growth and decline of the ability across the age range).
- The REF *W* values continually increase to a single local maximum and then decrease across the remaining ages.
- The SDLo and SDUp curves reflect ability variations that change slowly or rapidly at different parts of the age range.
- The calculated standard scores for the norming sample adhere to expected trends, where the median standard score is reasonably close to 100, and the 10th and 90th percentiles of the standard-score distribution are reasonably close to the standard-normal values of 81 and 119, respectively.
- A plot of the standard score values associated with a specific *W*-Ability score behaves as expected, without undesirable "inversions."

Step 7: Apply modifications to REF W, SDLo, and SDUp curves as needed. When necessary, any issues identified in Step 6 were resolved by applying modifications to the candidate curves. Decisions about modifications required a balance between (a) obtaining the desired properties for the curves and (b) obtaining the desired sets of standard scores in the norming sample. When these goals were in conflict, decisions were made in favor of the desired properties of the curves, because any norming sample contains an imperfect representation of the population parameters. The new, modified candidate was then evaluated against the same criteria and more modifications were implemented if necessary. In this way, the team iterated on the REF *W*, SDLo, and SDUp curves until they were satisfied that the curves, and the resulting standard scores, met the criteria in Step 6.

Step 8: Create the final norm table. Once the team agreed that no additional modifications were needed, the final candidate REF *W*, SDLo, and SDUp values were converted into a norm table for each test.

Age-Based Norms for Clusters

Cluster-level *W*-Ability scores for each examinee were calculated from the final imputed test-level *W*-Ability score matrix. When a test score was missing for an examinee after imputation (due to trimming or filtering), cluster score(s) requiring that test were not computed, and the cluster score(s) was left missing in the final cluster *W*-score matrix.

The REF *W* column of the norm table for each cluster was constructed by averaging the REF *W* values for the cluster's component tests at each month of age. The SDLo and SDUp columns for clusters were constructed using nearly the same process as was used for the SDLo and SDUp columns for the tests:

- *W*-Ability scores for each cluster were bootstrap resampled using examinee weights as sampling probabilities, generating 251 resamples as was done for the test-level norm tables.
- Within each resample, examinees were ordered by age in months and divided into blocks of 250 examinees each, with each block overlapping by 50 examinees.
- For each cluster, summary statistics were calculated for each of the overlapping blocks of examinees within each resample, providing the median age in months and *SD*s for the lower (SDLo) and upper (SDUp) halves of the score distributions. The values of SDLo and SDUp were then plotted as a function of age across all 251 resamples. (Median ability values were not calculated within the blocks because the cluster REF *W* values are derived directly from the REF *W* values for the component tests.)
- The thread clouds for SDLo and SDUp were each fit with a piecewise linear model. A smoothing spline was applied to generate an initial candidate curve.
- The candidate norm table, comprised of the REF *W* values calculated from the component tests and the cluster-specific candidate values for SDLo and SDUp, was then used to calculate standard scores for all examinees in the norming sample.
- The scaled scores were evaluated, and candidates were modified until the properties of both the curves and the standard score distributions met the same criteria as described above for tests.

Grade-Based Norms for Tests and Clusters

The linear relationship between age (in months) and grade (in tenths of the school year) for the school-age examinees in the norming study was derived using the ages and grade placements from the examinees in the norming sample who were enrolled in K–12 schooling at the time of testing. The grade-based REF *W* values for tests were calculated by applying this age-to-grade relationship to the age-based REF *W* values. The grade-based REF *W*s for clusters were calculated as the grade-by-grade average of the REF *W*s for the component tests. For grade-based test norms and grade-based cluster norms, the SDLo and SDUp values were calculated in the same way as the age-based SDLo and SDUp values for clusters, by starting with the examinees' grade placements and *W*-Ability scores, resampling the examinees, dividing them into overlapping blocks, and fitting curves through summary statistics from the overlapping blocks.

Comparison Norms

A *comparison score* describes the difference between an examinee's performance on a target measure and that same examinee's performance on a predictor measure, within the context of the difference between the target and predictor performance that is observed for the examinee's age or grade peers (i.e., the "difference score"). The WJ V comparison procedures express the examinee's difference score as a percentile rank (i.e., base rate), showing the relative standing of the examinee's difference score among the distribution of difference scores earned by the examinee's age or grade peers. While it is mathematically possible to perform comparisons between predicted and obtained scores for test batteries that are normed separately (using established correlations between scores from separate test batteries), the comparison score norms in the WJ V were developed using standard scores from the same individuals on the predictor and target measures. These data-based comparisons are possible on the WJ V because the scaled scores for the predictor and target measures were "conormed," or developed on the same norming sample. Conorming is a long-standing feature of the Woodcock-Johnson family of tests.

The difference score percentile rank is a transformation of the z score, which takes the general form of

$$z_{DIFF} = \frac{(O_{DIFF(SS)} - E_{DIFF(SS)})}{SD_{DIFF(SS)}}, \tag{1}$$

where $O_{DIFF(SS)}$ is the observed difference of standard scores, $E_{DIFF(SS)}$ is the expected difference, and $SD_{DIFF(SS)}$ is some measure of variation of the difference score within the population. Examinees are expected to perform as predicted ($E_{DIFF(SS)} = 0$). Because comparison scores are based on the differences of standard scores, and standard scores are calculated from norm table values that have already incorporated examinee weights, examinee weights were not applied in the development of the comparison score norms.

The WJ V battery provides 144 comparisons within two major types of difference scores (intra-ability and ability/achievement) across two norm bases (age and grade). Predictors can be clusters or pools (groups of tests); targets are always clusters. The procedure for developing the comparison norms was the same for each of the 144 comparisons. First, a prediction equation was developed using a piecewise linear model to predict an examinee's target cluster standard score from their age (or grade) and standard score on the predictor measure. Then, norm tables were developed that allow the evaluation of the magnitude of the difference in the context of the difference scores from the examinee's age (or grade) peers.

Reliability

The *Standards for Educational and Psychological Testing* (AERA et al., 2014) specify that test developers should report reliability information for each score—both individual test scores and composite scores—that will be interpreted by test users. Additionally, test developers should present reliability indices to support all the proposed uses and interpretations of the test scores. In accordance with the *Standards*, several different types of reliability coefficients were computed for the WJ V tests and clusters. Additional reliability evidence supports the equivalence of the first and second forms of tests and the stability of scores from first and second administrations of tests.

Marginal Test Reliability

The standard-score variance and standard-score standard error of measurement (*SEM*) associated with the number-correct score earned by examinees in the WJ V norming sample were inserted into the following equation to obtain a marginal reliability coefficient (r_{11}) for each test in six different age groups:

$$r_{11} = 1 - \frac{SEM^2}{SD^2_{observed}}. \tag{2}$$

There are two benefits of using standard scores ($M = 100$, $SD = 15$) to compute and interpret reliability coefficients. First, most users rely on standard scores for the interpretation of an examinee's performance on the WJ V. Second, as the name implies, standard scores are standardized to a common metric, where the expected standard deviation for any age-group subset drawn from the norming sample is 15. By holding the value of $SD^2_{observed}$ in Equation 2 relatively constant, differences in reliability coefficients between age groups on any test can be interpreted as being mostly due to differences in the average measurement precision (i.e., *SEM*) for examinees in that age group, rather than simply as an artifact of *W*-score range restriction. The marginal reliability coefficients for all WJ V tests (except the RAN tests[5]) are reported in Table 13 across six age groups. Reliability is reported in Table 13 for only the *recommended* administration age range for each test.

Table 13.
Marginal IRT Reliability for WJ V Tests

Test	Location in WJ V	Age 4–5		6–9		10–14		15–19		20–49		50–80+		Median r_{11}
		n	r_{11}	*n*	r_{11}	*n*	r_{11}	*n*	r_{11}	*n*	r_{11}	*n*	r_{11}	
Oral Vocabulary	COG	203[a]	0.79	948	0.86	1,256	0.87	1,198	0.87	905	0.89	866	0.88	0.87
Matrices	COG	—	—	949	0.82	1,256	0.85	1,198	0.85	905	0.85	866	0.87	0.85
Spatial Relations	COG	444	0.84	949	0.83	1,256	0.86	1,198	0.85	905	0.84	866	0.85	0.84
Story Recall	COG	443	0.80	949	0.90	1,256	0.91	1,198	0.93	905	0.93	866	0.92	0.92
Semantic Word Retrieval	COG	444	0.77	949	0.88	1,256	0.90	1,198	0.92	905	0.92	866	0.91	0.91
Verbal Attention	COG	438	0.76	949	0.80	1,256	0.78	1,198	0.80	905	0.80	866	0.80	0.80
Number-Pattern Matching	COG	394	0.92	943	0.97	1,255	0.97	1,198	0.97	905	0.97	866	0.97	0.97
Verbal Analogies	COG	443	0.75	949	0.84	1,256	0.83	1,198	0.84	905	0.85	866	0.86	0.84
Analysis-Synthesis	COG	196[a]	0.87	942	0.88	1,256	0.86	1,198	0.86	905	0.86	866	0.89	0.87
Block Rotation	COG	444	0.79	949	0.82	1,256	0.83	1,198	0.84	905	0.84	866	0.83	0.83
Story Comprehension	COG/ACH	—	—	949	0.88	1,256	0.89	1,198	0.88	905	0.89	866	0.87	0.88
Phonemic Word Retrieval	COG	212[a]	0.68	949	0.85	1,256	0.91	1,198	0.92	905	0.94	866	0.94	0.92
Numbers Reversed	COG	202[a]	0.77	947	0.84	1,256	0.88	1,198	0.90	905	0.90	866	0.90	0.89
Letter-Pattern Matching	COG	418	0.92	947	0.96	1,256	0.96	1,198	0.96	905	0.96	866	0.96	0.96
General Information	COG	208[a]	0.84	949	0.83	1,256	0.82	1,198	0.84	905	0.87	866	0.84	0.84
Concept Formation	COG	—	—	949	0.90	1,256	0.90	1,198	0.87	905	0.89	866	0.92	0.90
Number Series	COG	435	0.85	949	0.90	1,256	0.91	1,198	0.92	905	0.91	866	0.91	0.91
Visual-Auditory Learning	COG	441	0.90	949	0.93	1,256	0.91	1,198	0.89	905	0.91	866	0.94	0.91
Visual Working Memory	COG	—	—	949	0.83	1,256	0.78	1,198	0.76	905	0.76	863	0.84	0.78
Symbol Inhibition	COG	—	—	721[b]	0.87	1,255	0.92	1,198	0.93	905	0.93	864	0.92	0.92
Picture Vocabulary	ACH	444	0.79	949	0.81	1,256	0.84	1,198	0.85	905	0.87	866	0.85	0.84
Letter-Word Identification	ACH	442	0.95	949	0.97	1,256	0.93	1,198	0.92	905	0.91	866	0.90	0.93
Calculation	ACH	—	—	946	0.91	1,256	0.93	1,198	0.93	905	0.93	866	0.92	0.93
Spelling	ACH	205[a]	0.89	948	0.96	1,256	0.94	1,198	0.93	905	0.91	866	0.91	0.92
Oral Comprehension	ACH	444	0.76	949	0.78	1,256	0.77	1,198	0.75	905	0.75	866	0.69	0.76
Word Attack	ACH	209[a]	0.80	949	0.94	1,256	0.91	1,198	0.87	905	0.85	866	0.88	0.88
Math Facts Fluency	ACH	—	—	941	0.90	1,255	0.97	1,198	0.96	905	0.95	866	0.95	0.95
Sentence Writing Accuracy	ACH	—	—	944	0.96	1,256	0.96	1,198	0.93	905	0.94	866	0.93	0.94
Passage Comprehension	ACH	212[a]	0.87	949	0.90	1,256	0.86	1,198	0.85	905	0.87	866	0.86	0.87
Applied Problems	ACH	444	0.83	949	0.85	1,256	0.88	1,198	0.89	905	0.88	866	0.87	0.88
Sentence Reading Fluency	ACH	161[a]	0.88	920	0.96	1,256	0.98	1,198	0.98	905	0.98	862	0.98	0.98
Written Language Samples	ACH	200[a]	0.77	946	0.89	1,256	0.91	1,198	0.91	905	0.92	866	0.91	0.91
Oral Language Samples	ACH	444	0.79	949	0.85	1,256	0.86	1,198	0.82	905	0.80	866	0.79	0.81
Sentence Writing Fluency	ACH	—	—	694[b]	0.90	1,252	0.93	1,194	0.94	904	0.94	865	0.94	0.94
Paragraph Reading Comprehension	ACH	—	—	948	0.94	1,256	0.90	1,198	0.84	905	0.84	866	0.86	0.86
Word Reading Fluency	ACH	—	—	912	0.95	1,255	0.97	1,198	0.97	905	0.98	865	0.97	0.97
Math Problem Identification	ACH	212[a]	0.79	948	0.88	1,256	0.89	1,198	0.91	905	0.89	866	0.89	0.89
Magnitude Comparison	ACH	422	0.94	949	0.97	1,256	0.97	1,198	0.98	905	0.97	865	0.97	0.97
Number Sense	ACH	212[a]	0.76	949	0.85	1,256	0.85	1,198	0.85	905	0.84	866	0.86	0.85
Spelling of Sounds	ACH	—	—	946	0.83	1,256	0.85	1,198	0.81	905	0.82	866	0.86	0.83

[5] Due to the rate-based scoring model underlying the RAN tests, marginal reliability coefficients would be uninterpretable. Instead, only test-retest reliability coefficients are reported for the RAN tests (see the "Test-Retest Reliability" section).

Table 13. *(cont.)*
Marginal IRT Reliability for WJ V Tests

Test	Location in WJ V	Age 4–5 *n*	Age 4–5 r_{11}	Age 6–9 *n*	Age 6–9 r_{11}	Age 10–14 *n*	Age 10–14 r_{11}	Age 15–19 *n*	Age 15–19 r_{11}	Age 20–49 *n*	Age 20–49 r_{11}	Age 50–80+ *n*	Age 50–80+ r_{11}	Median r_{11}
Oral Reading	ACH	—	—	938	0.93	1,256	0.89	1,198	0.83	905	0.80	866	0.80	0.83
Reading Recall	ACH	—	—	943	0.87	1,256	0.93	1,198	0.93	905	0.93	866	0.94	0.93
Academic Vocabulary	ACH	212[a]	0.66	949	0.83	1,256	0.89	1,198	0.91	905	0.91	866	0.91	0.90
Academic Facts	ACH	212[a]	0.79	949	0.80	1,256	0.85	1,198	0.87	905	0.88	866	0.87	0.86
Letter Writing Fluency	ACH	400	0.88	947	0.95	—	—	—	—	—	—	—	—	0.91
Nonsense Word Repetition	VTL	444	0.87	949	0.84	1,256	0.82	1,198	0.81	905	0.82	866	0.85	0.83
Animal-Number Sequencing	VTL	210[a]	0.80	948	0.83	1,256	0.83	1,198	0.85	905	0.86	866	0.86	0.84
Sound Reversal	VTL	—	—	710[b]	0.81	1,255	0.88	1,198	0.87	905	0.86	863	0.88	0.87
Understanding Directions	VTL	444	0.89	949	0.88	1,256	0.79	1,198	0.65	905	0.68	866	0.77	0.78
Sound Blending	VTL	442	0.84	949	0.87	1,256	0.86	1,198	0.85	905	0.87	866	0.89	0.87
Memory for Words	VTL	442	0.86	949	0.84	1,256	0.82	1,198	0.80	905	0.79	866	0.81	0.82
Segmentation	VTL	439	0.88	949	0.92	1,256	0.90	1,198	0.88	905	0.90	866	0.91	0.90
Sentence Repetition	VTL	444	0.89	949	0.84	1,256	0.79	1,198	0.78	905	0.79	866	0.78	0.79
Sound Deletion	VTL	212[a]	0.87	949	0.86	498[c]	0.72	—	—	—	—	—	—	0.86
Sound Substitution	VTL	—	—	941	0.87	751[d]	0.78	—	—	—	—	—	—	0.83

[a] Reliability based on age 5 only.
[b] Reliability based on ages 7 to 9 only.
[c] Reliability based on ages 10 and 11 only.
[d] Reliability based on ages 10 to 12 only.

Cluster Reliabilities

Reliabilities for all the WJ V cluster scores except RAN–Reading and RAN–Math were calculated using Mosier's (1943) equation for the reliability of an equally weighted composite:

$$r_{cc} = 1 - \frac{\sum SD_j^2 - \sum SD_j^2 r_{jj}}{\sum SD_j^2 + 2\sum SD_j\, SD_k r_{jk}}, \tag{3}$$

where r_{cc} is the reliability of the cluster, SD_j is the standard deviation of examinee standard scores on test j, SD_k is the standard deviation of examinee ability scores on test k, r_{jj} is the reliability of test j, and r_{jk} is the correlation between test j and k (test weights have been canceled out in Equation 3). For all clusters except RAN–Reading and RAN–Math, the marginal test reliability coefficients from Table 13 were used to calculate the cluster reliabilities. For the RAN–Reading and RAN–Math clusters, the reliabilities are the correlations between the cluster *W* scores computed from the first and second administrations of the RAN tests in the CAR study. Table 14 contains the composite reliabilities for all non-RAN clusters for six age groups, computed from the WJ V norming sample. Table 15 contains the reliabilities for the RAN–Reading and RAN–Math clusters, computed from the CAR study examinees. In both of these tables, reliability is reported for only the *recommended* administration age range for each cluster.

Table 14.
Composite Reliability for WJ V Clusters

Cluster	Location in WJ V	Age 4–5 *n*	Age 4–5 r_{11}	Age 6–9 *n*	Age 6–9 r_{11}	Age 10–14 *n*	Age 10–14 r_{11}	Age 15–19 *n*	Age 15–19 r_{11}	Age 20–49 *n*	Age 20–49 r_{11}	Age 50–80+ *n*	Age 50–80+ r_{11}	Median r_{11}
General Intellectual Ability (GIA)	COG	—	—	942	0.96	1,255	0.97	1,198	0.97	905	0.97	866	0.97	0.97
Brief Intellectual Ability (BIA)	COG	—	—	948	0.91	1,256	0.91	1,198	0.91	905	0.92	866	0.92	0.91
Gf-Gc Composite	COG	—	—	941	0.94	1,256	0.94	1,198	0.94	905	0.95	866	0.95	0.94
Comprehension-Knowledge	COG	203[b]	0.87	948	0.91	1,256	0.91	1,198	0.91	905	0.93	866	0.92	0.91

Table 14. *(cont.)*
Composite Reliability for WJ V Clusters

		Age												
	Location	4–5		6– 9		10–14		15–19		20–49		50–80+		
Test	in WJ V	*n*	r_{11}	*n*	r_{11}	*n*	r_{11}	*n*	r_{11}	*n*	r_{11}	*n*	r_{11}	Median r_{11}
Fluid Reasoning	COG	—	—	942	0.89	1,256	0.90	1,198	0.90	905	0.91	866	0.92	0.90
Auditory Working Memory Capacity	COG	201[b]	0.88	947	0.88	1,256	0.89	1,198	0.91	905	0.90	866	0.91	0.90
Cognitive Processing Speed	COG	386	0.95	942	0.98	1,255	0.98	1,198	0.98	905	0.98	866	0.98	0.98
Retrieval Fluency	COG	212[b]	0.83	949	0.91	1,256	0.94	1,198	0.95	905	0.96	866	0.95	0.94
Long-Term Storage	COG	—	—	949	0.93	1,256	0.94	1,198	0.94	905	0.95	866	0.94	0.94
Visual Processing	COG	444	0.88	949	0.89	1,256	0.90	1,198	0.90	905	0.90	866	0.90	0.90
Cognitive Efficiency	COG	389	0.89	943	0.90	1,255	0.91	1,198	0.92	905	0.92	866	0.92	0.92
Phonemic Retrieval Fluency[a]	COG/VTL	—	—	906	0.90	1,239	0.94	1,188	0.95	900	0.96	852	0.95	0.95
Oral Language	ACH	—	—	949	0.93	1,256	0.94	1,198	0.93	905	0.93	866	0.92	0.93
Listening Comprehension	ACH	—	—	949	0.89	1,256	0.89	1,198	0.88	905	0.88	866	0.86	0.88
Oral Expression	ACH	444	0.85	949	0.87	1,256	0.89	1,198	0.88	905	0.89	866	0.87	0.88
Vocabulary	ACH	212[b]	0.84	949	0.89	1,256	0.92	1,198	0.93	905	0.94	866	0.93	0.92
Brief Reading	ACH	212[b]	0.96	949	0.97	1,256	0.94	1,198	0.93	905	0.93	866	0.93	0.93
Basic Reading Skills	ACH	209[b]	0.93	949	0.97	1,256	0.96	1,198	0.94	905	0.93	866	0.94	0.94
Reading Fluency	ACH	—	—	893	0.97	1,255	0.99	1,198	0.99	905	0.99	861	0.99	0.99
Reading Comprehension	ACH	—	—	948	0.95	1,256	0.93	1,198	0.91	905	0.92	866	0.92	0.92
Brief Math	ACH	—	—	946	0.93	1,256	0.94	1,198	0.95	905	0.95	866	0.94	0.94
Math Calculation Skills	ACH	—	—	938	0.94	1,255	0.96	1,198	0.97	905	0.96	866	0.96	0.96
Number Concepts	ACH	206[b]	0.89	949	0.93	1,256	0.94	1,198	0.94	905	0.94	865	0.94	0.94
Math Problem Solving	ACH	212[b]	0.90	948	0.92	1,256	0.94	1,198	0.94	905	0.94	866	0.93	0.93
Brief Writing	ACH	—	—	943	0.96	1,256	0.96	1,198	0.95	905	0.96	866	0.95	0.96
Basic Writing Skills	ACH	—	—	944	0.98	1,256	0.97	1,198	0.96	905	0.96	866	0.95	0.96
Spelling Skills	ACH	—	—	945	0.94	1,256	0.93	1,198	0.91	905	0.91	866	0.93	0.93
Written Expression	ACH	—	—	694[c]	0.93	1,252	0.94	1,194	0.95	904	0.95	865	0.95	0.95
Broad Achievement	ACH	—	—	682[c]	0.98	1,252	0.98	1,194	0.98	904	0.98	861	0.98	0.98
Academic Skills/Brief Achievement	ACH	—	—	946	0.98	1,256	0.97	1,198	0.97	905	0.96	866	0.96	0.97
Academic Fluency	ACH	—	—	682[c]	0.96	1,252	0.98	1,194	0.98	904	0.98	861	0.98	0.98
Academic Applications	ACH	200[b]	0.92	946	0.95	1,256	0.95	1,198	0.94	905	0.95	866	0.95	0.95
Academic Knowledge	ACH	212[b]	0.85	949	0.90	1,256	0.93	1,198	0.94	905	0.94	866	0.94	0.93
Phoneme-Grapheme Knowledge	ACH	—	—	946	0.93	1,256	0.92	1,198	0.90	905	0.90	866	0.92	0.92
Phonological Awareness	VTL	437	0.90	949	0.93	1,256	0.92	1,198	0.91	905	0.92	866	0.93	0.92
Phonological Manipulation	VTL	—	—	941	0.92	498[d]	0.86	—	—	—	—	—	—	0.89
Auditory Memory Span	VTL	442	0.92	949	0.90	1,256	0.87	1,198	0.86	905	0.87	866	0.87	0.87

Note. Cluster reliabilities were calculated from the contributing test reliabilities using Mosier's (1943) equation for a weighted composite reliability.
[a] The alternate-form reliability coefficient for the Rapid Phoneme Naming test was used in the computation of the cluster reliability.
[b] Reliability based on age 5 only.
[c] Reliability based on ages 7 to 9 only.
[d] Reliability based on ages 10 and 11 only.

Table 15.
Reliability for WJ V RAN Clusters

Cluster	Location in WJ V	Age 4–19		Age 20–80+	
		n	r_{12}	*n*	r_{12}
RAN–Reading	VTL	160[a]	0.88	74	0.84
RAN–Math	VTL	687	0.87	319	0.79

Note. Cluster reliabilities were calculated as Pearson correlations between the cluster *W* scores obtained from the first and second administrations of the RAN tests during the CAR study. Plausible test *W* scores were imputed for missing tests for the examinees who were not administered all the tests in the clusters.

[a] Reliability based on ages 6 to 19.

Alternate-Forms Reliability

The WJ V provides parallel second forms of 27 tests: 5 tests in the Cognitive battery, 17 tests in the Achievement battery, and 5 tests in the Virtual Test Library. Second forms can be used (a) if the administration of a speeded test is spoiled due to an unexpected interruption during the timed portion of the test or (b) when an examinee who was administered the first form of a test in a prior evaluation is being retested. Alternate-form reliability coefficients for the WJ V tests with parallel second forms were obtained from the sample of individuals who participated in the CAR study that ran concurrently with the WJ V norming study. Most (95.7%) of the CAR examinees were administered all the tests in their assigned administration set within the same testing session. Some younger examinees were scheduled across two separate testing sessions; however, even in these cases, the time elapsed between the administration of the two forms of each test was generally less than 24 hours. Table 16 contains the alternate-form reliability coefficients for the 24 tests that contain a parallel second form, for ages 4 to 19 and ages 20 to 80+.

Table 16.
Alternate-Form Reliability Coefficients for Parallel First and Second Forms of WJ V Tests

Test	Location in WJ V	Age 4–19		Age 20–80+	
		n	r_{12}	*n*	r_{12}
Number-Pattern Matching	COG STD	388	0.94	171	0.87
Letter-Pattern Matching	COG STD	386	0.92	161	0.86
Symbol Inhibition	COG EXT	267[e]	0.81	134	0.78
Letter-Word Identification[a]	ACH STD	373	0.99	146	0.85
Calculation[a]	ACH STD	323[d]	0.96	160	0.88
Spelling[a]	ACH STD	341[c]	0.98	150	0.88
Word Attack[a]	ACH STD	356[c]	0.95	163	0.85
Math Facts Fluency[a, b]	ACH STD	162[e]	0.97	77	0.93
Sentence Writing Accuracy	ACH STD	303[d]	0.97	140	0.88
Passage Comprehension[a]	ACH STD	355[c]	0.96	164	0.64
Applied Problems[a]	ACH STD	400	0.96	172	0.85
Sentence Reading Fluency[a, b]	ACH STD	173[c]	0.94	76	0.93
Written Language Samples[a]	ACH STD	320[c]	0.95	149	0.69
Sentence Writing Fluency[a, b]	ACH STD	156[e]	0.95	83	0.93
Paragraph Reading Comprehension[a]	ACH STD	304[d]	0.93	149	0.68
Word Reading Fluency[a, b]	ACH STD	159[d]	0.92	72	0.77
Magnitude Comparison	ACH EXT	361	0.97	155	0.88
Number Sense	ACH EXT	334[c]	0.95	142	0.84

Table 16. *(cont.)*
Alternate-Form Reliability Coefficients for Parallel First and Second Forms of WJ V Tests

Test	Location in WJ V	Age			
		4–19		20–80+	
		n	r_{12}	*n*	r_{12}
Oral Reading[a]	ACH EXT	345[d]	0.93	176	0.79
Rapid Picture Naming	VTL	376	0.88	164	0.80
Rapid Letter Naming	VTL	369	0.95	149	0.80
Rapid Phoneme Naming[a, b]	VTL	162[d]	0.82	72	0.90
Rapid Number Naming	VTL	364	0.97	145	0.89
Rapid Quantity Naming	VTL	397	0.94	175	0.89

Note. Reliability coefficients are Pearson correlations between the first and second forms of each test.
[a] The order of form presentation was counterbalanced during the CAR study.
[b] The reliability coefficients for this speeded test were computed using data from only those examinees who were administered the first form of the test before the second form.
[c] Reliability based on ages 5 to 19.
[d] Reliability based on ages 6 to 19.
[e] Reliability based on ages 7 to 19.

Test-Retest Reliability

In the case of the WJ V Semantic Word Retrieval, Phonemic Word Retrieval, and Letter Writing Fluency tests—in which the second form of the test is identical to the first form—the test-retest reliability coefficients provide indices of how similar the examinee scores are rank-ordered between the first and second administrations of the test. Test-retest reliability coefficients for the WJ V tests with identical second forms were obtained from the sample of individuals who participated in the CAR study that ran concurrently with the WJ V norming study. Table 17 contains the test-retest reliability coefficients between the W-ability scores obtained from the first and second administrations of these three tests for two broad age groups.

Table 17.
Test-Retest Reliability Coefficients for Identical First and Second Forms of WJ V Speeded Tests

Test	Location in WJ V	Age			
		4–19		20–80+	
		n	r_{12}	*n*	r_{12}
Semantic Word Retrieval	COG STD	352	0.95	143	0.85
Phonemic Word Retrieval	COG STD	353[a]	0.92	162	0.89
Letter Writing Fluency	ACH EXT	141[b]	0.94	—	—

Note. Reliability coefficients are Pearson correlations between the first and second administrations of each test.
[a] Reliability based on ages 6 to 19.
[b] Reliability based on ages 4 to 9.

Impact of Prior Exposure

Although not a characteristic of the test forms themselves, prior exposure to a test task can impact an individual's performance on subsequent administrations of the test—even when the content of the individual items is different between the two forms. This is sometimes referred to as *carryover effect* (Allen & Yen, 2001) or *practice effect*. For the WJ V tests, prior exposure is most likely to impact examinee performance on tests that require the examinee to perform simple tasks repeatedly under timed conditions. Examples of such tests in the WJ V are the processing speed (Gs) tests and the rapid automatized naming (RAN) tests. Additionally, the retrieval fluency (Gr) tests, particularly Semantic Word Retrieval, may be easier for some examinees on the second administration because they reflect on the strategy they used during the first administration and may then employ a more efficient or effective cognitive strategy on the subsequent administration.

The impact of practice on a second test administration can be quantified using Cohen's *d*, or the difference in mean scores between the two forms divided by the pooled standard deviation of scores from both forms. Effect sizes of 0.2, 0.5, and 0.8 are considered small, medium, and large, respectively (Cohen, 2013). Positive effect sizes indicate higher mean scores on the second administration, whereas negative effect sizes indicate lower mean scores on the second administration. Most (96.9%) of the examinees in the CAR study (described earlier in this document) were administered the first form of the tests followed by the second form of the tests within the same day (minimum = 0.03 hours, maximum = 43 days,[6] median = 0.98 hours). Table 18 contains the effect sizes, by age group, for the mean score changes from the first to second administration of each test.

Table 18.
Effect Sizes for Mean Score Differences Between First and Second Administration of the Speeded WJ V Tests

Test	Location in WJ V	Age			
		4–19		20–80+	
		n	Effect Size	*n*	Effect Size
Number-Pattern Matching	COG STD	388	–0.02	171	–0.10
Semantic Word Retrieval	COG STD	352	0.06	143	–0.10
Letter-Pattern Matching	COG STD	386	–0.02	161	0.00
Phonemic Word Retrieval	COG STD	353[b]	0.03	162	–0.07
Symbol Inhibition	COG EXT	267[d]	0.03	134	0.11
Math Facts Fluency[a]	ACH STD	162[c]	0.10	77	0.12
Sentence Reading Fluency[a]	ACH STD	173[b]	0.08	76	0.23
Sentence Writing Fluency[a]	ACH STD	156[d]	0.13	83	0.08
Word Reading Fluency[a]	ACH STD	159[c]	0.23	72	0.27
Magnitude Comparison	ACH EXT	361	0.01	155	–0.09
Letter Writing Fluency	ACH EXT	141[e]	0.13	—	—
Rapid Picture Naming	VTL	376	–0.01	164	0.03
Rapid Letter Naming	VTL	369	–0.02	149	0.27
Rapid Phoneme Naming[a]	VTL	162[c]	0.57	72	0.52
Rapid Number Naming	VTL	364	–0.01	145	0.15
Rapid Quantity Naming	VTL	397	0.02	175	0.20

Note. Effect sizes are Cohen's *d* standardized differences between the means of the first and second administration of each test.
[a] The two forms of these tests were presented in counterbalanced order during the CAR study; the effect sizes reported here were computed from only those examinees who were administered the first form of the test before the second form.
[b] Effect size based on ages 5 to 19.
[c] Effect size based on ages 6 to 19.
[d] Effect size based on ages 7 to 19.
[e] Effect size based on ages 4 to 9.

[6] A small number of examinees (approximately 3% of the total sample) was administered the second form of the test more than 1 day after the first form due to unexpected scheduling conflicts.

Validity Evidence

Following the framework consistent with that outlined in the *Standards for Education and Psychological Testing* (AERA et al., 2014), this section presents a summary of the key evidence presented in the WJ V Technical Manual (LaForte et al., 2025) to support the use and interpretation of the WJ V test and cluster scores. Space constraints in this document preclude a full discussion of the breadth and depth of validity evidence presented there; as a result, only high-level summary evidence is presented in this document. In cases where analyses were conducted across multiple age groups and results are presented in figures, this document contains only the results for the age 10 to 14 group; however, users who are interested in more details, or results from other age groups, may consult Chapter 6 and Appendices D through J of the WJ V Technical Manual.

Content, Facet, and Cognitive Complexity Characteristics of the WJ V Tests

The WJ V includes tests measuring a complex set of abilities constituting intellectual ability, oral language ability, and achievement under CHC theory. Evidence supporting the content or substantive validity for the WJ V scores is provided via the specification of test and cluster content according to contemporary CHC research and theories. This aspect of the WJ V validity argument builds upon the theoretical frameworks presented in the four prior editions of the battery, which were based on successive revisions to the Cattell-Horn Extended Gf-Gc and Cattell-Horn-Carroll (CHC) theories of cognitive abilities (McGrew, 2005, 2009, 2023; Schneider & McGrew, 2012, 2018).

Content Validity

Content validity evidence "usually takes the form of consensual professional judgments about the relevance of item content to the specified domain and about the representativeness with which the test content covers the domain content" (Messick, 1989, p. 36). In the WJ V, all broad CHC abilities are represented by at least two tests measuring narrow CHC abilities. For a majority of WJ IV tests retained in the WJ V, support for the CHC content classifications was provided in the WJ IV Technical Manual (McGrew et al., 2014). These prior classifications were based on multiple iterations of cross-battery CHC expert consensus of WJ III and WJ IV tests (Flanagan et al., 2006; Flanagan et al., 2007; Flanagan et al., 2013; McGrew, 1997; McGrew & Flanagan, 1998). For the WJ V, author-assigned CHC broad and narrow ability test classifications were compared to the consensus classifications assigned by an independent panel of experts as part of a larger CHC test classification pilot study (Flanagan, Ortiz, & Alfonso, in press). Three levels of broad CHC ability test classification agreement were calculated between WJ V authors' and CHC experts' broad ability classifications. Complete agreement across all tests was 77% (46 of 60 tests). Partial agreement was 22% (13 of 60 tests). Together, these account for a 98% total agreement rate (59 of 60 tests). Similarly, three levels of agreement were calculated based on the WJ V authors' and CHC experts' narrow ability classifications. Complete agreement was 63% (38 of 60 tests) and partial agreement was 30% (18 of 60 tests). Together, these account for a 93% total agreement rate (56 of 60 tests). The total agreement for both the broad (98%) and narrow (93%) ability classifications is considered high, providing CHC-based content validity evidence for the WJ V tests.

Test Content or Stimulus (Facet) Characteristics

Research has demonstrated, within academic curriculum domains, the potential for evaluating content validity via empirical methods. Li and Sireci (2013) demonstrated the value of using multidimensional scaling (MDS), augmented by cluster analysis, to evaluate the correspondence between empirically identified content dimensions and professional judgment. McGrew et al.

(2014) and Meyer and Reynolds (2022) have demonstrated how MDS can provide content validity evidence for combined cognitive and achievement test data. The most valuable MDS output is a visual-spatial figure, or map, where the relations between the variables are represented by distances and facets (i.e., dimensions) on the map. These maps help users understand the key facets among the variables. In contrast to factor analysis, interpretation of MDS is more qualitative and subjective. However, MDS has unique potential for evaluating the content validity of tests as it provides information about "both the content and processes underlying performance on diverse cognitive tasks" (McGrew, 2005, p. 172; McGrew et al., 2014).

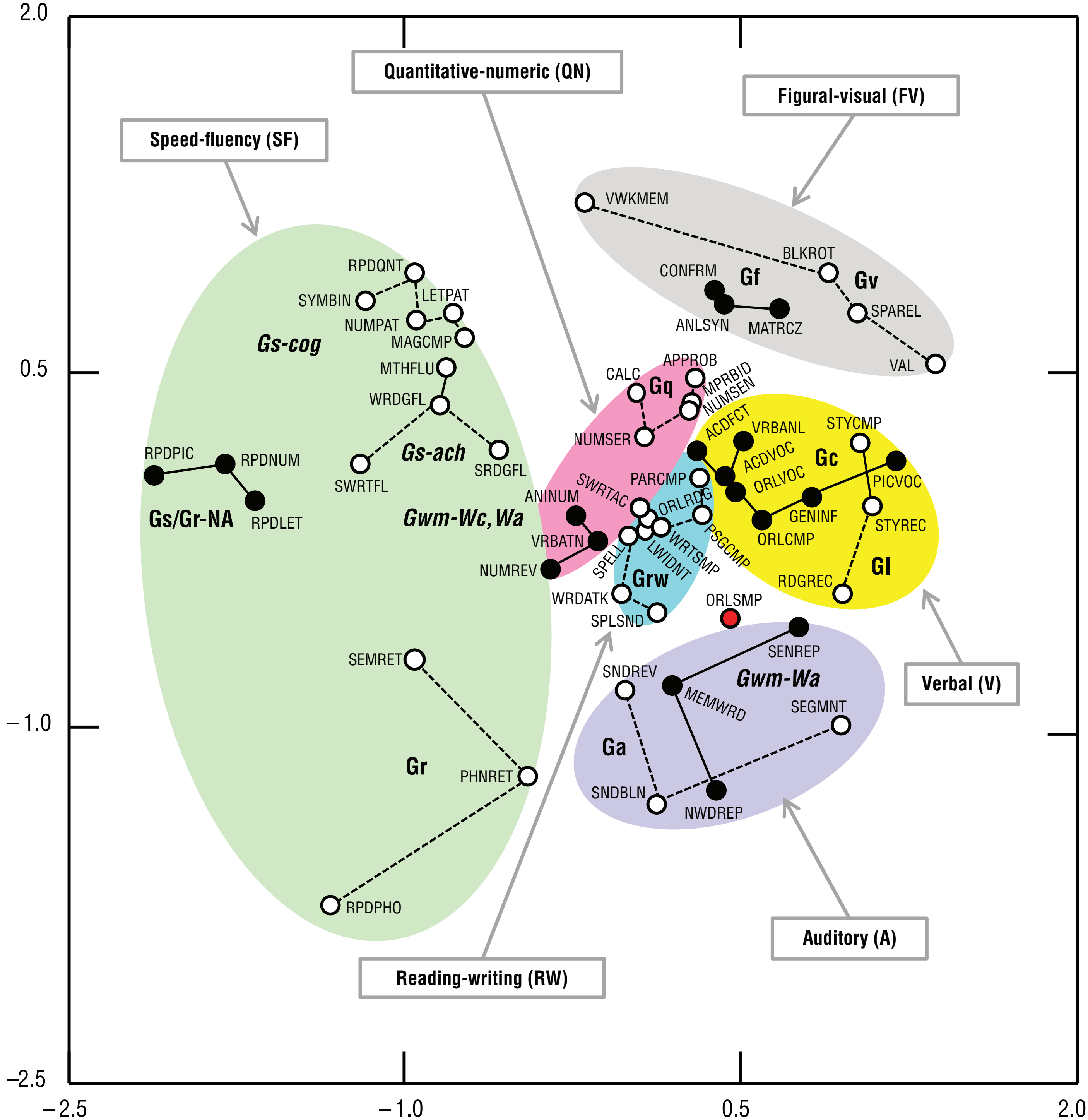
Following the precedent set with the WJ IV (McGrew et al., 2014), Guttman's Radex two-dimensional MDS procedure was applied to the WJ V norm data. The correlations between all COG, ACH, and VTL tests included in the WJ V norming study were analyzed with a Guttman Radex two-dimensional MDS procedure (SYSTAT Version 13.1, 2009) at each of six WJ V age groups. Figure 4 presents the MDS results for all WJ V COG, ACH, and VTL tests for the age 10 to 14 group model development (MD) Sample A. All other age-group results can be found in Appendix H of the WJ V Technical Manual.

In the context of this discussion, the mass of information visually portrayed in Figure 4 should be overlooked in favor of a big picture understanding. The tests that share either cognitive process or stimulus content characteristics are typically in close spatial proximity and are connected by lines (solid or dashed) or are organized into larger groups or facets by the superimposed shaded ovals. The interpretation of spatially similar tests per the CHC model (Schneider & McGrew, 2018) is included in Figure 4 and discussed later in this chapter. The six large, shaded ovals are relevant to the current content validity discussion. See Tables 3 through 5 for detailed descriptions of the WJ V tests that aided interpretation of the content facets in Figure 4. The MDS analysis of the WJ V tests provides empirical support for six broad types of shared stimulus content facets (viz., verbal, auditory, figural-visual, quantitative-numeric, speed-fluency, and reading-writing).

Cognitive Complexity of Tests

One WJ V revision goal was to increase the cognitive complexity demands of certain revised or new WJ V tests and clusters. Two different approaches are typically used to achieve this goal (McGrew et al., 2014). The first approach, which is a common approach in applied test development, is to deliberately design factorially complex CHC tests, or tests that deliberately include the influence of two or more narrow CHC abilities. In this approach, construct-irrelevant variance (Benson, 1998; Messick, 1995) is not deliberately minimized or eliminated. Although tests that measure more than one narrow CHC ability typically have lower validity as indicators of CHC abilities, they tend to lend support to other types of validity evidence (e.g., higher predictive validity). Verbal Analogies (a mixed measure of Gc and Gf) and Visual Working Memory (a mixed measure of Gwm and Gv) are examples of factorially complex WJ V tests.

Figure 4.
CHC Cognitive Operations and Content Facet Interpretation of MDS (Guttman Radex) for Age 10 to 14 years Model Development (MD) Sample A, All Tests (n = 630)



Notes. Black circles with solid connecting lines and white circles with dashed connecting lines are used to aid in the visual interpretation of broad CHC test dimensions. Italic font for Gs-cog, Gs-ach, Gwm-Wc,Wa, and Gwm-Wa indicates possible Gs and Gwm substructures. Shaded ovals represent content and speed-fluency MDS facets. Oral Language Samples (ORLSMP; red circle) is not a clear member of a specific CHC dimension. Understanding Directions is omitted from this analysis.

The second approach to enhancing the cognitive complexity of tests is to maintain the CHC factor purity of tests or clusters (as much as possible) while concurrently and deliberately increasing the complexity of information processing demands of the tests within the specific broad or narrow CHC domain (McGrew, 2012). As described by Lohman and Lakin (2011), the cognitive complexity of the abilities measured by tests can be increased by (a) increasing the number of cognitive component processes, (b) including differences in speed of component processing, (c) increasing the number of more important component processes (e.g., inference), (d) increasing the demands of attentional control and working memory, or (e) increasing the demands on adaptive functions (assembly, control, and monitoring). This second form of cognitive complexity, not to be confused with factorial complexity, is defined as the inclusion of test tasks that place greater demands on cognitive information processing (i.e., cognitive load), that require greater allocation of key cognitive resources (viz., working memory or attentional control), and that invoke the involvement of more cognitive control or executive functions. Per this second form of cognitive complexity, the objective is to design a test that is more cognitively complex within a CHC domain, not to deliberately make it a mixed measure of two or more CHC abilities.

One form of evidence for determining the cognitive complexity of tests is to inspect their locations on the Guttman Radex MDS model. Tests closest to the center of the two-dimensional MDS plots are often interpreted as being more cognitively complex (Cohen et al., 2006; Marshalek et al., 1983; Tucker-Drob & Salthouse, 2009), and the cognitive complexity decreases the further away a test is from the center of a plot. The MDS of the WJ V tests did not show this complexity pattern in the age 10 to 14 group (see Figure 4) or any of the other age-group samples. It is hypothesized that the MDS of joint cognitive and achievement tests may not be appropriate for accurately classifying the degree of cognitive complexity of a large set of combined cognitive and achievement tests per MDS Radex principles. Twenty-six of the 56 WJ V tests (46%) included in the MDS analysis reported in Figure 4 are from the ACH battery. The inclusion of such a disproportionately large number of reading, writing, and math achievement tests in the MDS analysis biases the center of the MDS figure toward academic achievement (Grw, Gq) as seen in Figure 4.

A more commonly accepted method for analyzing the cognitive complexity of tests is to examine each test's loading on a single general intelligence factor (i.e., psychometric *g*) extracted from the collection of tests (Kaufman, 1979; Meyer & Reynolds, 2022). The rationale is that more cognitively complex tests require abstract reasoning and problem solving and invoke a wider range of elementary cognitive processes (Jensen, 1998; Stankov, 2000, 2005), which in turn is reflected in relatively higher psychometric *g* loadings. Typically, but not with 100% congruence, factor analysis results-based test psychometric *g* classifications often correspond to cognitive complexity as determined by MDS methods (Marshalek et al., 1983). Like the *g* loadings for the tests across all prior editions of the WJ presented in WJ V Technical Manual Chapter 2, all WJ V COG and VTL tests were analyzed with principal component analyses at separate age groups.[7] Each test's loading on the first unrotated principal component by age group is presented in Table 19.

[7] This analysis was not completed with the age 4 to 5 group due to the more restricted set of available test indicators at this age range compared to the other five analysis age groups.

Table 19.
Psychometric g *Loadings (First Unrotated Principal Component) for WJ V COG and VTL Tests Common Across Five Age Groups*

Tests	Broad CHC Abilities	Location in WJ V	Age 6–9 (*n* = 949)	Age 10–14 (*n* = 1,256)	Age 15–19 (*n* = 1,198)	Age 20–49 (*n* = 905)	Age 50–80+ (*n* = 866)	Median *g* Loading
Oral Vocabulary	Gc	COG	0.76	0.73	0.74	0.78	0.78	**0.76**
Verbal Analogies	Gc/Gf	COG	0.71	0.71	0.72	0.79	0.76	**0.72**
Verbal Attention	Gwm	COG	0.73	0.67	0.72	0.72	0.71	**0.72**
Animal-Number Sequencing	Gwm	VTL	0.68	0.69	0.72	0.71	0.71	0.71
Understanding Directions	Gwm/Gf	VTL	0.69	0.70	0.73	0.76	0.70	0.70
Concept Formation	Gf	COG	0.65	0.68	0.69	0.70	0.68	0.68
Number Series	Gq/Gf	COG	0.70	0.68	0.68	0.67	0.68	0.68
General Information	Gc	COG	0.62	0.65	0.68	0.72	0.70	0.68
Numbers Reversed	Gwm	COG	0.71	0.67	0.67	0.67	0.70	0.67
Analysis-Synthesis	Gf	COG	0.66	0.66	0.67	0.70	0.67	0.67
Story Recall	Gl	COG	0.67	0.64	0.66	0.65	0.66	**0.66**
Matrices	Gf	COG	0.55	0.63	0.61	0.66	0.67	**0.63**
Sentence Repetition	Gwm	VTL	0.62	0.61	0.61	0.63	0.63	0.62
Story Comprehension	Gl	COG	0.58	0.61	0.62	0.66	0.59	0.61
Spatial Relations	Gv	COG	0.61	0.62	0.62	0.58	0.61	**0.61**
Sound Blending	Ga	VTL	0.58	0.57	0.60	0.65	0.62	0.60
Visual-Auditory Learning	Gv/Gf	COG	0.66	0.58	0.60	0.60	0.63	0.60
Letter-Pattern Matching	Gs	COG	0.51	0.59	0.67	0.66	0.59	0.59
Number-Pattern Matching	Gs	COG	0.45	0.59	0.65	0.63	0.55	**0.59**
Phonemic Word Retrieval	Gr	COG	0.59	0.57	0.58	0.63	0.63	0.59
Memory for Words	Gwm	VTL	0.60	0.59	0.58	0.59	0.58	0.59
Sound Reversal	Ga	VTL	0.56	0.58	0.57	0.58	0.63	0.58
Semantic Word Retrieval	Gr	COG	0.56	0.55	0.61	0.60	0.58	**0.58**
Segmentation	Ga	VTL	0.66	0.56	0.50	0.58	0.64	0.58
Block Rotation	Gv	COG	0.57	0.59	0.57	0.58	0.61	0.58
Rapid Quantity Naming	Gs/Gr/Gv	VTL	0.43	0.58	0.65	0.60	0.57	0.58
Symbol Inhibition	Gs	COG	0.49	0.52	0.61	0.60	0.57	0.57
Nonsense Word Repetition	Ga/Gwm	VTL	0.59	0.56	0.55	0.54	0.57	0.56
Rapid Letter Naming	Gs/Gr	VTL	0.40	0.51	0.56	0.58	0.58	0.56
Visual Working Memory	Gwm/Gv	COG	0.49	0.56	0.57	0.54	0.53	0.54
Rapid Phoneme Naming	Gs/Gr/Ga	VTL	0.39	0.43	0.47	0.55	0.54	0.47
Rapid Number Naming	Gs/Gr	VTL	0.33	0.47	0.52	0.53	0.47	0.47
Rapid Picture Naming	Gs/Gr	VTL	0.31	0.43	0.48	0.53	0.45	0.45

Notes. Bold font indicates tests in the COG GIA cluster. Tests are sorted and listed in descending order of median *g*-loading values. Kaufman's (1979, 1990) original system of classifying median *g* loadings is reflected in the shading: .70 or above (dark gray shading) = good; .51 to .69 (no shading) = ample; ≤ .50 (light gray shading) = poor. Instead of Kaufman's verbal labels, we use McGrew and Flanagan's (1998) suggested labels of high, medium, and low, as the Kaufman system implies a "good-bad" continuum that can be misleading (i.e., a test with a poor/low *g* loading may still be useful and valuable for other interpretation purposes).

The magnitude of the *g* loadings presented in Table 19 can be used to evaluate the relative cognitive complexity of each WJ V COG test. For example, Oral Vocabulary (Gc; median *g* = .76) and Verbal Analogies (Gc/Gf; median *g* = .72) are the most cognitively complex measures (when defined by high psychometric *g* loadings) and are featured as the two-test Gc cluster. In contrast, the other COG primary Gc test, General Information, has a median *g* loading of .68 (medium). As designated by bold font, three of the eight tests (Oral Vocabulary, Verbal Analogies, and Verbal Attention) that contribute to the COG GIA cluster are classified as high indicators of psychometric *g* in the WJ V. The remaining five GIA tests all have medium *g* loadings. None of the Ga tests in the WJ V VTL has high psychometric *g* classifications—all are medium (median *g* loadings between .58 and .60). Three of the five VTL rapid naming (RAN) tests (viz., Rapid Phoneme Naming, Rapid Number Naming, and Rapid Picture Naming) are classified as low (median *g* loadings from 0.45 to 0.47) indicators of psychometric *g*. Rapid Letter Naming (0.56) and Rapid Quantity Naming (0.58) appear to have relatively more complex cognitive processing demands than the other three RAN tests.

Developmental Patterns of WJ V Ability Clusters

The WJ V tests and clusters display average score changes consistent with the developmental growth and decline of cognitive and achievement abilities across the life span.[8] Divergent growth curves provide validity evidence for the existence of distinct, unique abilities (Baltes et al., 1999; Blair, 2006; Breit et al., 2024; Carroll, 1993; Horn, 1991; Horn & Noll, 1997; Kaufman, 2001; Salthouse, 2012). The WJ V growth curves illustrate that the unique abilities measured by distinctly different WJ V CHC clusters (e.g., Retrieval Fluency-Gr compared to Comprehension-Knowledge-Gc) follow different developmental courses or trajectories from childhood to geriatric levels.

Cross-sectional growth curves for most WJ V COG, VTL, and ACH clusters are presented in Figures 5 through 8; the WJ V COG GIA cluster serves as a common reference curve across all four figures. The growth curves were constructed using age 6 years, 0 months as a starting point and subtracting the norm-based Reference-*W* (REF *W*) score at age 6 years, 0 months from all other REF *W*s for that cluster up through age 90+. Age 6 years, 0 months was selected as the starting point because all WJ V clusters have normative REF *W* scores at and above this age. This procedure produced growth curves with a common origin of zero. To remove any minor age-related fluctuations, the curves were subjected to additional statistical smoothing. Additionally, although the range of the *y*-axis values varies by cluster, the *y*-axis range is held constant (0 to 110) to allow comparison of curves across the four figures. Figures 5 and 6 present the COG and VTL broad, narrow, and clinical clusters that are not mixed measures of multiple CHC ability tests (e.g., Gf-Gc Composite, BIA, Cognitive Efficiency). Given that a significant number of COG and oral language ACH clusters measure different aspects of Gc, the COG Comprehension-Knowledge and ACH Academic Knowledge, Vocabulary, Oral Expression, Listening Comprehension, and Oral Language clusters are presented as a set of Gc-related growth curves in Figure 7. Figure 8 presents the COG GIA and all reading, writing, and math clusters.

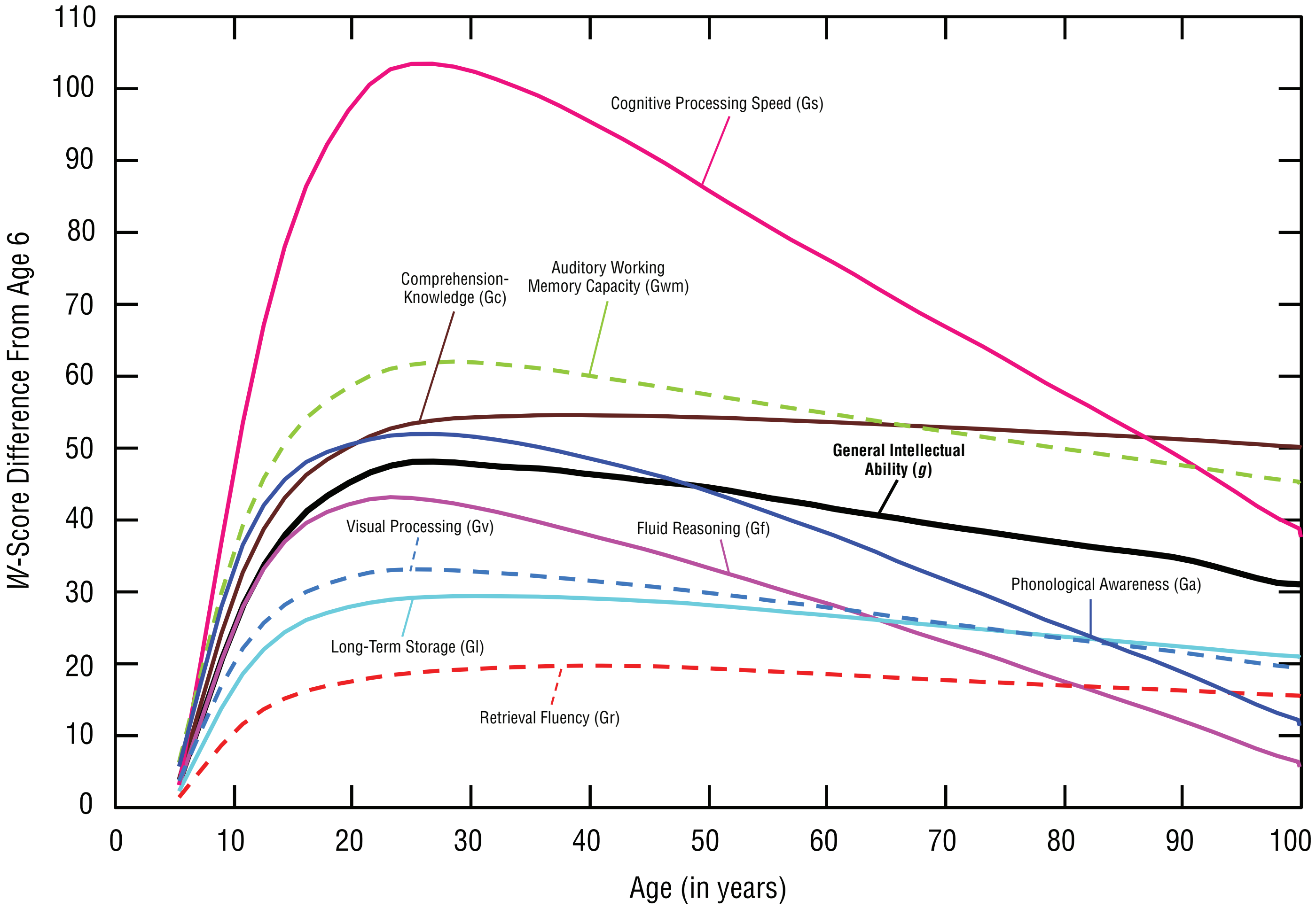
General (*g*), Broad, and Narrow WJ V COG Cluster Growth Curves

Growth curves for the General Intellectual Ability (GIA, psychometric *g*) and eight broad CHC COG clusters are presented in Figure 5. The patterns of relative growth and decline of the WJ V CHC cognitive clusters differ markedly, providing additional evidence supporting the interpretation of the clusters as measures of distinct abilities. For example, the Retrieval Fluency (Gr) cluster demonstrates less developmental change than any of the other WJ V COG abilities. The growth curves for the Long-Term Storage (Gl), Visual Processing (Gv), and Fluid Reasoning (Gf) clusters are also consistent with abilities that develop more as a function of informal and

[8] Summary descriptive statistics for all tests and clusters are available in Appendices B and C of the WJ V Technical Manual.

indirect learning experiences. Fluid Reasoning (Gf) shows a much more dramatic decrease in average level of performance with increasing age (starting approximately between ages 25 to 30) compared to the Retrieval Fluency (Gr), Long-Term Storage (Gl), and Visual Processing (Gv) clusters. In contrast, the growth curve for the Comprehension-Knowledge (Gc) cluster is an example of a measure in which direct and more formalized learning is an important causal factor. This Gc cluster reaches its apex later than the Gr, Gl, Gv, Gf, and Ga clusters and demonstrates continual growth (or maintenance) well into late adulthood.

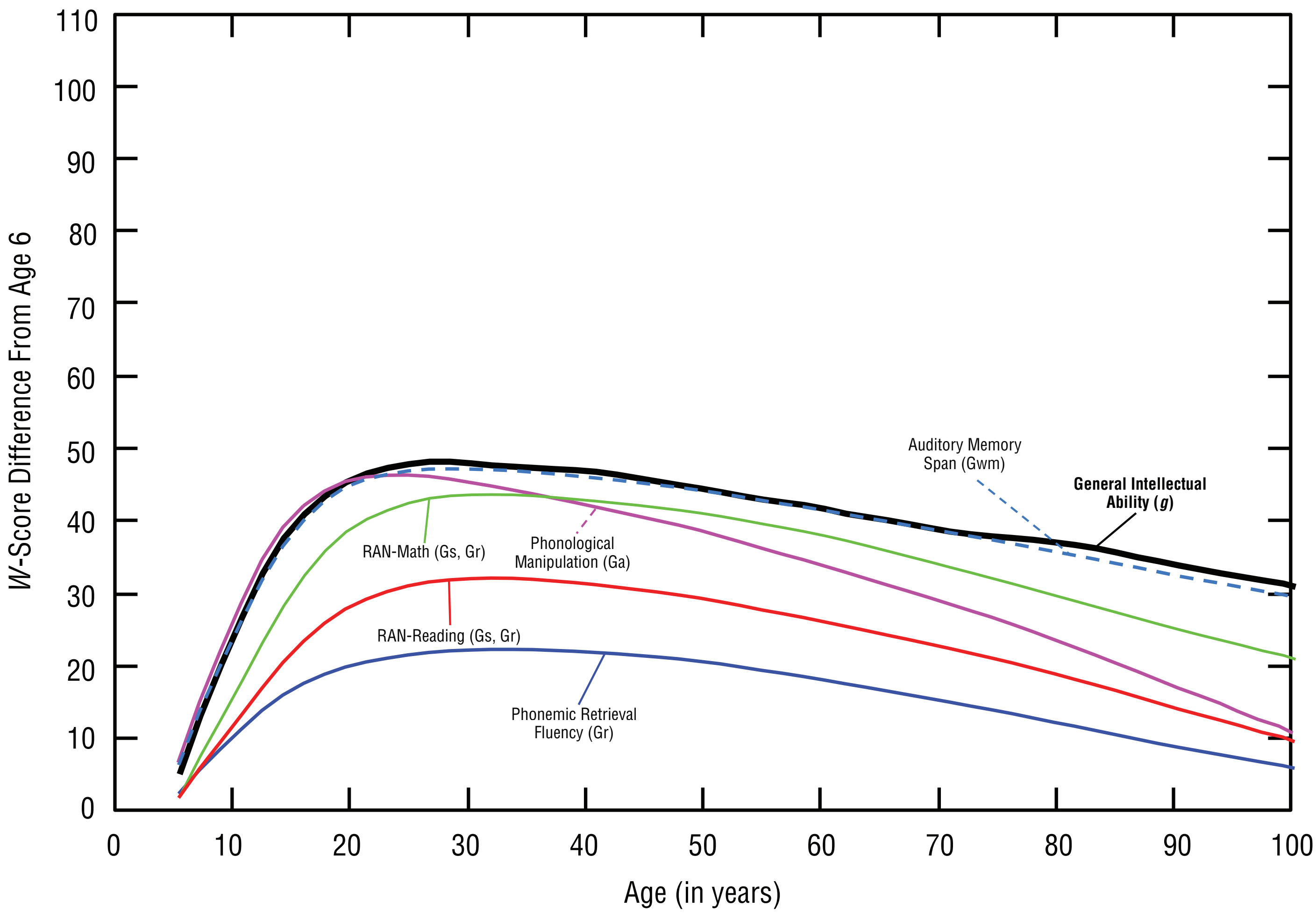
Figure 5.
Plot of WJ V COG GIA and Eight CHC Factor Cluster W-Score Difference Curves by Age



VTL Narrow Ability and Clinical Cluster Growth Curves

Figure 6 presents the growth curves for five of the VTL narrow ability and clinical clusters. The Auditory Memory Span (Gwm) cluster curve mirrors the GIA (psychometric g) curve. The Auditory Memory Span (see Figure 6) and Auditory Working Memory Capacity (see Figure 5) cluster growth curves are noticeably different, providing evidence that these two distinct WJ V working memory clusters measure different aspects of Gwm. The Phonological Manipulation (Ga) cluster growth curve in Figure 6 is similar in shape to the other WJ V Ga cluster (Phonological Awareness) growth curve shown in Figure 5. As expected, given their classification as measures of Gr, the Phonemic Retrieval Fluency (Gr) curve in Figure 6 is like the Retrieval Fluency (Gr) curve in Figure 5, although Phonemic Retrieval Fluency (Gr) demonstrates a more rapid rate of decline than Retrieval Fluency (Gr) does, suggesting that these two related Gr clusters may measure similar cognitive abilities until middle age and beyond. Finally, the two RAN cluster curves diverge from one another in early rate of growth and approximate age of plateau, suggesting that these clinically formed clusters measure different cognitive constructs.

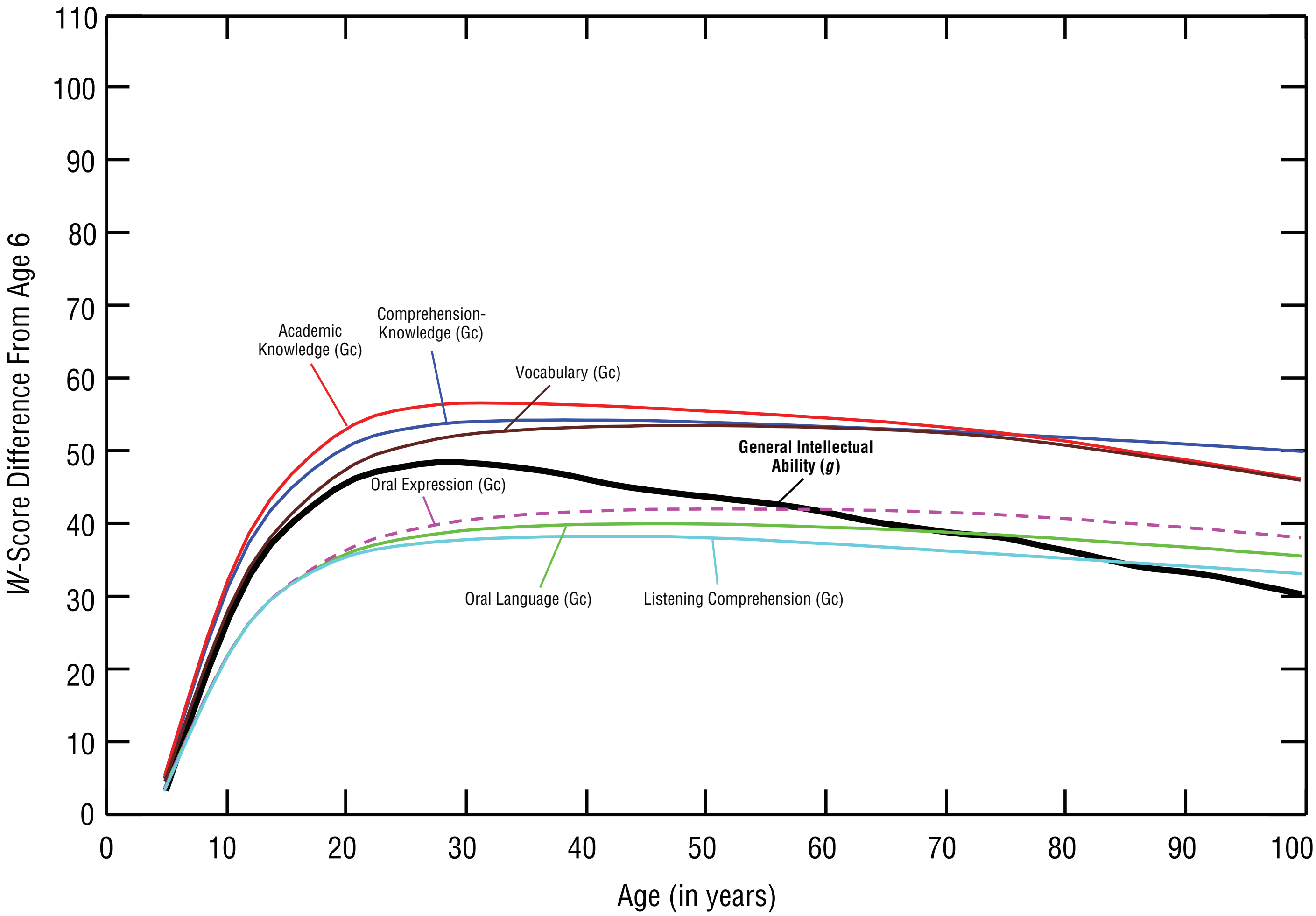
Figure 6.
Plot of WJ V COG GIA and Five VTL Cluster
W-Score Difference Curves by Age



COG and ACH Gc or Language-Related Cluster Growth Curves

Six Gc-related WJ V cluster growth curves are presented in Figure 7. Given that all six clusters measure different aspects of Gc (see Chapter 2 of the WJ V Technical Manual), the similarity of the shape of the six cluster curves, especially when compared to those displayed in Figure 5, is not unexpected. Four primary conclusions can be drawn from the WJ V curves in Figure 7. First, compared to the GIA (psychometric g) reference curve, after reaching their respective asymptotes (at roughly the same age range), all six Gc-related clusters tend to maintain the same general level of performance across adulthood, except for some minor slow decline in abilities in late adulthood. Second, the clusters comprised of tests typically included in cognitive ability or intelligence batteries (i.e., Comprehension-Knowledge, Academic Knowledge, Vocabulary) show similar rates of growth during early childhood through young adulthood, followed by relatively strong maintenance of performance from middle to late adulthood. Third, the Oral Expression and Listening Comprehension clusters (as well as the Oral Language cluster, which is a combination of the tests from these two clusters) "hang together" much like the cognitive Gc-cluster measures hang together. Although Listening Comprehension and Oral Expression are considered measures of Gc abilities, they are abilities that are developmentally different from the more traditional cognitive Gc measures of comprehension-knowledge, vocabulary, and general academic knowledge. Finally, divergence of the Oral Language cluster curve from the GIA (psychometric *g*) cluster curve suggests that the WJ V Oral Language cluster represents primarily Gc, distinct from the GIA cluster that is a broader and more comprehensive mixture of Gc, Gf, Gv, Ga, Gwm, Gl, Gr, and Gs.

Figure 7.
Plot of WJ V COG GIA and All COG/ACH Gc Cluster W-Score Difference Curves by Age



WJ V ACH Cluster Growth Curves

Growth curves for 12 WJ V ACH clusters are presented in Figure 8. Close examination of the ACH curves reveals several developmental patterns of interest. First, the most obvious is how all ACH cluster growth curves are distinctly different from the GIA (psychometric *g*) cluster. All ACH cluster growth curves (when compared to the GIA cluster) display steeper early growth, much higher asymptotes, and—except for the clusters that contain speeded tests—less absolute decline with age. These are classic patterns for abilities that are most influenced by formal educational opportunities, which is the pragmatic distinction between achievement and cognitive abilities. Second, compared to the COG cluster growth curves in Figures 5 and 6, most achievement clusters do not demonstrate as much absolute decline across the age span as the cognitive ability clusters do; the achievement abilities are generally maintained at higher levels into the older age ranges. Third, the subareas of reading show distinctly different growth curve trends. Reading decoding skills (measured by the Basic Reading Skills cluster) grow rapidly and are maintained at a higher level throughout the life span than reading comprehension abilities are. In contrast, fluent reading skills (measured by the Reading Fluency cluster) display rapid early growth like Basic Reading Skills do but then decline more rapidly after approximately age 30. Fourth, within writing, Basic Writing Skills demonstrates a growth curve similar in shape to (but lower than) Basic Reading Skills. The Basic Writing Skills, Written Expression, and Spelling Skills clusters all demonstrate asymptotes at approximately ages 20 to 40, but Basic Writing Skills shows much more absolute relative growth than the other two writing clusters. The three clusters diverge in maintenance (or rate of decline) after approximately ages 40 to 50. Basic Writing Skills maintains better in middle to late adulthood, while Spelling Skills and, especially, Written Expression show more rapid decline during the same age period. As noted above for the Reading Fluency cluster, the more rapid rate of decline for Written Expression is likely because this two-test cluster includes one test that measures general cognitive processing speed (Gs) variance. Finally, two of the three math achievement clusters (Math Calculation Skills and Math Problem Solving) demonstrate very similar growth curves in terms of the rate of early growth, age of asymptote, and rate of decline. The Number Concepts cluster, a measure of core cognitive-related math abilities, demonstrates a relatively longer rate of early growth and a higher asymptote (between approximately 20 to 30 years of age) than the other two math clusters, followed by a much steeper rate of decline starting at 40 to 50 years of age. Given that the speeded Magnitude Comparison test is included in the two-test Number Concepts cluster, it is likely that the different shape of the Number Concepts cluster growth curve is influenced by the general cognitive processing speed (Gs) variance of this test.

Interpretation of the achievement growth curves is difficult as several of the curves reflect the influence of cognitive processing speed (Gs). Because of the achievement/cognitive ability factorial complexity of many of the ACH growth curves, the curves presented in Figure 8 are best interpreted as indicators of the rate of growth of the manifest WJ V achievement clusters and should be interpreted only cautiously as reflecting latent achievement ability constructs.

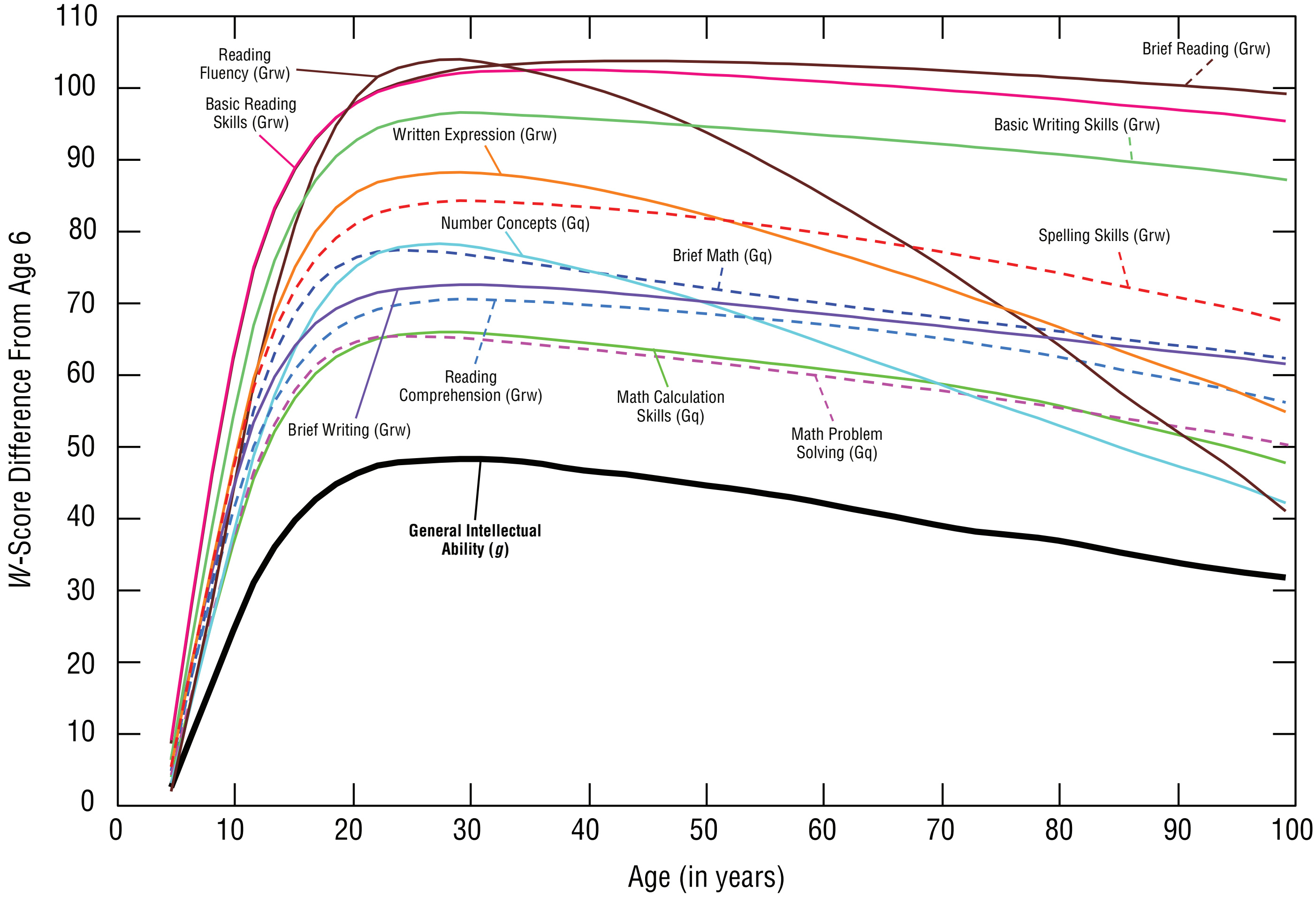
Internal Structural Validity Evidence for the WJ V

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

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 V provide empirical support for several inferences about the relations between the WJ V scores.[9] 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 is lower than those reported among the primary achievement clusters, providing evidence that the WJ V 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.

[9] Complete correlation matrices for all tests and clusters are reported in Appendices D and E, respectively, of the WJ V Technical Manual for six broad age group samples.

Figure 8.
Plot of WJ V COG GIA and ACH Cluster W-Score Difference Curves by Age

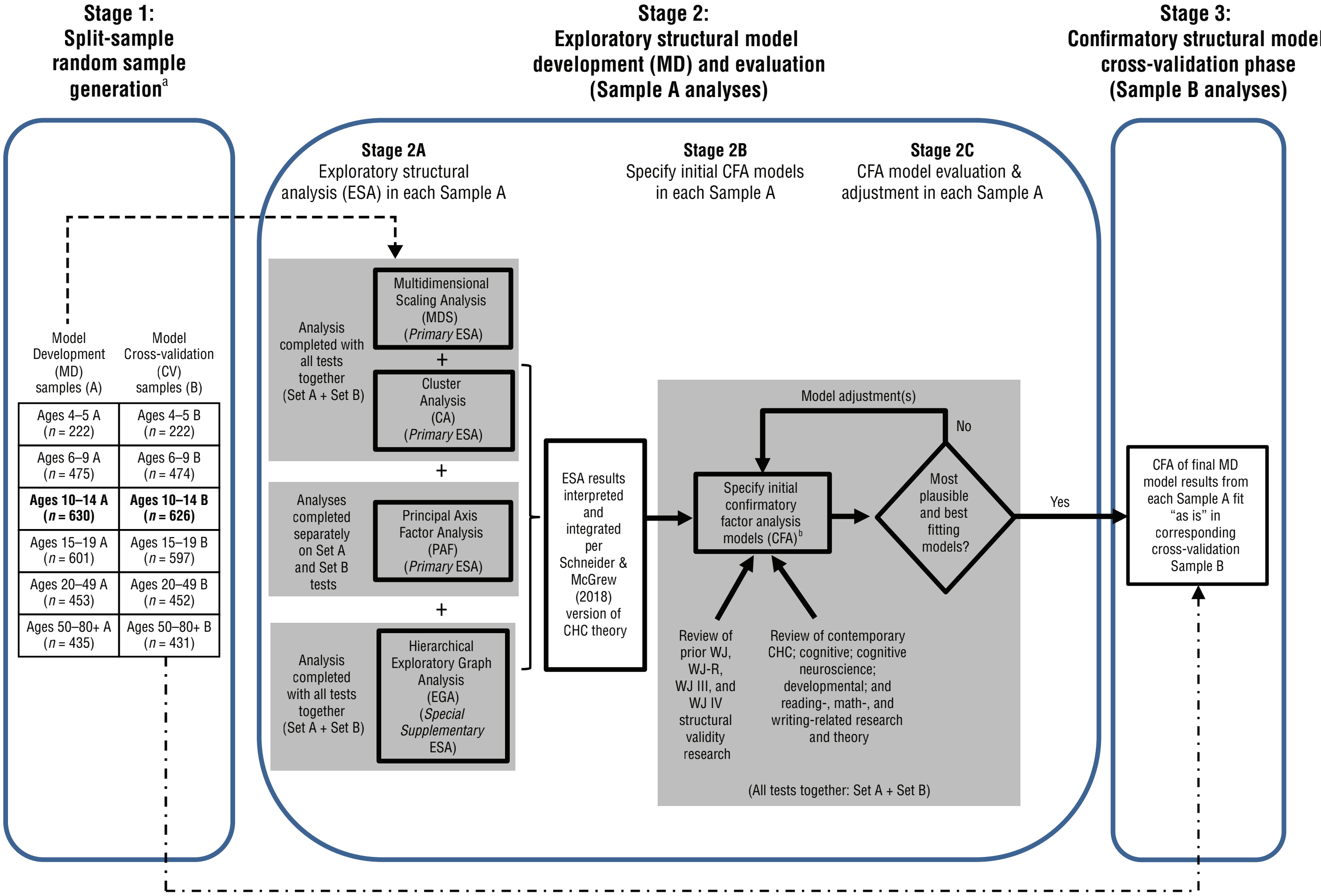


Multivariate statistical procedures (e.g., cluster analysis, multidimensional scaling, exploratory and confirmatory factor analysis, psychometric network analysis) are particularly helpful in uncovering the statistical patterns of convergent and discriminative relations between the WJ V individual tests in the correlation matrices reported in Appendix D. The primary indicator of internal (structural) validity for educational and psychological tests is the "degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based" (AERA et al., 2014, p. 16). The WJ V structural analysis has the advantage of building on the structural analyses reported in the WJ (Woodcock, 1978), WJ-R (McGrew et al., 1991), WJ III (McGrew & Woodcock, 2001), and WJ IV (McGrew et al., 2014) technical manuals as well as independent analysis of some of the WJ batteries by Carroll (1993, 2003). A series of exploratory multidimensional scaling (MDS), cluster analysis (CA), principal axis factor (PAF), psychometric network analysis (exploratory graph analysis, PNA-EGA), and confirmatory factor analysis (CFA) procedures were conducted during the early stages of WJ V data collection. These formative analyses (N = 3,910) were completed at ages 4 to 5, 6 to 9, 10 to 19, 20 to 39, and 40 to 80+. The objective of the formative analyses was to determine how the new or revised tests related to the broad CHC abilities. This step, which also included logical and psychological content analyses of the tests, informed decisions regarding the possible modification or elimination of tests. At this point, the relationship between the WJ V tests and the CHC broad factors, and a preliminary WJ V organizational structure, was established.

When norming data collection was complete ($N = 5{,}837$), a unique three-stage structural validity analysis framework was used to evaluate the structural validity of the WJ V tests and clusters. These procedures were like those employed in the WJ IV structural validity analysis (McGrew et al., 2014). The three-stage process that was used to investigate the internal structural validity of the WJ V battery and to establish the final organizational structure of the assessment system is displayed in Figure 9. A summary of the process is provided here; however, readers are encouraged to consult Chapter 6 of the WJ V Technical Manual for further details.

Figure 9.

Three-Stage Structural Validity Procedures for the WJ V Battery



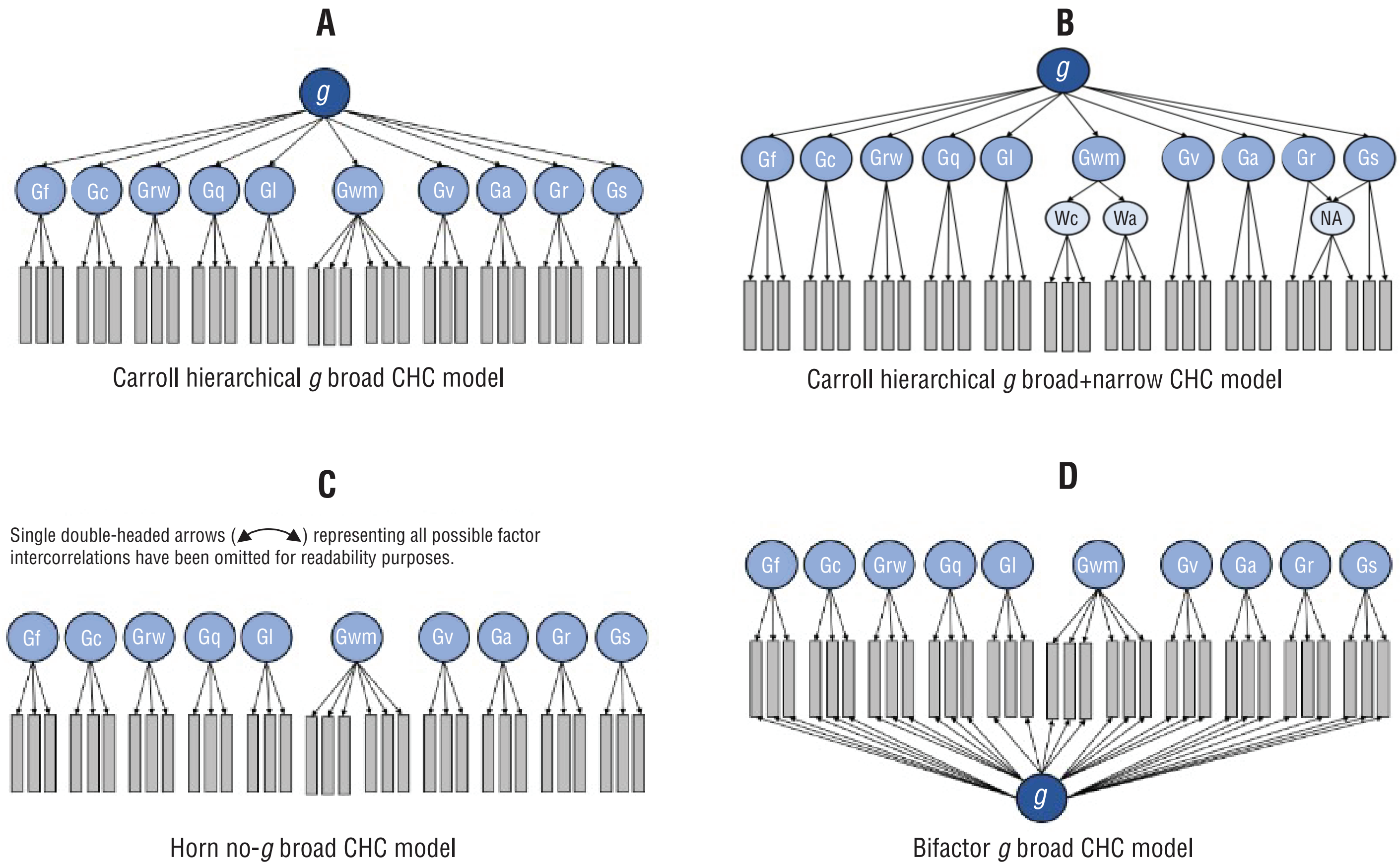
Note. Bold font designates initial target age (10 to 14 years) sample.

[a] "Quick norms" *W*-Difference scores were used as a proxy for age-based standard scores for all structural analyses. See the WJ V Technical Manual for more details.

[b] Three classes of models were evaluated: (1) Carroll hierarchical *g* broad CHC, (2) Carroll hierarchical *g* broad+narrow CHC, and (3) Horn no-*g* broad CHC.

As illustrated in Figure 9, the WJ V norming sample was divided into six age-differentiated groups. Each sample was randomly split into model development (MD; sample A) and model cross-validation (CV; sample B) samples of approximately equal size (see Stage 1 in Figure 9). Each of the six MD samples was analyzed with four different exploratory multivariate methods—multidimensional scaling (MDS) analysis, cluster analysis (CA), principal axis factor analysis (PAF), and hierarchical exploratory graph analysis (EGA) (see Stage 2A in Figure 9). The use of four methodological lenses allows for the detailed exploration of the relations among the complete collection of WJ V tests. These analyses were first conducted using the ages 10 to 14 sample as an exemplar. The results from this age group were then used as an approximate starting model (see Stage 1 in Figure 9) 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 MDS, CA, PAF, and EGA results from Stage 2A. A comprehensive review of contemporary CHC and neuroscience research as well as structural validity research on all four prior editions of the Woodcock-Johnson tests was integrated with the exploratory results from Stage 2A to specify the initial WJ V MG CFA models (see Stage 2B in Figure 9). The synthesis of the complete set of exploratory structural analyses was used to specify the initial mapping of WJ V tests to three[10] major families of CHC structural models (Carroll, 2003; McGrew et al., 2023) that vary primarily in the specification (or lack thereof) of a psychometric g factor. The different families of CHC models are conceptually represented in Figure 10.

Figure 10.
Conceptual Representation of Four Types of CHC Structural Models



Note. Rectangles represent tests, colored ovals represent latent factors, and arrows represent latent factor loadings.

After evaluating these models in Stage 2B, it was concluded that all three models were plausible explanations of the CHC structure of the WJ V system of tests across ages. These findings support the stability and generalizability of the structural validity of the WJ V measurement model operationalized as three types of CHC models. In Stage 3, the two models were taken "as is" and cross-validated with the exemplar age group CV sample (see Figure 9). The WJ V CFA

[10] Note that Model D in Figure 10 (bifactor g broad CHC model) was not among the models specified in Stage 2B in Figure 9. In this model, the variance associated with a dominant psychometric g factor is first extracted from all individual tests. The residual variance is modeled as 10 uncorrelated (orthogonal) CHC broad factors. As noted by Reynolds and Keith (2017), "bifactor models may serve as a useful *mathematical convenience* for partitioning variance in test scores" (p. 45; emphasis added). The bifactor g model preordains "that the statistically significant lion's share of IQ battery test variance must be of the form of a dominant psychometric g factor (Decker et al., 2021)" (McGrew et al., 2023, p. 3). Of the four families of CHC structural models, it is the model that most supports the statistical and conceptual importance of general intelligence and the preeminence of the full-scale or global IQ score over broad CHC test scores (e.g., see Dombrowski et al., 2019; Farmer, McGill, Dombrowski, & Canivez, 2020; Farmer et al., 2021)—a theoretical position inconsistent with the position of the WJ V authors and with Woodcock's legacy.

models were evaluated for overall statistical model fit and for size, statistical significance, and interpretability of all model parameter estimates (Brown, 2006). Although the Horn no-g broad CHC model might be the preferred model per the parsimony principle,[11] the more complex (defined by degrees of freedom) Carroll hierarchical g broad+narrow CHC model offers potentially important insights regarding the structure of the WJ V battery, possible clinically relevant interpretations, and potential new insights into the CHC theories of intelligence. It is the position of the WJ V authors, based on nearly three decades of CHC-theory-based refinement from the WJ-R through the WJ V and the extant factor structure research supporting the CHC model as the consensus structure of human cognitive abilities (Carroll, 1993; Euler et al., 2023; McGrew, 1997, 2005, 2009; Schneider & McGrew, 2012, 2018), that the CHC structural model—with a primary focus on the validity of the broad CHC constructs—is the model for which structural validity information should be provided in the WJ V Technical Manual.

The CHC test factor loadings for the Horn no-g broad CHC model (Model C in Figure 10) and the broad CHC ability and test factor loadings for the Carroll hierarchical g broad CHC model (Model A in Figure 10) are presented in Tables 20 and 21, respectively. In these tables, the significant cross validation (CV, Sample B) model parameters are presented for each of the six different age groups followed by the median value across all age groups.

Table 20.
*Test Factor Loadings for CHC Ability Factors for the Horn No-*g *Broad CHC Model in Cross-Validation (CV) Samples by Age Group*

Test	CHC Broad Ability	Age 4–5 (*n* = 222)	6–9 (*n* = 474)	10–14 (*n* = 626)	15–19 (*n* = 597)	20–49 (*n* = 452)	50–80+ (*n* = 431)	Median Factor Loading
Oral Vocabulary	Gc	0.83	0.88	0.85	0.87	0.90	0.90	0.87
Academic Facts	Gc	0.90	0.88	0.87	0.86	0.87	0.88	0.88
Academic Vocabulary	Gc	0.87	0.86	0.85	0.86	0.89	0.87	0.86
General Information	Gc	0.74	0.72	0.76	0.78	0.83	0.79	0.77
Oral Comprehension	Gc	0.68	0.71	0.75	0.74	0.78	0.73	0.74
Picture Vocabulary	Gc	0.76	0.67	0.70	0.72	0.79	0.74	0.73
Verbal Analogies	Gc	0.56	0.55	0.60	0.44	0.63	0.54	0.56
Oral Language Samples	Gc	—	***0.11***	—	—	0.36	0.40	0.36
Paragraph Reading Comprehension	Gc		0.20	0.39	0.37	0.36	0.39	0.37
Passage Comprehension	Gc		0.32	0.32	0.26	0.32	—	0.32
Sentence Repetition	Gc	—	—	0.49	—	—	—	0.49
Concept Formation	Gf	0.65	0.74	0.80	0.79	0.78	0.77	0.78
Analysis-Synthesis	Gf	0.64	0.70	0.71	0.74	0.75	0.76	0.72
Matrices	Gf	0.54	0.63	0.68	0.69	0.72	0.73	0.69
Understanding Directions	Gf	0.67	0.35	0.52	0.27	0.39	0.39	0.39
Visual-Auditory Learning	Gf	—	0.39	0.28	0.28	—	0.28	0.28
Verbal Analogies	Gf	0.26	0.26	0.21	0.37	0.25	0.28	0.26
Math Problem Identification	Gf	—	—	0.21	0.22	0.25	—	0.22
Spatial Relations	Gv	0.79	0.75	0.82	0.79	0.79	0.82	0.79
Block Rotation	Gv	0.64	0.79	0.74	0.81	0.78	0.73	0.76
Visual-Auditory Learning	Gv	0.67	0.34	0.36	0.37	0.65	0.41	0.39
Visual Working Memory	Gv	—	—	0.25	0.29	0.36	0.44	0.32
Rapid Quantity Naming	Gv	—	0.26	0.25	0.25	0.19	0.23	0.25

[11] Also known as Occam's Razor, the parsimony principle states that "given two models with similar fit to the data, the simpler model is preferred" (Kline, 2011, p. 102).

Table 20. *(cont.)*
*Test Factor Loadings for CHC Ability Factors for the Horn No-*g *Broad CHC Model in Cross-Validation (CV) Samples by Age Group*

Test	CHC Broad Ability	Age 4–5 (n = 222)	6–9 (n = 474)	10–14 (n = 626)	15–19 (n = 597)	20–49 (n = 452)	50–80+ (n = 431)	Median Factor Loading
Letter-Pattern Matching	Gv	—	0.11	0.07	0.10	0.13	0.11	0.11
Sound Blending	Ga	0.61	0.58	0.65	0.71	0.73	0.70	0.68
Sound Reversal	Ga		0.66	0.66	0.65	0.68	0.70	0.66
Segmentation	Ga	0.65	0.63	0.65	0.60	0.67	0.71	0.65
Spelling of Sounds	Ga	0.66	0.69	0.35	0.55	0.43	0.54	0.54
Nonsense Word Repetition	Ga	—	0.37	0.41	0.31	0.52	0.58	0.41
Word Attack	Ga	0.14	0.26	0.23	0.29	0.13	0.18	0.21
Rapid Phoneme Naming	Ga		—	***0.03***	0.16	0.26	—	0.16
Sound Deletion	Ga	0.67	0.77					0.72
Sound Substitution	Ga	0.69	0.77					0.73
Story Recall	Gl	0.77	0.75	0.72	0.70	0.80	0.78	0.76
Story Comprehension	Gl		0.75	0.75	0.78	0.81	0.72	0.75
Oral Language Samples	Gl	0.70	0.49	0.65	0.66	0.29	0.23	0.57
Phonemic Word Retrieval	Gr	0.60	0.82	0.79	0.74	0.76	0.78	0.77
Semantic Word Retrieval	Gr	0.78	0.69	0.66	0.76	0.74	0.70	0.72
Rapid Phoneme Naming	Gr		0.45	0.52	0.47	0.44	0.65	0.47
Rapid Letter Naming	Gr	***−0.02***	0.18	0.27	0.24	0.39	0.28	0.26
Rapid Picture Naming	Gr	0.18	0.19	0.19	0.25	0.28	0.37	0.22
Rapid Number Naming	Gr	—	—	0.15	***0.10***	0.28	0.15	0.15
Verbal Attention	Gwm	0.83	0.78	0.78	0.82	0.81	0.76	0.79
Animal-Number Sequencing	Gwm	0.65	0.72	0.75	0.78	0.74	0.76	0.75
Numbers Reversed	Gwm	0.78	0.77	0.72	0.71	0.71	0.78	0.75
Memory for Words	Gwm	0.62	0.64	0.60	0.68	0.62	0.65	0.63
Sentence Repetition	Gwm	0.56	0.60	0.19	0.60	0.62	0.58	0.59
Understanding Directions	Gwm	—	0.38	0.23	0.50	0.45	0.38	0.38
Visual Working Memory	Gwm	0.48	0.51	0.34	0.36	0.27	0.22	0.35
Nonsense Word Repetition	Gwm	0.57	0.20	0.19	0.32	***0.15***	—	0.20
Symbol Inhibition	Gwm		—	***0.07***	***0.03***	0.15	0.18	0.11
Number-Pattern Matching	Gs	0.54	0.70	0.77	0.84	0.83	0.85	0.80
Letter-Pattern Matching	Gs	0.54	0.64	0.71	0.74	0.73	0.73	0.72
Word Reading Fluency	Gs		0.48	0.65	0.73	0.73	0.67	0.67
Magnitude Comparison	Gs	0.76	0.58	0.62	0.64	0.65	0.67	0.64
Sentence Reading Fluency	Gs		0.51	0.61	0.65	0.67	0.61	0.61
Symbol Inhibition	Gs		0.57	0.55	0.66	0.58	0.54	0.57
Rapid Number Naming	Gs	0.55	0.53	0.51	0.58	0.47	0.56	0.54
Rapid Quantity Naming	Gs	0.52	0.41	0.53	0.57	0.60	0.52	0.53
Rapid Letter Naming	Gs	0.52	0.47	0.49	0.52	0.40	0.54	0.51
Sentence Writing Fluency	Gs		0.37	0.47	0.51	0.50	0.40	0.47
Math Facts Fluency	Gs		0.40	0.42	0.51	0.52	0.46	0.46
Rapid Picture Naming	Gs	0.20	0.32	0.41	0.40	0.41	0.33	0.36
Letter Writing Fluency	Gs	0.67						0.67
Applied Problems	Gq	0.88	0.86	0.88	0.89	0.89	0.88	0.88

Table 20. *(cont.)*
*Test Factor Loadings for CHC Ability Factors for the Horn No-*g *Broad CHC Model in Cross-Validation (CV) Samples by Age Group*

Test	CHC Broad Ability	Age 4–5 (*n* = 222)	6–9 (*n* = 474)	10–14 (*n* = 626)	15–19 (*n* = 597)	20–49 (*n* = 452)	50–80+ (*n* = 431)	Median Factor Loading
Number Sense	Gq	0.88	0.87	0.86	0.85	0.86	0.89	0.87
Calculation	Gq	0.72	0.80	0.84	0.86	0.82	0.81	0.81
Number Series	Gq	0.78	0.79	0.76	0.81	0.81	0.80	0.79
Math Problem Identification	Gq	0.86	0.87	0.70	0.71	0.67	0.88	0.79
Math Facts Fluency	Gq		0.32	0.38	0.31	0.31	0.38	0.32
Magnitude Comparison	Gq	—	0.14	0.21	0.28	0.25	0.23	0.23
Letter-Word Identification	Grw	0.94	0.96	0.87	0.89	0.91	0.91	0.91
Spelling	Grw	0.89	0.90	0.82	0.83	0.87	0.85	0.86
Oral Reading	Grw		0.85	0.82	0.83	0.84	0.84	0.84
Sentence Writing Accuracy	Grw		0.80	0.81	0.76	0.80	0.81	0.80
Written Language Samples	Grw		0.79	0.73	0.70	0.74	0.77	0.74
Word Attack	Grw	0.75	0.62	0.62	0.60	0.72	0.68	0.65
Reading Recall	Grw		0.72	0.59	0.59	0.63	0.60	0.60
Paragraph Reading Comprehension	Grw		0.63	0.49	0.51	0.49	0.71	0.51
Passage Comprehension	Grw		0.60	0.47	0.48	0.53	0.45	0.48
Spelling of Sounds	Grw	—	—	0.38	0.18	0.38	0.20	0.29
Sentence Writing Fluency	Grw		0.26	0.23	0.22	0.23	0.35	0.23
Sentence Reading Fluency	Grw		0.22	0.21	0.21	0.26	0.27	0.22
Word Reading Fluency	Grw		0.30	0.15	0.09	0.16	0.18	0.16

Notes. Gray cells designate tests that were not included in the analyses at specific age groups. Blank cells designate tests that were included in the CV (Sample B) model but that were not specified on the CHC factor (based on the MD Sample A final model). Bold, italic font designates parameters that were significant in the model development sample (A) but not in the cross-validation sample (B).

Table 21.
CHC Broad Ability and Test Factor Loadings for CHC Factors in the Carroll Hierarchical g *Broad CHC Model in Cross-Validation (CV) Samples by Age Group*

CHC Broad Ability/Test	CHC Broad Ability	Age 4–5 (*n* = 222)	6–9 (*n* = 474)	10–14 (*n* = 626)	15–19 (*n* = 597)	20–49 (*n* = 452)	50–80+ (*n* = 431)	Median Factor Loading
Gc (Comprehension-Knowledge)	*g*	0.86	0.91	0.89	0.91	0.92	0.94	0.91
Gf (Fluid Reasoning)	*g*	0.91	0.89	0.87	0.89	0.87	0.89	0.89
Gq (Quantitative Knowledge)	*g*	0.89	0.93	0.89	0.88	0.85	0.86	0.89
Gwm (Auditory Working Memory Capacity)	*g*	0.92	0.89	0.82	0.87	0.88	0.83	0.87
Ga (Auditory Processing)	*g*	0.82	0.86	0.83	0.74	0.84	0.84	0.84
Gl (Long-Term Storage)	*g*	0.82	0.81	0.83	0.83	0.84	0.54	0.82
Grw (Reading & Writing)	*g*	0.63	0.79	0.80	0.78	0.81	0.86	0.80
Gv (Visual Processing)	*g*	0.72	0.70	0.74	0.72	0.69	0.70	0.71
Gr (Retrieval Fluency)	*g*	0.81	0.68	0.67	0.64	0.73	0.68	0.68
Gs (Cognitive Processing Speed)	*g*	0.72	0.58	0.58	0.65	0.60	0.57	0.59
Academic Facts	Gc	0.90	0.88	0.87	0.87	0.88	0.88	0.88
Oral Vocabulary	Gc	0.83	0.88	0.85	0.86	0.89	0.90	0.87
Academic Vocabulary	Gc	0.87	0.85	0.85	0.86	0.89	0.87	0.86

Table 21. *(cont.)*
CHC Broad Ability and Test Factor Loadings for CHC Factors in the Carroll Hierarchical g *Broad CHC Model in Cross-Validation (CV) Samples by Age Group*

CHC Broad Ability/Test	CHC Broad Ability	Age 4–5 ($n = 222$)	Age 6–9 ($n = 474$)	Age 10–14 ($n = 626$)	Age 15–19 ($n = 597$)	Age 20–49 ($n = 452$)	Age 50–80+ ($n = 431$)	Median Factor Loading
General Information	Gc	0.74	0.72	0.76	0.78	0.84	0.80	0.77
Picture Vocabulary	Gc	0.77	0.67	0.70	0.72	0.80	0.75	0.74
Oral Comprehension	Gc	0.68	0.71	0.75	0.73	0.78	0.73	0.73
Verbal Analogies	Gc	0.51	0.52	0.63	0.48	0.63	0.65	0.58
Passage Comprehension	Gc		0.31	0.40	0.36	0.36	0.38	0.36
Oral Language Samples	Gc	—	***0.09***	0.35	0.23	0.28	0.34	0.28
Paragraph Reading Comprehension	Gc		0.17	0.30	0.23	0.32	—	0.27
Sentence Repetition	Gc	—	—	0.48	—	—	—	0.48
Concept Formation	Gf	0.63	0.73	0.79	0.80	0.78	0.80	0.79
Analysis-Synthesis	Gf	0.65	0.71	0.71	0.73	0.75	0.76	0.72
Matrices	Gf	0.54	0.63	0.68	0.69	0.72	0.71	0.68
Understanding Directions	Gf	0.67	0.37	0.53	0.28	0.38	0.43	0.41
Visual-Auditory Learning	Gf	***0.07***	0.44	0.28	0.29	0.35	0.26	0.28
Verbal Analogies	Gf	0.31	0.29	0.18	0.32	0.25	0.17	0.27
Math Problem Identification	Gf	—	—	0.26	0.27	0.27	0.28	0.27
Spatial Relations	Gv	0.82	0.78	0.81	0.81	0.82	0.82	0.82
Block Rotation	Gv	0.64	0.78	0.74	0.80	0.79	0.73	0.76
Visual-Auditory Learning	Gv	0.60	0.32	0.40	0.38	0.38	0.46	0.39
Visual Working Memory	Gv	—	—	0.37	0.31	0.33	0.44	0.35
Rapid Quantity Naming	Gv	—	0.20	0.23	0.23	0.15	0.22	0.22
Letter-Pattern Matching	Gv	—	0.09	0.07	0.11	0.12	0.10	0.10
Sound Reversal	Ga		0.67	0.68	0.67	0.71	0.74	0.68
Segmentation	Ga	0.70	0.65	0.63	0.60	0.69	0.73	0.67
Sound Blending	Ga	0.62	0.58	0.63	0.69	0.71	0.68	0.65
Spelling of Sounds	Ga	0.63	0.45	0.37	0.55	0.42	0.44	0.45
Nonsense Word Repetition	Ga	—	0.33	0.42	0.27	0.43	0.24	0.33
Word Attack	Ga	0.22	0.22	0.24	0.30	0.17	0.17	0.22
Rapid Phoneme Naming	Ga		—	***0.10***	0.23	0.33	***0.11***	0.17
Sound Deletion	Ga	0.64	0.76					0.70
Sound Substitution	Ga	0.67	0.78					0.72
Story Recall	Gl	0.78	0.78	0.76	0.73	0.78	0.77	0.77
Story Comprehension	Gl		0.71	0.76	0.77	0.81	0.73	0.76
Oral Language Samples	Gl	0.69	0.53	0.33	0.45	0.39	0.31	0.42
Phonemic Word Retrieval	Gr	0.70	0.83	0.79	0.77	0.81	0.81	0.80
Semantic Word Retrieval	Gr	0.70	0.68	0.66	0.77	0.72	0.71	0.70
Rapid Phoneme Naming	Gr		0.44	0.47	0.39	0.38	0.53	0.44
Rapid Letter Naming	Gr	0.16	0.18	0.28	0.18	0.34	0.27	0.23
Rapid Picture Naming	Gr	***0.14***	0.18	0.21	0.16	0.24	0.33	0.20
Rapid Number Naming	Gr	—	—	0.13	—	0.22	0.15	0.15
Verbal Attention	Gwm	0.83	0.79	0.78	0.82	0.82	0.78	0.80
Animal-Number Sequencing	Gwm	0.65	0.72	0.76	0.78	0.74	0.78	0.75
Numbers Reversed	Gwm	0.78	0.77	0.71	0.71	0.71	0.79	0.74

Table 21. *(cont.)*
CHC Broad Ability and Test Factor Loadings for CHC Factors in the Carroll Hierarchical g *Broad CHC Model in Cross-Validation (CV) Samples by Age Group*

CHC Broad Ability/Test	CHC Broad Ability	Age						Median Factor Loading
		4–5 (*n* = 222)	6–9 (*n* = 474)	10–14 (*n* = 626)	15–19 (*n* = 597)	20–49 (*n* = 452)	50–80+ (*n* = 431)	
Memory for Words	Gwm	0.63	0.64	0.60	0.68	0.62	0.68	0.64
Sentence Repetition	Gwm	0.57	0.61	0.19	0.60	0.62	0.60	0.60
Understanding Directions	Gwm	—	0.37	0.21	0.49	0.46	0.36	0.37
Visual Working Memory	Gwm	0.48	0.50	—	0.35	0.31	0.22	0.35
Nonsense Word Repetition	Gwm	0.55	0.23	0.20	0.36	0.23	0.41	0.30
Symbol Inhibition	Gwm		—	***0.08***	***0.05***	0.14	0.16	0.11
Number-Pattern Matching	Gs	0.58	0.71	0.79	0.54	0.85	0.85	0.75
Letter-Pattern Matching	Gs	0.59	0.65	0.72	0.73	0.73	0.74	0.72
Word Reading Fluency	Gs		0.51	0.63	0.74	0.72	0.67	0.67
Magnitude Comparison	Gs	0.78	0.59	0.64	0.64	0.66	0.67	0.65
Sentence Reading Fluency	Gs		0.52	0.60	0.65	0.67	0.61	0.61
Symbol Inhibition	Gs		0.58	0.53	0.65	0.58	0.57	0.58
Rapid Number Naming	Gs	0.53	0.50	0.54	0.64	0.54	0.57	0.54
Rapid Quantity Naming	Gs	0.52	0.39	0.53	0.57	0.62	0.52	0.53
Rapid Letter Naming	Gs	0.35	0.46	0.51	0.58	0.48	0.57	0.50
Math Facts Fluency	Gs		0.37	0.41	0.51	0.51	0.46	0.46
Sentence Writing Fluency	Gs		0.36	0.45	0.50	0.50	0.41	0.45
Rapid Picture Naming	Gs	0.23	0.33	0.41	0.48	0.46	0.38	0.40
Letter Writing Fluency	Gs	0.64						0.64
Visual Working Memory	Gs	—	—	0.30	—	—	—	0.30
Applied Problems	Gq	0.87	0.86	0.88	0.89	0.89	0.88	0.88
Number Sense	Gq	0.89	0.87	0.87	0.86	0.86	0.89	0.87
Calculation	Gq	0.71	0.80	0.84	0.86	0.82	0.81	0.81
Number Series	Gq	0.78	0.79	0.76	0.81	0.81	0.80	0.80
Math Problem Identification	Gq	0.85	0.87	0.66	0.67	0.67	0.65	0.67
Math Facts Fluency	Gq		0.34	0.39	0.32	0.31	0.40	0.34
Magnitude Comparison	Gq	—	0.13	0.20	0.27	0.23	0.23	0.23
Letter-Word Identification	Grw	0.98	0.96	0.88	0.90	0.91	0.91	0.91
Spelling	Grw	0.84	0.89	0.81	0.83	0.87	0.84	0.84
Oral Reading	Grw		0.86	0.82	0.83	0.84	0.84	0.84
Sentence Writing Accuracy	Grw		0.80	0.81	0.77	0.80	0.81	0.80
Written Language Samples	Grw		0.79	0.75	0.71	0.75	0.78	0.75
Word Attack	Grw	0.71	0.68	0.63	0.60	0.70	0.70	0.69
Reading Recall	Grw		0.74	0.62	0.61	0.64	0.62	0.62
Paragraph Reading Comprehension	Grw		0.66	0.51	0.54	0.50	0.72	0.54
Passage Comprehension	Grw		0.62	0.47	0.49	0.53	0.47	0.49
Spelling of Sounds	Grw	—	0.27	0.38	0.21	0.39	0.33	0.33
Sentence Writing Fluency	Grw		0.26	0.24	0.21	0.22	0.33	0.24
Sentence Reading Fluency	Grw		0.22	0.23	0.21	0.26	0.27	0.23
Word Reading Fluency	Grw		0.29	0.17	0.09	0.17	0.18	0.17

Notes. g = psychometric *g* (see text). Gray cells designate tests that were not included in the analyses at specific age groups. Blank cells designate tests that were included in the CV (Sample B) model but that were not specified on the CHC factor (based on the MD Sample A final model). Bold, italic font designates parameters that were significant in the model development sample (A) but not in the cross-validation sample (B).

The test-to-CHC broad factor measurement models for the Horn no-g broad CHC and Carroll hierarchical g broad CHC models were nearly identical. Whether the correlations between the broad CHC latent factors are specified as latent variable correlations or are accounted for by the specification of a higher-order psychometric g factor makes little to no difference in the CHC broad ability interpretations for the WJ V tests. Thus, in the discussion and interpretation of the CHC characteristics of the WJ V tests, only one of these two models is necessary; the Horn no-g broad CHC model (Model C in Figure 10) serves as the *primary WJ V test-broad CHC measurement model*.[12]

Relationship of WJ V Scores to Other Measures

The necessary and sufficient conditions for construct validity are met when structurally valid measures display expected convergent and divergent relationships with measures of constructs external to the focal measures (Benson, 1998; Benson & Hagtvet, 1996; Cronbach & Meehl, 1955). A summary of the results from several concurrent validity studies are presented. These studies examined the relationships between WJ V COG, ACH, and VTL cluster scores (and some tests) with select composites and subtests from other commercially available cognitive and achievement test batteries. Key correlations[13] are presented in this section; refer to the WJ V Technical Manual (LaForte et al., 2025) for complete study results.

Correlations for the WJ V With Other Measures of Intelligence

The WJ V COG scores were examined in five studies that included the other commercially available intelligence batteries. Table 22 presents correlations for the WJ V COG GIA (*g*), Gf-Gc Composite, and Brief Intellectual Ability (BIA) cluster scores with the composite measures of general intelligence (*g*) from the external measures. The .66 to .85 correlations for the WJ V GIA cluster with the general intelligence total scores from the other intelligence batteries support the conclusion that the WJ V 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 V 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 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 22.
Correlations for Select WJ V COG Measures and Other Measures of Cognitive Abilities

Other Measures	*N*	WJ V COG General Intellectual Ability (GIA)	WJ V COG Gf-Gc Composite	WJ V COG Brief Intellectual Ability (BIA)
Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV) Full-Scale IQ (FSIQ)	100[a]	0.85	0.83	0.81
Wechsler Intelligence Scale for Children–Fifth Edition (WISC-V) Full-Scale IQ (FSIQ)	97[b]	0.82	0.85	0.85
Wechsler Preschool and Primary Scale of Intelligence–Fourth Edition (WPPSI-IV) Full-Scale IQ (FSIQ)	47[c]	0.77	0.68	0.73
Kaufman Assessment Battery for Children–Second Edition Normative Update (KABC-II NU) Mental Processing Index (*g*)	49[d]	0.79	0.82	0.77
Reynolds Intellectual Assessment Scales–Second Edition (RIAS-2) Composite Intelligence Index (*g*)	51[e]	0.66	0.76	0.73

[a] Pairwise comparisons based on 98–100 examinees.
[b] Pairwise comparisons based on 92–97 examinees.
[c] Pairwise comparisons based on 43–44 examinees.
[d] Pairwise comparisons based on 47–49 examinees.
[e] Pairwise comparisons based on 48–51 examinees.

[12] The more complex Carroll hierarchical g broad+narrow CHC model (Model B in Figure 10) does, however, offer potentially important WJ V structural insights that may be relevant for WJ V clinical interpretation. See the WJ V Technical Manual (LaForte et al., 2025) for the structural analysis results and a complete discussion of this model.
[13] The results of the concurrent validity study for the MMSE-2 and DRS-2 with an older adult (50+) sample are not included in this document due to space constraints but are reported in detail in the WJ V Technical Manual (LaForte et al., 2025).

Correlations for Select WJ V COG and VTL Cluster Scores With Other Measures of Phonological Processing and Rapid Automatized Naming

A single study investigated the concurrent correlations between the WJ V phonological awareness (Ga and select Gwm abilities), rapid automatized naming (RAN; Gs/Gr), and select tests from two external batteries measuring similar CHC-classified abilities. The CHC classifications of the external measure variables were based on the WJ V authors' CHC task analysis of the tests and the Gs/Gr classification of similar rapid naming tests in the WJ V structural validity analysis. Results are presented in Table 23.

The WJ V Phonological Awareness (Ga) and Phonological Manipulation (Ga) clusters, as expected, displayed moderate correlations (.35 to .66) with the CTOPP-2 Phonological Awareness (Ga), Alternative Phonological Awareness (Ga), and Phonological Memory (Gwm/Ga) composite scores. The WJ V RAN–Reading and RAN–Math clusters (Gs/Gr) displayed moderate to strong correlations with the CTOPP-2 Rapid Symbolic Naming Composite. As expected, the WJ V Auditory Memory Span cluster (Gwm-Wa) correlated the highest with the CTOPP-2 Phonological Memory Composite.

Table 23.
Summary Statistics and Correlations for WJ V Auditory and RAN Measures With CTOPP-2 and RAN/RAS Measures

	CTOPP-2 Composites				RAN/RAS Tests (Gs/Gr)					
WJ V Clusters	Phonological Awareness (Ga)	Alternative Phonological Awareness (Ga)	Phonological Memory (Gwm/Ga)	Rapid Symbolic Naming (Gs/Gr)	Objects (Gs/Gr)	Colors (Gs/Gr)	Numbers (Gs/Gr)	Letters (Gs/Gr)	2-Set Letters & Numbers (Gs/Gr)	2-Set Letters & Numbers & Colors (Gs/Gr)
Phonological Awareness (Ga)	0.60	0.56	0.45	0.29	–0.04	–0.05	–0.12	0.11	0.02	–0.08
Phonological Manipulation (Ga)	0.66	0.63	0.35	0.32	0.08	–0.02	0.02	0.22	0.23	0.15
RAN–Reading (Gs/Gr)	0.47	0.36	0.20	0.69	0.52	0.55	0.58	0.69	0.56	0.55
RAN–Math (Gs/Gr)	0.42	0.48	0.27	0.75	0.53	0.72	0.66	0.64	0.51	0.50
Auditory Memory Span (Gwm)	0.37	0.52	0.60	0.00	0.14	–0.02	0.14	0.17	0.10	0.08

Notes. CTOPP-2 sample *n* = 47; pairwise correlations based on 46–47 examinees. RAN/RAS sample *n* = 41; pairwise correlations based on 40–41 examinees. Shading indicates cells in which correlations are expected to be the highest, based on shared CHC content classifications.

Correlations for Select WJ V ACH Clusters With Other Measures of Achievement

The WJ V ACH scores were examined in four school-age samples of examinees who were administered two commercially available external measures of achievement. Table 24 presents the correlations between the WJ V achievement and oral language clusters and select tests and composites from the KTEA-3 (ages 8 to 12 and 13 to 18) and the WIAT-4 (ages 6 to 12 and 13 to 18). The pattern of convergent and discriminative WJ V validity correlations in the KTEA-3 samples supports the interpretation of the primary WJ V reading, math, writing, and oral language clusters for both age groups. Likewise, in both WIAT-4 samples, the pattern of convergent and discriminative WJ V validity correlations supports the interpretation of the WJ V reading, math, and writing clusters.

Table 24.

Correlations for Select WJ V ACH Measures and Other Measures of Achievement

	WJ V Reading Clusters				WJ V Math Clusters				WJ V Writing Clusters		WJ V Oral Language Clusters				WJ V Cross-Domain Clusters			
Other Measures	**Brief Reading**	**Basic Reading Skills**	**Reading Comprehension**	**Reading Fluency**	**Basic Math**	**Math Calculation Skills**	**Math Problem Solving**	**Number Concepts**	**Brief Writing**	**Written Expression**	**Oral Language**	**Listening Comprehension**	**Oral Expression**	**Vocabulary**	**Academic Skills**	**Academic Fluency**	**Academic Applications**	**Phoneme-Grapheme Knowledge**
***Kaufman Tests of Educational Achievement, Third Edition* (KTEA-3)**																		
Ages 8 to 12 (*N* = 49)[a]																		
Reading Index	0.86	0.82	0.83	0.65	0.74	0.63	0.67	0.35	0.74	0.73	0.58	0.53	0.63	0.71	0.86	0.69	0.85	0.59
Math Index	0.64	0.67	0.61	0.64	0.94	0.89	0.83	0.58	0.65	0.70	0.44	0.39	0.52	0.58	0.80	0.70	0.79	0.61
Written Expression	0.67	0.57	0.70	0.56	0.72	0.73	0.59	0.36	0.78	0.77	0.55	0.49	0.62	0.65	0.69	0.60	0.82	0.48
Oral Expression	0.61	0.61	0.63	0.55	0.60	0.52	0.60	0.23	0.67	0.71	0.60	0.58	0.62	0.67	0.59	0.61	0.71	0.65
Listening Comprehension	0.63	0.62	0.64	0.65	0.66	0.54	0.71	0.46	0.70	0.63	0.77	0.75	0.75	0.71	0.67	0.61	0.74	0.55
Ages 13 to 18 (*N* = 49)[a]																		
Reading Index[a]	0.91	0.89	0.90	0.52	0.78	0.77	0.77	0.72	0.71	0.57	0.66	0.59	0.70	0.67	0.88	0.53	0.82	0.77
Math Index[a]	0.80	0.83	0.76	0.56	0.96	0.94	0.87	0.76	0.76	0.63	0.58	0.54	0.60	0.57	0.89	0.61	0.84	0.76
Written Expression	0.81	0.79	0.78	0.51	0.69	0.70	0.65	0.61	0.82	0.64	0.67	0.64	0.67	0.65	0.82	0.56	0.78	0.78
Oral Expression	0.66	0.53	0.68	0.32	0.45	0.40	0.47	0.50	0.57	0.53	0.54	0.49	0.58	0.49	0.53	0.37	0.63	0.53
Listening Comprehension	0.75	0.75	0.74	0.50	0.55	0.56	0.57	0.54	0.59	0.45	0.61	0.59	0.61	0.63	0.67	0.43	0.65	0.65
***Wechsler Individual Achievement Test, Fourth Edition* (WIAT-4)**																		
Ages 6 to 12 (*N* = 50)																		
Reading	0.81	0.80	0.79	0.52	0.56	0.57	0.52	0.35	0.74	0.63	0.58	0.48	0.66	—	0.81	0.60	0.73	0.78
Decoding	0.86	0.90	0.79	0.50	0.53	0.52	0.49	0.34	0.78	0.61	0.52	0.39	0.64	—	0.86	0.57	0.75	0.85
Mathematics	0.60	0.56	0.71	0.43	0.85	0.81	0.77	0.65	0.69	0.68	0.62	0.53	0.63	—	0.73	0.59	0.79	0.58
Math Fluency	0.34	0.36	0.39	0.40	0.65	0.79	0.60	0.61	0.44	0.64	0.43	0.41	0.42	—	0.47	0.63	0.58	0.34
Oral Language	0.62	0.57	0.74	0.57	0.67	0.66	0.61	0.54	0.69	0.54	0.76	0.69	0.75	—	0.69	0.61	0.74	0.56
Ages 13 to 18 (*N* = 49)[b]																		
Reading	0.92	0.90	0.75	0.50	0.73	0.60	0.77	0.62	0.79	0.61	0.74	0.62	0.77	—	0.86	0.58	0.80	0.78
Decoding	0.89	0.95	0.58	0.40	0.61	0.51	0.64	0.52	0.70	0.57	0.57	0.46	0.62	—	0.83	0.54	0.64	0.86
Mathematics	0.74	0.70	0.69	0.42	0.92	0.80	0.87	0.68	0.63	0.47	0.66	0.61	0.66	—	0.86	0.53	0.75	0.57
Math Fluency	0.51	0.63	0.31	0.48	0.71	0.85	0.65	0.67	0.38	0.42	0.33	0.28	0.36	—	0.77	0.63	0.45	0.54
Oral Language	0.79	0.63	0.63	0.43	0.61	0.40	0.70	0.55	0.71	0.57	0.81	0.79	0.74	—	0.67	0.48	0.80	0.46

Notes. KTEA-3 Written Expression, Oral Expression, and Listening Comprehension measures are subtests; all other external measures are indexes. Shading indicates cells in which correlations are expected to be the highest, based on shared achievement or oral language domain classifications.

[a] Pairwise correlations based on 47–49 examinees.

[b] Pairwise correlations based on 48–49 examinees.

Performance of Clinical Samples on the WJ V

The relationship between WJ V 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 V tests were administered to individuals within several clinical groups (see Tables 11 and 12). The comprehensiveness of the WJ V 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, as shown in Table 25. For example, the gifted and ID groups displayed large differences on all tests and clusters administered, with most WJ V cluster standard scores for the gifted group in the 113 to 116 range and most cluster standard scores for the ID/MR group in the 60 to 70 range. With some exceptions, the three LD groups displayed mean WJ V COG cluster scores and relevant ACH cluster scores that were generally 1 or more standard deviations below the mean. Complete results and interpretation of the WJ V clinical validity studies are presented in Chapter 6 of the WJ V Technical Manual (LaForte et al., 2025).

Table 25.
Standard Score Summary Statistics for Select WJ V Clusters for Clinical Validity Study Groups

	Gifted			ID			SLD–Reading			SLD–Writing			SLD–Math			Language Impairment			ADHD			ASD		
Clusters	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
General Intellectual Ability (GIA) (g)	73	113.4	11.1	20	58.9	24.3	70	85.1	13.9	15	84.3	15.0	29	80.3	14.3	23	87.5	26.0	37	95.8	12.9	39	82.0	20.6
Brief Intellectual Ability (BIA)	84	113.8	11.4	20	60.4	17.7	84	84.6	16.1	15	81.9	18.5	33	80.9	13.7	23	85.1	27.3	50	95.0	14.9	46	83.3	18.9
Comprehension-Knowledge (Gc)	85	113.6	12.0	20	63.1	22.1	88	85.3	17.5	15	86.6	17.9	37	83.8	15.4	23	85.7	24.3	50	97.2	13.8	48	88.9	19.3
Cognitive Efficiency (CE)	84	105.1	11.7	20	66.2	18.9	85	87.6	10.4	15	85.0	15.7	34	83.0	15.8	23	89.0	22.7	50	93.6	12.4	47	81.2	17.3
Academic Skills/Brief Achievement (Grw, Gq)	83	112.9	10.1	20	60.3	25.5	86	77.4	14.5	15	78.5	15.5	35	84.4	14.2	22	83.4	24.7	44	90.9	14.3	47	88.2	19.7
Gf-Gc Composite (Gf, Gc)	85	116.0	11.6	19	61.6	16.7							32	78.9	12.8									
Fluid Reasoning (Gf)	85	116.6	12.9	19	67.9	12.0							32	79.7	13.1									
Cognitive Processing Speed (Gs)							85	90.0	9.6	15	89.9	14.9	34	87.4	14.3				49	94.9	11.8			
Auditory Working Memory Capacity (Gwm)							86	85.4	13.0				36	83.3	16.0				50	94.5	15.1	48	82.6	22.6
Academic Knowledge (Gc)	85	114.5	9.6	20	61.3	29.4																		
Vocabulary (Gc)	85	112.3	10.3	20	65.7	30.1																		
Listening Comprehension (Gc)										15	88.6	23.2				23	87.8	18.5						
Long-Term Storage (Gl)										15	88.3	23.4				23	90.3	18.7						
RAN–Math (Gs, Gr)													36	89.4	13.1				50	88.7	9.3			
Brief Reading (Grw)							88	78.8	17.5															
Basic Reading Skills (Grw)							88	78.5	14.7															
Reading Fluency (Grw)							87	80.8	9.9															
Reading Comprehension (Grw)							85	81.4	16.1															
Brief Writing (Grw)										13	81.9	15.3												
Basic Writing Skills (Grw)										15	77.3	14.3												
Spelling Skills (Grw, Ga)										15	78.9	12.1												
Written Expression (Grw)										13	82.5	15.8												
Brief Math (Gq)													37	77.5	14.3									
Math Calculation Skills (Gq)													37	82.5	12.8									
Number Concepts (Gq)													36	84.6	9.7									
Math Problem Solving (Gq)													37	77.2	14.0									
Visual Processing (Gv)													32	85.8	15.2									
Retrieval Fluency (Gr)																24	95.3	18.7						
Phonemic Retrieval Fluency (Gr)																24	97.8	19.9						
Oral Language (Gc)																23	88.8	17.8						
Oral Expression (Gc, Gl)																24	89.2	22.0						

Notes. ID = intellectual disability, SLD = specific learning disability, ADHD = attention-deficit/hyperactivity disorder, ASD = autism spectrum disorder. Colored fonts refer to rounded standard-score integer values (e.g., 89.7 rounds to 90): purple ≤ 70, red = 70–79, blue = 80–89, black = 90–109, and green ≥ 109.

Summary

The procedures used to develop and validate the WJ V have produced a diagnostic system that can be used with confidence in a variety of settings. Throughout the design and development of the WJ V, 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 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 V Technical Manual was designed to provide test users with a comprehensive resource for evaluating the validity of the scores and interpretations from the WJ V battery for measuring an individual's level of functioning. Interested users should consult the WJ V Technical Manual (LaForte et al., 2025) for more in-depth details about the technical characteristics of the test.

References

Allen, M. J., & Yen, W. M. (2001). *Introduction to Measurement Theory*. Waveland Press.

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

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology, 107*(3), 868.

Baltes, P. B., Staudinger, U. M., & Lindenberger, U. (1999). Lifespan psychology: Theory and application to intellectual functioning. *Annual Review of Psychology, 50*(1), 471–507.

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

Benson, J., & Hagtvet, K. A. (1996). The interplay between design, data analysis and theory in the measurement of coping. In M. Zeidner & N. Endler (Eds.), *Handbook of coping: Theory, research, applications* (pp. 83–106). Wiley.

Berlin, L., Bohlin, G., & Rydell, A. M. (2004). Relations between inhibition, executive functioning, and ADHD symptoms: A longitudinal study from age 5 to 8½ years. *Child Neuropsychology, 9*(4), 255–266.

Berninger, V. W. (2020). *Revised user guide for PAL reading and writing, Second Edition (PAL II RW)*. Pearson.

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences, 21*(4), 327–336.

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences, 29*(2), 109–125.

Booth, J. N., Boyle, J. M. E., & Kelly, S. W. (2010). Do tasks make a difference? Accounting for heterogeneity of performance of children with reading difficulties on tasks of executive function: Findings from a meta-analysis. *British Journal of Developmental Psychology, 28*(1), 133–176.

Breit, M., Scherrer, V., Tucker-Drob, E. M., & Preckel, F. (2024). The stability of cognitive abilities: A meta-analytic review of longitudinal studies. *Psychological Bulletin, 150*(4), 399–439.

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

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-analytic studies*. Cambridge University Press.

Carroll, J. B. (2003). The higher stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–22). Pergamon Press.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Taylor Francis.

Cohen, A., Fiorello, C., & Farley, F. (2006). The cylindrical structure of the Wechsler Intelligence Scale for Children–IV: A retest of the Guttman model of intelligence. *Intelligence, 24*, 587–591.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.

Decker, S. L. (2021). Don't use a bifactor model unless you believe the true structure is bifactor. *Journal of Psychoeducational Assessment, 39*(1), 39–49.

Decker, S. L., Bridges, R. M., Luedke, J. C., & Eason, M. J. (2021). Dimensional evaluation of cognitive measures: Methodological confounds and theoretical concerns. *Journal of Psychoeducational Assessment, 39*(1), 3–27.

Demetriou, E. A., DeMayo, M. M., & Guastella, A. J. (2019). Executive function in autism spectrum disorder: History, theoretical models, empirical findings, and potential as an endophenotype. *Frontiers in Psychiatry, 10*, 753–753.

Diamond, A. (2005). Attention-deficit disorder (attention-deficit/hyperactivity disorder without hyperactivity): A neurobiologically and behaviorally distinct disorder from attention-deficit/hyperactivity disorder (with hyperactivity). *Developmental Psychopathology, 17*, 807–825.

Dombrowski, S. C., McGill, R. J., & Morgan, G. B. (2019). Monte Carlo modeling of contemporary intelligence test (IQ) factor structure: Implications for IQ assessment, interpretation, and theory. *Assessment, 28*(3), 977–993.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428–1446.

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

Euler, M. J., Vehar, J. V., & Guevara, J. E. (2023). Theories of Intelligence. In *Handbook of clinical child psychology: Integrating theory and research into practice* (pp. 289–323). Springer International Publishing.

Farmer, R. L., McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2020). Why questionable assessment practices remain popular in school psychology: Instructional materials as pedagogic vehicles. *Canadian Journal of School Psychology, 36*(2), 98–114.

Farmer, R. L., Zaheer, I., Duhon, G. J., & Ghazal, S. (2021). Reducing low-value practices a functional-contextual consideration to aid in de-implementation efforts. *Canadian Journal of School Psychology, 36*(2), 153–165.

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

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

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (in press). *Essentials of cross-battery assessment* (4th ed.). 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.). Wiley.

Folstein, M. F., Folstein, S. E., & Fanjiang, G. (2010). User's Manual. *Mini-Mental State Examination* (2nd ed.). PAR.

Hegarty, M. (2010). Components of spatial intelligence. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 52, pp. 265–297). Academic Press.

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock, Technical Manual. *Woodcock-Johnson Psycho-Educational Battery–Revised* (pp. 197–232). Riverside Assessments, LLC.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). Guilford Press.

Individuals with Disabilities Education Improvement Act of 2004, Pub. L. No. 108–446, 118 Stat. 2647 (2004) Enacted H.R. 1350. Final regulations implementing IDEA 2004 were published in the Federal Register, Monday, August 14, 2006, pp. 46540–46845.

Jensen, A. R. (1998). *The* g *factor: The science of mental ability*. Praeger.

Jewsbury, P. A., Jia, Y., & Gonzalez, E. J. (2024). Considerations for the use of plausible values in large-scale assessments. *Large-Scale Assessments in Education, 12*, Article 24.

Jurica, P. J., Leitten, C. L., & Mattis, S. (2001). DRS-2 Professional Manual. *Dementia Rating Scale-2*. PAR.

Kaufman, A. S. (1979). *Intelligent testing with the WISC-R*. Wiley.

Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence*. Allyn & Bacon.

Kaufman, A. S. (2001). WAIS-III IQs, Horn's theory, and generational changes from young adulthood to old age. *Intelligence, 29*(2), 131–167.

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

Kaufman, A. S., & Kaufman, N. L. (2014). *Kaufman Test of Educational Achievement* (3rd ed.) Pearson.

Kaufman, A. S., Kaufman, N. L., Drozdick, L. W., & Morrison, J. (2018). Manual Supplement. *Kaufman Assessment Battery for Children* (2nd ed.) Normative Update. Pearson.

Kent, S. C., & Wanzek, J. (2016). The relationship between component skills and writing quality and production across developmental levels. *Review of Educational Research, 86*(2), 570–601.

Kirby, J. R., Georgiou, G., Martinussen, R., & Parrila, R. (2010). Naming speed and reading: From prediction to instruction. *Reading Research Quarterly, 45*, 341–362.

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

Koponen, T., Georgiou, G., Salmi, P., Leskinen, M., & Aro, M. (2017). A meta-analysis of the relation between RAN and mathematics. *Journal of Educational Psychology, 109*(7), 977.

LaForte, E. M. (in press). Development and preliminary validation of a new WJ V test of executive functioning for school-age children. *Journal of Psychoeducational Assessment*.

LaForte, E. M., Dailey, D., & McGrew, K. S. (2025). Technical Manual. *Woodcock-Johnson V*. Riverside Assessments, LLC.

Li, X., & Sireci, S. (2013). A new method for analyzing content validity data using multidimensional scaling. *Educational and Psychological Measurement, 73*(3), 365–385.

Lohman, D. F., & Lakin, J. (2011). Reasoning and intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (2nd ed., pp. 419–441). Cambridge University Press.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7*(2), 107–127.

Mather, N., McGrew, K. S., LaForte, E. M., & Wendling, B. J. (2025a). *Woodcock-Johnson V Tests of Achievement*. Riverside Assessments, LLC.

Mather, N., McGrew, K. S., LaForte, E. M., & Wendling, B. J. (2025b). *Woodcock-Johnson V Virtual Test Library*. Riverside Assessments, LLC.

McCallum, R. S. (Ed.). (2017). *Handbook of nonverbal assessment* (2nd ed.). Springer.

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). Guilford Press.

McGrew, K. S. (2005). The Cattell-Horn-Carroll (CHC) theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 136–202). Guilford Press.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research [Editorial]. *Intelligence, 37*, 1–10.

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.

McGrew, K. S. (2023). Carroll's three-stratum (3S) cognitive ability theory at 30 years: Impact, 3S-CHC theory clarification, structural replication, and cognitive-achievement psychometric network analysis. *Journal of Intelligence, 11*, 32.

McGrew, K. S., Dailey, D., & Schrank, F. (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). Riverside Assessments, LLC.

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

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). Technical Manual. *Woodcock-Johnson IV*. Riverside Assessments, LLC.

McGrew, K. S., Mather, N., & LaForte, E. M. (2025). *Woodcock-Johnson V Tests of Cognitive Abilities*. Riverside Assessments, LLC.

McGrew, K. S., Mather, N., LaForte, E. M., & Wendling, B. J. (2025). *Woodcock-Johnson V*. Riverside Assessments, LLC.

McGrew, K. S., Schneider, W. J., Decker, S. L., & Bulut, O. (2023). A psychometric network analysis of CHC intelligence measures: Implications for research, theory, and interpretation of broad CHC scores "beyond *g*." *Journal of Intelligence, 11*, 19.

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

McGrew, K. S., & Woodcock, R. W. (2001). Technical Manual. *Woodcock-Johnson III*. Riverside Assessments, LLC.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–100).

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

Meyer, E. M., & Reynolds, M. R. (2022). Multidimensional scaling of cognitive ability and academic achievement scores. *Journal of Intelligence, 10*(4), Article 117.

Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement, 29*(2), 133–161.

Miyake, A., Emerson, M. J., & Friedman, N. P. (2000). Assessment of executive functions in clinical settings: Problems and recommendations. *Seminars in Speech and Language, 17*(2), 169–183.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science: A Journal of the American Psychological Society, 21*(1), 8–14.

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

NCS Pearson. (2020). *Wechsler Individual Achievement Test* (4th ed.).

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology, 63*, 427–452.

Pahor, A., Stavropoulos, T., Jaeggi, S. M., & Seitz, A. R. (2019). Validation of a matrix reasoning task for mobile devices. *Behavior Research Methods, 51*(5), 2256–2267.

Parkin, J. R. (2021). The simple views of reading and writing: Frameworks for interpretation of the Woodcock-Johnson IV. *Journal of Psychoeducational Assessment, 39*(7), 832–847.

Raven J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology, 41*(1), 1–48.

Ray, K., Dally, K., Rowlandson, L., Tam, K. L., & Lane, A. E. (2022). The relationship of handwriting ability and literacy in kindergarten: A systematic review. *Reading and Writing, 35*, 1119–1155.

Reynolds, C. R., & Kamphaus, R. W. (2015). *Reynolds Intellectual Assessment Scales* (2nd ed.) and *Reynolds Intellectual Screening Test* (2nd ed.). Psychological Assessment Resources.

Reynolds, M. R., & Keith, T. Z. (2017). Multi-group and hierarchical confirmatory factor analysis of the Wechsler Intelligence Scale for Children–Fifth Edition: What does it measure? *Intelligence, 62*, 31–47.

Salthouse, T. (2012). Consequences of age-related cognitive declines. *Annual Review of Psychology, 63*(1), 201–226.

Schneider, W. J. (2016). Strengths and weaknesses of the Woodcock-Johnson IV Tests of Cognitive Abilities: Best practice from a scientist–practitioner perspective. In D. P. Flanagan & V. C. Alfonso (Eds.), *WJ IV clinical use and interpretation* (pp. 191–210). Academic Press.

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). Guilford Press.

Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73–163). Guilford Press.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014). *Woodcock-Johnson IV Tests of Oral Language*. Riverside Assessments, LLC.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV*. Riverside Assessments, LLC.

Schrank, F. A., McGrew, K. S., & Mather, N. (2015). *The WJ IV composite and its use in the identification of specific learning disabilities* (Woodcock-Johnson IV Assessment Service Bulletin No. 3). Riverside Assessments, LLC.

Siegler, R. S., & Braithwaite, D. W. (2017). Numerical development. *Annual Review of Psychology, 68*, 187–213.

Smith-Spark, J. H., Henry, L. A., Messer, D. J., Edvardsdottir, E., & Zięcik, A. P. (2016). Executive functions in adults with developmental dyslexia. *Research in Developmental Disabilities, 53–54*, 323–341.

Stankov, L. (2000). Structural extensions of a hierarchical view on human cognitive abilities. *Learning and Individual Differences, 12*(1), 35–51.

Stankov, L. (2005). g factor: Issues of design and interpretation. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 279–293). Sage Publications.

Swanson, H. L., Trainin, G., Necoechea, D. M., & Hammill, D. D. (2003). Rapid naming, phonological awareness, and reading: A meta-analysis of the correlation evidence. *Review of Educational Research, 73*(4), 407–440.

SYSTAT (2009) (Version 13.1) [Computer software]. Grafiti, LLC.

Titz, C., & Karbach, J. (2014). Working memory and executive functions: Effects of training on academic achievement. *Psychological Research, 78*, 852–868.

Tucker-Drob, E. M., & Salthouse, T. A. (2009). Confirmatory factor analysis and multidimensional scaling for construct validation of cognitive abilities. *International Journal of Behavioral Development, 33*(3), 277–285.

U.S. Office of Management and Budget. (1997). *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*. https://www.govinfo.gov/content/pkg/FR-1997-10-30/pdf/97-28653.pdf

Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (2015). Verbal and visual-spatial working memory and mathematical ability in different domains throughout primary school. *Memory and Cognition, 43*(3), 367–378.

Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? *IERI monograph series, 2*(1), 9–36.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A., (2013). *Comprehensive Test of Phonological Processing* (2nd ed.). Pearson.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). Pearson. Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence* (4th ed.). Pearson.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence* (4th ed.). Pearson.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children* (5th ed.). Pearson.

Wolf, M., & Denckla, M. B. (2005). *Rapid Automatized Naming and Rapid Alternating Stimulus Test*. Pro-ed.

Woodcock, R. W. (1978). *Development and standardization of the Woodcock-Johnson Psycho-Educational Battery*. Riverside Assessments, LLC.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery–Revised*. Riverside Assessments, LLC.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Riverside Assessments, LLC.

Woodcock, R. W., McGrew, K. S., Schrank, F. A., & Mather, N. (2001, 2007). *Woodcock-Johnson III Normative Update*. Riverside Assessments, LLC.

Riverside Insights®

800.323.9540
https://www.riversideinsights.com/woodcock_johnson_v