



Why WJ IV™ and WJ V™ Scores Should Not Be Directly Compared

A Guide for Practitioners Transitioning Between Editions

Riverside Insights®

The Woodcock-Johnson® V (WJ V; McGrew, Mather, LaForte, & Wendling, 2025) represents a fundamental revision and reorganization of the Woodcock-Johnson IV (WJ IV; Schrank, McGrew, & Mather, 2014). While the names of many tests and clusters appear familiar to experienced WJ IV users, the WJ V is not simply an update of the WJ IV. It is a new battery with new tests, new item content, new cluster compositions, a new normative sample, and an entirely new mode of administration. Comparing scores from the WJ IV to scores from the WJ V as if they were equivalent is inappropriate and can lead to serious interpretive errors. This bulletin explains why, with specific examples drawn from the WJ V Technical Manual (LaForte, Dailey, & McGrew, 2025).

Copyright © 2026 by Riverside Assessments, LLC. All rights reserved. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, or by any information storage or retrieval system without the prior written permission of Riverside Assessments, LLC. Permission is hereby granted to individuals having permissible access to these materials to photocopy entire pages from this publication for personal, non-commercial use and not for resale or any commercial activity. Requests for information on other matters regarding duplication of this work should be sent via email to permissions@riversideinsights.com or addressed via mail to Riverside Insights, Attention: Permissions, One Pierce Place, Suite 900W, Itasca, Illinois 60143.

Riverside Insights, the Riverside Insights logo, Riverside Score, Woodcock-Johnson, and the Woodcock-Johnson logo are registered trademarks of Riverside Assessments, LLC. WJ IV, WJ V, and the WJ V logo are trademarks of Riverside Assessments, LLC.

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:

Riverside Insights. (2026). *Why WJ IV and WJ V Scores Should Not Be Directly Compared: A Guide for Practitioners Transitioning Between Editions*. Riverside Assessments, LLC.

Why WJ IV and WJ V Scores Should Not Be Directly Compared

The WJ V Is a New Battery, Not Just a New Edition

When a new edition of a familiar test is created, practitioners sometimes assume that the new edition is essentially the same instrument with updated norms. That assumption does not apply to the WJ V. The WJ V Technical Manual explicitly states that “The WJ V represents the culmination of approximately 50 years of theory, research, and applied test development and innovation; it is a significant revision and reorganization of the WJ IV” (LaForte et al., 2025, p. 70). The battery introduces new tests, restructured clusters, modernized CHC theory alignment, and a complete shift from paper-based to digital administration.

This matters because practitioners who began using the WJ V in 2025 may be tempted to compare new WJ V scores to WJ IV scores obtained in prior evaluations or to treat the two WJ V editions as interchangeable when tracking an individual’s progress over time. The WJ V Technical Manual is direct on this point:

“Users should exercise extreme caution when interpreting score differences between similarly named WJ IV and WJ V tests and clusters.” (LaForte et al., 2025, p. 70)

Seven Reasons Scores Will Differ Between Editions

The WJ V Technical Manual (LaForte et al., 2025, pp. 70–71) identifies seven general sources of score differences that may occur when comparing scores from the WJ IV to the WJ V.

1. **Differences in test construct coverage, item content, and scoring keys.** Many WJ V tests that share names with WJ IV tests contain new items, revised response formats, and updated scoring criteria. The construct being measured may be the same in name, but the specific tasks may have changed.
2. **Differences in scaling methods.** The WJ V uses the *W*-score metric with updated item calibration based on new normative data. So WJ V scores are not on the same scale as WJ IV scores, even when they look identical on the surface (e.g., both are expressed as standard scores with $M = 100$, $SD = 15$).
3. **Differences in standardized administration procedures and instructions.** Examiner administration scripts, starting points, discontinuation rules, and item presentation sequences have all been updated for the WJ V. Administration differences can directly affect an examinee’s performance.
4. **Differences in mode of administration.** The WJ IV was a paper-and-pencil battery. The WJ V is administered entirely via the digital platform (*Riverside Score*®). This is not a cosmetic change. The WJ V development team recognized early that “potential differences in test instructions, stimulus presentation, and examinee response mode between paper and digital versions may introduce construct-irrelevant differences that render scores from the two versions incomparable” (LaForte et al., 2025, p. 94). The decision to make the WJ V digital only was made precisely because the two formats could not be assumed equivalent.

5. **Differences in the composition of cluster scores.** This is perhaps the most important source of non-equivalence. As detailed later in this bulletin, virtually every cluster in the WJ V has a different composition than its WJ IV counterpart—even when the cluster shares the same name.
6. **Differences in the reference population.** The WJ V was standardized on a new, nationally representative sample collected from 2021–2023. The WJ IV normative sample was collected over a decade earlier. Even if the tests were identical, scores derived from different norm samples are not directly comparable.
7. **Norm obsolescence (the Flynn Effect).** Populations tend to score higher on older norms over time (Floyd et al., 2021; McGrew, 2015). This means that a person who takes the WJ IV today will likely obtain higher standard scores than they would on the newly normed WJ V, even if their underlying abilities are unchanged. The Flynn Effect is a well-documented psychometric phenomenon, not a reflection of any flaw in either battery.

In addition to these seven systematic sources of score variance, scores for the same individual across editions may also differ due to random measurement error or genuine change—growth or decline—in the individual’s measured abilities over time. Any or all of these sources of variance may be operating simultaneously when a practitioner compares WJ IV and WJ V scores.

Cluster Compositions Have Changed: What Looks the Same Is Not the Same

The most consequential—and most easily overlooked—source of non-equivalence between the WJ IV and WJ V is the change in cluster compositions. The WJ V Technical Manual notes that for clusters carried forward from the WJ IV, “there are many significant modifications to test names, content, and scoring, as well as to cluster names and compositions” (LaForte et al., 2025, p. 70). A cluster with the same name that appears to measure the same construct may contain entirely different tests or may use the same tests in a different context.

General Intellectual Ability (GIA)

The WJ IV GIA consisted of seven tests, differentially weighted to optimize the estimate of psychometric *g*. The WJ V GIA consists of eight equally weighted tests. Only three tests are shared between the two versions: Oral Vocabulary (Gc), Verbal Attention (Gwm), and Story Recall (Gl). Every other GIA test has changed, as shown in Table 1.

Table 1.
Differences Between GIA in WJ IV and WJ V

WJ IV	WJ V
Number Series (Gf): COG Standard battery	Matrices (Gf): COG Standard battery; replaces Number Series
Letter-Pattern Matching (Gs): COG Standard battery	Number-Pattern Matching (Gs): COG Standard battery; replaces Letter-Pattern Matching
Visualization (Gv): Two-part subtest (Spatial Relations + Block Rotation)	Spatial Relations (Gv): Spatial Relations and Block Rotation are now both full separate tests in the COG Standard battery
Phonological Processing (Ga): Three-part subtest (Word Access + Word Fluency + Substitution)	Removed from the GIA entirely
—Not present	Semantic Word Retrieval (Gr): Added; from the WJ IV OL battery
—Not present	Verbal Analogies (Gc/Gf): Added

The changes shown in Table 1 are not minor refinements. Replacing Number Series with Matrices and removing Phonological Processing from the GIA were identified in simulation studies as the primary drivers of changes in the GIA's relationship to achievement scores (LaForte et al., 2025,). The WJ IV GIA had a median correlation of .70 with achievement clusters; the WJ V GIA has a median correlation of .65—a difference that reflects deliberate changes in cluster composition, not lower validity. The WJ V authors determined that Number Series introduced a degree of achievement contamination into the GIA (Gq variance) and that the Phonological Processing test, as a domain-specific predictor of reading and writing, was inflating GIA-achievement correlations. These are improvements, but they mean that a WJ IV GIA score and a WJ V GIA score are measuring psychometric g through a meaningfully different combination of tests.

Comprehension-Knowledge (Gc)

The WJ IV Comprehension-Knowledge (Gc) cluster consisted of Oral Vocabulary and General Information. In the WJ V, General Information has been replaced with Verbal Analogies, a mixed measure of Gc and Gf. This change was made because research suggested that General Information was “more of a vocabulary test than a general knowledge test” (Schneider, 2016), biasing the WJ IV Gc cluster toward lexical knowledge. The WJ V Gc cluster is designed to better reflect both language knowledge and reasoning. A Gc score from the WJ IV and a Gc score from the WJ V reflect different combinations of abilities.

Phonological Awareness—A Cluster That Moved Entirely

Perhaps the clearest example of why direct cross-edition comparison is inappropriate involves what is arguably the most clinically significant cluster in reading-related assessments: Phonological Awareness.

In the WJ IV, the closest cluster was Phonetic Coding, found in the Oral Language (OL) battery. It consisted of Sound Blending and Segmentation.

In the WJ V, the Phonological Awareness cluster consists of the same two tests—Sound Blending and Segmentation—but it is now located in the Virtual Test Library (VTL), a separate supplemental battery. It is no longer part of the OL or COG battery. As the WJ V Technical Manual notes, the name change from Phonetic Coding to Phonological Awareness “reflects the authors' desire to modernize the Ga cluster names (based on contemporary clinical and research literature) rather than aligning with factor analysis terminology that is more than 30 years old” (LaForte et al., 2025, p. 58).

A WJ IV Phonetic Coding cluster score and a WJ V Phonological Awareness cluster score come from the same two tests but from different batteries, different scoring contexts, different norms, and a different platform (paper vs. digital). They should not be compared directly.

It is also critical to understand that what appeared as Phonological Processing in the WJ IV COG was a three-part composite test (Word Access, Word Fluency, and Substitution). Word Access has been dropped entirely from the WJ V. Word Fluency has been renamed Phonemic Word Retrieval and is now a standalone COG test. Substitution has been renamed Sound Substitution and moved to the VTL. The WJ IV Phonological Processing test no longer exists in the WJ V in any form. Any comparison of WJ IV Phonological Processing test scores to WJ V scores is comparing scores to a construct that has been fundamentally restructured.

Broad CHC Cognitive Clusters

With the exception of Auditory Working Memory Capacity (formerly Short-Term Working Memory in the WJ IV), the composition of every two-test broad CHC cluster changed, either partially or completely, from the WJ IV to the WJ V. Table 2 summarizes changes to selected clusters.

Table 2.
*Differences Between WJ IV
and WJ V Clusters*

WJ IV	WJ V
Comprehension-Knowledge (Gc): Oral Vocabulary + General Information	Comprehension-Knowledge (Gc): Oral Vocabulary + Verbal Analogies
Short-Term Working Memory (Gwm): Verbal Attention + Numbers Reversed	Auditory Working Memory Capacity (Gwm): Verbal Attention + Numbers Reversed; same composition but name changed
Cognitive Processing Speed (Gs): Letter-Pattern Matching + Pair Cancellation	Cognitive Processing Speed (Gs): Number-Pattern Matching + Letter-Pattern Matching; Pair Cancellation removed from the WJ V entirely
Long-Term Retrieval (Glr): Story Recall + Visual-Auditory Learning	Long-Term Storage (Glr): Story Recall + Story Comprehension; Visual- Auditory Learning moved to COG Extended battery; ability abbreviation changed from Glr to Gl
Auditory Processing (Ga): Phonological Processing + Nonword Repetition	Phonological Awareness (Ga): Sound Blending + Segmentation; moved to VTL battery
Phonetic Coding: Sound Blending + Segmentation; in OL battery	Phonological Awareness (Ga): Sound Blending + Segmentation; the same tests; new cluster name; moved to VTL battery; new norms

A Note on Tests That Appear in Both Editions

Even tests that carry the same name and appear to be the same across editions may have been modified. The transition from the WJ IV to the WJ V included the following changes.

- The WJ IV Visualization test consisted of two subtests (Spatial Relations + Block Rotation). In the WJ V, these are two separate, full-length tests. Scores from the WJ IV Visualization subtest cannot be meaningfully compared to the WJ V Spatial Relations or Block Rotation scores.
- The WJ V standalone COG Phonemic Word Retrieval test appeared as the Word Fluency subtest of the WJ IV Phonological Processing test. The WJ V test is not the same as the WJ IV subtest.
- Word Access, another subtest of the WJ IV Phonological Processing test, was dropped entirely from the WJ V.
- WJ IV COG Extended tests Picture Recognition and Pair Cancellation were not included anywhere in the WJ V.
- The WJ IV OL Retrieval Fluency test was renamed Semantic Word Retrieval and moved to the WJ V COG battery, where it now contributes to the GIA cluster.

Just as we do not expect test and cluster score interpretations to remain static across editions, the WJ V Technical Manual notes that “we also do not assume that all validity evidence gathered for one edition of a test is automatically appropriate for supporting score inferences from another edition” (LaForte et al., 2025, p. 71). New evidence was gathered for the WJ V, and it is that evidence—not the WJ IV evidence base—that supports WJ V score interpretations.

What Should Practitioners Do?

The WJ V is a technically sophisticated, contemporary battery with strong validity evidence. The goal of this bulletin is to highlight the changes and upgrades in the WJ V to ensure that practitioners understand what the WJ V is and what it is not. This information will help practitioners interpret WJ V scores accurately and communicate those interpretations to the families, teams, and organizations they serve.

When transitioning from WJ IV to WJ V, practitioners should do the following.

- Treat WJ V scores as new baseline data, not as a continuation of a WJ IV score history. A student who received a WJ IV GIA of 85 in 2022 and a WJ V GIA of 88 in 2025 has not necessarily gained 3 IQ points because the two scores are not on the same scale and the clusters are not composed of the same tests.
- Be especially cautious with cluster comparisons when the cluster name is the same but the composition has changed. The GIA, Gc, Gs, Glr/Gl, and Ga clusters all have meaningfully different compositions in the WJ IV and WJ V.
- Recognize that the WJ V Phonological Awareness cluster in the VTL is the conceptual successor to WJ IV Phonetic Coding cluster in the OL battery, but it does not provide the same score. Do not compare the scores for these two clusters in psychoeducational reports.
- Consult Chapter 2 of the WJ V Technical Manual for a full description of changes from the WJ IV to the WJ V, including changes to specific test names, cluster compositions, and interpretive procedures.
- When writing reports for individuals who have both WJ IV and WJ V histories, clearly note that the two editions use different tests, norms, and administration formats and that direct comparison of scores across editions is not appropriate.

Summary

The WJ V is a significant revision of the WJ IV, not a renorming of it. Although many test and cluster names are preserved across editions, the changes in item content, test composition, cluster membership, administration mode, and normative samples are substantial enough that WJ IV and WJ V scores cannot be treated as equivalent. The WJ V Technical Manual (LaForte et al., 2025, p. 70) itself cautions practitioners to “exercise extreme caution when interpreting score differences between similarly named WJ IV and WJ V tests and clusters.”

Key reasons that WJ IV and WJ V scores should not be directly compared include:

- Items, scoring keys, and administration procedures have changed.
- The mode of administration shifted from paper to digital.
- Most cluster compositions changed, and in some cases, they changed substantially (e.g., the GIA).
- Some tests were removed entirely (e.g., Pair Cancellation, Picture Recognition, Word Access).
- Some tests were restructured (e.g., Visualization split into Spatial Relations and Block Rotation).
- Some clusters changed batteries (e.g., Phonological Awareness/Phonetic Coding moved from WJ IV OL to WJ V VTL).
- The WJ V normative sample is new, and norm obsolescence (the Flynn Effect) adds additional score differences.

Each edition of the Woodcock-Johnson should be interpreted within its own framework, using its own evidence base, its own cluster definitions, and its own normative data. WJ V results stand on their own as current, valid, and technically sound evidence about an examinee’s abilities, and practitioners who understand the differences between editions will be best positioned to communicate that evidence with clarity and confidence.

References

- Floyd, R. G., Reynolds, M. R., Farmer, R. L., & Kranzler, J. H. (2021). Are the general factors from different child and adolescent intelligence tests the same? Results from a five-sample, six-test analysis. *School Psychology Review*, 42, 383–401.
- LaForte, E. M., Dailey, D., & McGrew, K. S. (2025). Technical Manual. *Woodcock-Johnson V*. Riverside Assessments, LLC.
- McGrew, K. (2010). The Flynn Effect and its critics: Rusty linchpins and lookin’ for g and Gf in some of the wrong places. *Journal of Psychoeducational Assessment*, 28(5), 448–468.
- McGrew, K. S., Mather, N., LaForte, E. M., & Wendling, B. J. (2025). *Woodcock-Johnson V*. Riverside Assessments, LLC.
- Schneider, W. J. (2016). Buros review of the *Woodcock-Johnson IV*. *Mental Measurements Yearbook*.
- Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV*. Riverside Assessments, LLC.



Riverside
Insights®

800.323.9540

https://www.riversideinsights.com/woodcock_johnson_v