



# Test Review

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

**Reviewed by:** Victor Villarreal, The University of Texas at San Antonio, TX, USA

DOI: 10.1177/0734282915569447

## Test Description

### General Description

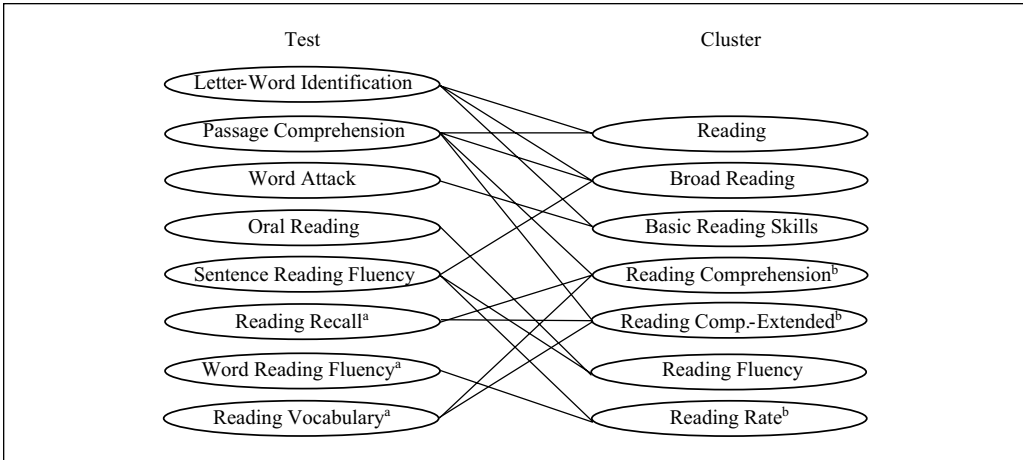
The *Woodcock-Johnson IV Tests of Achievement* (WJ IV ACH; Schrank, Mather, & McGrew, 2014a) is an individually administered measure containing tests of reading, mathematics, written language, and academic knowledge. Areas of reading, mathematics, and written language each include tests of basic skills, fluency, and application. Academic knowledge includes tests of science, social studies, and humanities. The test authors note that the WJ IV ACH can be used to assist with determining an individual's academic strengths and weaknesses, diagnosing specific abilities and disabilities, and educational planning (Schrank et al., 2014a). When used in conjunction with the WJ IV Tests of Cognitive Abilities (WJ IV COG; Schrank, McGrew, & Mather, 2014) and the WJ IV Tests of Oral Language (WJ IV OL; Schrank, Mather, & McGrew, 2014b), it can also be used to evaluate variations between an individual's achievement and cognitive and linguistic abilities. Interpretation of WJ IV test batteries is based on the Cattell–Horn–Carroll (CHC) theory of cognitive abilities (see Schneider & McGrew, 2012).

The WJ IV ACH was published by Riverside in 2014; the previous version, the WJ III Tests of Achievement (Woodcock, McGrew, & Mather, 2001), was published in 2001. There are seven new tests in the WJ IV ACH; however, it no longer includes tests of oral language abilities as these are now published separately. Examiners should have knowledge of exact WJ IV ACH administration and scoring procedures; graduate-level training in educational and psychological assessment is recommended (Schrank et al., 2014a). The test may be administered to individuals from age 2 to over 90 years. Most of the tests in the WJ IV ACH require 5 to 10 min to administer; however, some require 15 to 20 min.

### Specific Description

The WJ IV ACH contains two test batteries. The Standard Battery contains 11 tests; there are three alternate and parallel forms (A, B, and C) of tests in this battery. The single version of the Extended Battery contains nine tests. Notably, Tests 1 through 6 are considered the core set of tests and are required for calculating intra-achievement variations (Schrank et al., 2014a). Administration of the WJ IV ACH yields up to 22 cluster scores for interpretation; tests in the Standard Battery form 15 cluster scores, and administration of the Extended Battery provides an additional 7 cluster scores.

Seven reading clusters are available. The *Reading* cluster is a measure of reading decoding and reading comprehension. The *Broad Reading* cluster is a measure of reading decoding, reading



**Figure 1.** Woodcock-Johnson IV Tests of Achievement reading tests and reading clusters.

<sup>a</sup>Tests come from the Extended Battery.

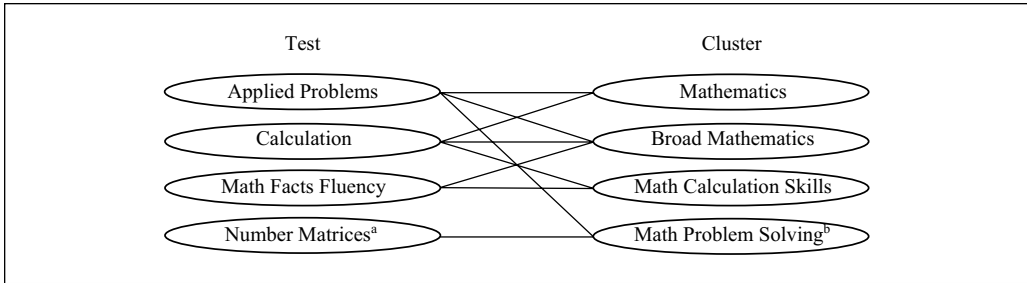
<sup>b</sup>Clusters require administration of tests from the Extended Battery.

speed, and reading comprehension. These first two clusters can be calculated by administering the core set of tests. The *Basic Reading Skills* cluster measures sight vocabulary, phonics, and structural analysis. The *Reading Comprehension* and *Reading Comprehension-Extended* clusters measure comprehension, reasoning, and vocabulary. *Reading Fluency* is a cluster that measures prosody, automaticity, and accuracy. *Reading Rate* measures automaticity with reading at the single word and sentence levels. Notably, the WJ IV ACH includes clusters that were not present in the WJ III ACH, including Reading Comprehension-Extended, Reading Fluency, and Reading Rate. See Figure 1 for a list of the names of the WJ IV ACH reading clusters and the tests that contribute to them.

Four math clusters are available. The *Mathematics* cluster provides a measure of problem solving and computational skill. The *Broad Mathematics* cluster is a measure of problem solving, number facility, automaticity, and reasoning. These first two clusters can be calculated by administering the core set of tests. *Math Calculation Skills* is a cluster that measures computational skills and automaticity with basic math facts. *Math Problem Solving* measures mathematical knowledge and reasoning. There are no significant differences in the math clusters of the WJ IV ACH and the WJ III ACH. See Figure 2 for a list of the names of the WJ IV ACH math clusters and the tests that contribute to them.

Four written language clusters are available. The *Written Language* cluster measures spelling and quality of expression. The *Broad Written Language* cluster measures spelling, writing fluency, and quality of expression. These first two clusters can be calculated by administering the core set of tests. *Basic Writing Skills* measures spelling and identifying and correcting errors in spelling, punctuation, capitalization, and word usage. *Written Expression* is a cluster that measures meaningful written expression and sentence writing fluency. There are no significant differences in the written language clusters of the WJ IV ACH and the WJ III ACH. See Figure 3 for a list of the names of the WJ IV ACH written language clusters and the tests that contribute to them.

Seven cross-domain clusters are available. Two of these clusters—*Brief Achievement* and *Broad Achievement*—are general academic proficiency clusters that measure performance in reading, writing, and math. The *Academic Skills*, *Academic Fluency*, and *Academic Applications* clusters contain tests of reading, math, and written language. *Academic Skills* is a measure of



**Figure 2.** Woodcock-Johnson IV Tests of Achievement mathematics tests and mathematics clusters.

<sup>a</sup>Tests come from the Extended Battery.

<sup>b</sup>Clusters require administration of tests from the Extended Battery.

basic achievement skill, and it can be calculated by administering the core set of tests; *Academic Fluency* is a measure of overall academic fluency; and *Academic Applications* is a measure of an individual's ability to apply academic skills to academic problems, and it can be calculated by administering the core set of tests. The *Academic Knowledge* cluster provides a broad sample of knowledge in science, social studies, and humanities. The *Phoneme–Grapheme Knowledge* cluster provides information about basic understanding of sound/symbol relationships.

### Scoring System

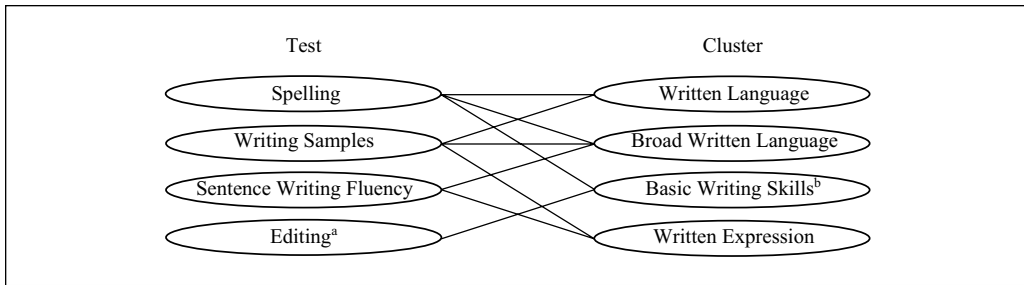
The paper Test Record booklet has built-in “Scoring Tables” that allow examiners to quickly estimate age- and grade-equivalent scores; however, these represent general estimates, so scores for interpretation should be attained from the online scoring program (<https://www.wjscore.com/>). Access to this program is provided with purchase of paper Test Records. In addition to precise age- and grade-equivalent scores, the scoring program can provide percentile ranks, cognitive-academic language proficiency (CALP) scores, relative proficiency index (RPI) scores, *W* scores, and standard scores. Examiners can also select normal curve equivalent (NCE) scores, stanine scores, *T*-scores, *z* scores, and proficiency range (e.g., average). Scores are provided for tests and clusters.

Calculations of actual and predicted discrepancies and variations can also be attained from the online scoring system. Using the WJ IV ACH, an examiner can determine intra-achievement and academic skills/academic fluency/academic applications variations across areas of reading, math, and written language. If used in conjunction with the other WJ IV test batteries (i.e., WJ IV COG and WJ IV OL), comparison procedures can be used to determine whether an examinee is achieving commensurate with his or her current levels of cognitive and oral language abilities.

Score reports can be output in PDF, web page, or Word formats. The online scoring program indicates that it allows users to delete test records. However, the delete functionality only hides the test record; the record remains in the database and can be restored if needed. Also, test records must be “committed” to utilize the scoring system; after committing test record information (i.e., raw scores and observations ratings), users can access reporting and score interpretation but only have 30 days to make changes to test data.

### Test Materials

The WJ IV ACH contains two easel Test Books, an Examiner's Manual, the Technical Manual on CD, Test Record and Examinee Response booklets, an audio recording on CD, and scoring guides. Examiners can also access a Report and Score Interpretation Guide through the scoring



**Figure 3.** Woodcock-Johnson IV Tests of Achievement written language tests and written language clusters.

<sup>a</sup>Tests come from the Extended Battery.

<sup>b</sup>Clusters require administration of tests from the Extended Battery.

website; this brief guide describes different report elements that can be selected for inclusion in the score report.

In the easel Test Books, administration and scoring directions face the examiner and the stimulus pictures and words face the examinee. General information and instructions specific to each test—including suggested starting points and basal/ceiling rules—are included in the Test Books. The Test Books are user friendly; verbal test instructions are highlighted in a differently colored font to assist with standardized administration, and instructions are written in a clear language. The easel format limits the ability of the examinee to see examiner information.

The Test Record booklet is used to record identifying information, general observations of behavior (e.g., attention and self-confidence), examinee responses, and raw scores. It also provides basal/ceiling rules and includes icons that indicate required materials (e.g., stopwatch) for each test. This booklet also includes Qualitative Observation Checklists for most of the tests in the standard Test Book. The checklists for each test are different, and they provide helpful information not available from the general test session observations checklist.

The Examiner's Manual includes descriptions, specific administration information, and scoring instructions for each test. It also includes reproducible test-by-test checklists that may be used as a self-study or observation tool; this is especially helpful for those learning to administer the WJ IV ACH, and they can also serve as a structured observation and evaluation tool for trainers. Finally, the manual includes the scoring guide for the Writing Samples Test.

## Technical Adequacy

### Test Construction

Development of the WJ IV ACH incorporated multiple stages including a review and update of the WJ III, creation of new tests and items, consultation with outside experts, and pilot testing and evaluation of items. Expert consultants included experienced teachers, university faculty, and psychologists; consultants assisted with new test and new item development.

Reviews and studies of the WJ III Tests of Achievement indicated that many subtests had inadequate floors and ceilings (Bradley-Johnson, Morgan, & Nutkins, 2004; Krasa, 2007). Notably, the authors of the WJ IV ACH note that one of the primary objectives of new item development was to extend the range of items at the very low and very high difficulty levels. In addition, for timed tests, items were added to reduce the number of examinees who would finish the test before the time limit. Adding new items also served to increase the item pool to allow for the formation of the three parallel forms of the Standard Battery.

## Item Analysis

Items utilized in the WJ IV ACH were evaluated using the item response theory measurement model. Specifically, calibration, item pool equating, and scaling were accomplished through the use of Rasch models (Rasch, 1980; Wright & Masters, 1982). Tests containing items that are scored dichotomously were calibrated using the dichotomous Rasch model; those containing multiple-point items were calibrated using the partial credit form of the Rasch model.

Expert reviewers examined item content for potential bias for multiple groups of people (i.e., women, individuals with certain disabilities, and individuals from cultural or linguistic minority backgrounds). Differential item functioning was also evaluated to provide an empirical review of item bias; in this case, items were evaluated by sex, race, and ethnicity. All items flagged in the differential item functioning analyses were reviewed by the test authors to identify sources of bias and, in most cases, were removed from final item pools.

## Standardization Sample

A stratified sample was used based on projections from a U.S. Census Bureau report from 2010. Depending on examinee age, samples were stratified based on census region, sex, country of birth, race, ethnicity, community type, parent education, type of school, type of college, educational attainment, employment status, and occupational level. Data were collected from 7,416 individuals from geographically diverse areas and were divided into four major sample levels. The preschool sample (ages 2 through 5 years) contained 664 children; the kindergarten through 12th-grade sample contained 3,891 examinees; the college/university sample contained 775 graduate and undergraduate students; and the adult sample contained 2,086 examinees.

Comparisons between the WJ IV norming sample and the U.S. Census projections were conducted at the major levels. The norming sample distribution matched the census data closely; however, individuals with higher education levels were overrepresented in the adult sample. Examinee weighting was applied during the test norm construction to account for such discrepancies; if an examinee belonged to a category that was overrepresented in the norming study sample, the examinee's partial weight for that variable was less than 1.0, and vice versa.

## Reliability

**Internal consistency.** Internal consistency reliabilities for all untimed tests with dichotomously scored items were calculated using the split-half procedure based on odd and even items. Reliabilities for these tests were primarily in the acceptable to excellent range (.84-.94). For tests containing multiple-point items, reliabilities were calculated from mean square error values; reliabilities on these tests were in the excellent range (.90-.96). Reliability estimates for the WJ IV ACH appear improved compared with those reported for the WJ III ACH (see Bradley-Johnson et al., 2004). In addition, reliabilities of WJ IV ACH cluster scores (.92-.97) are higher than they are for individual tests and meet minimum expectations for scores used to make important decisions (Ysseldyke & Nelson, 2012), so cluster scores are recommended for interpretation.

**Test-retest.** Reliabilities for speeded tests were based on a test-retest model with a 1 day time frame. In most cases, test-retest correlations were in the acceptable to excellent range (.83-.95), indicating adequate test-retest stability. Reliabilities for speeded tests also appear to be improvements over those from the WJ III ACH, and the reliabilities of cluster scores that include speeded tests are in the acceptable range for making important decisions.

*Alternative forms equivalence.* As previously noted, tests in the Standard Battery are available in three parallel forms. Items were selected for each form so that the item difficulty gradient was approximately equal for each and so that each contained equal representation of the intended test content. Content-area curriculum experts provided consultation on the comparability of the three forms, and equivalence was also evaluated by comparing test characteristic curves. Empirical evidence supports the equivalence of the alternate forms.

### Validity

*Content validity.* Content was designed to cover core curricular areas and achievement specified in federal legislation. In addition to content review by the test authors and content-area experts, multidimensional scaling (MDS) was used as a supplemental empirical tool. MDS provides information about content and processes underlying performance on diverse tasks; the Technical Manual (McGrew, LaForte, & Schrank, 2014) provides detailed information on the results of MDS analyses of the WJ IV tests, and results suggest adequate content validity.

*Construct validity.* Reported intercorrelations indicate that correlations are higher among related WJ IV ACH tests than among unrelated WJ IV ACH tests. Correlations are especially high among related WJ IV ACH clusters. This is expected, as many of the clusters utilize the same tests. For example, Test 1 (Letter-Word Identification) is utilized in deriving scores for the *Reading*, *Broad Reading*, and *Basic Reading Skills* clusters. Confirmatory multivariate statistical methods indicated that reading and writing tests demonstrated moderate to high factor loadings on the CHC Reading and Writing domain, supporting the validity of the reading and writing clusters. Moderate to strong math test loadings on the CHC Quantitative Knowledge domain also provided validity evidence. Notably, factor analyses were conducted to see how individual tests loaded on CHC broad factors (see Schneider & McGrew, 2012); factor analyses were not conducted to determine how individual tests loaded on WJ IV ACH clusters.

*Concurrent validity.* Five studies examined the relationship between WJ IV ACH scores and scores from the Kaufman Test of Educational Achievement–Second Edition (KTEA-II; Kaufman & Kaufman, 2004), the Wechsler Individual Achievement Test–Third Edition (WIAT-III; Wechsler, 2009), and the Oral and Written Language Scales–Written Expression (OWLS-WE; Carrow-Woolfolk, 1996). WJ IV ACH clusters generally showed their highest correlations with the measures of the same KTEA-II and WIAT-III domain composites. The WJ IV ACH written language clusters demonstrated moderate to strong correlations with the OWLS-WE total score. Overall, these correlations provide evidence of adequate concurrent validity. See the WJ IV ACH Technical Manual for more detailed results of these analyses.

*Clinical validity.* The Technical Manual also provides results of a clinical validity study that examined the relationship between test scores and group membership status. Of particular relevance to the WJ IV ACH is the examination of test scores for examinees identified as having learning disabilities (LDs) in reading, math, or writing. The LD-reading group was the only LD group with mean reading test scores consistently below 80; in those instances where specific reading test scores were common across the three LD groups, the LD-reading group's mean score was lower than the mean group scores for both the LD-writing and the LD-math groups. However, there were no clear-cut differences in scores in math and written language tests between the three LD groups. These results not only provide some additional evidence of validity for the reading tests but also demonstrate that interpretation of test information should be done in conjunction with other relevant information.

## Commentary and Recommendations

The WJ IV ACH assesses core curricular areas and achievement specified in federal legislation. The types of items and range of difficulty of tests seem appropriate for the stated population, and it appears especially useful for determining the academic achievement of students at the primary and secondary levels. The WJ IV ACH has been developed with a large, nationally representative sample. Information regarding reliability and validity is detailed and meets minimum requirements for tests used to make important decisions (e.g., diagnosing disabilities).

A particular strength of the WJ IV ACH is that it has been co-normed with the WJ IV COG and the WJ IV OL. This is particularly useful for professionals conducting comprehensive evaluations that require assessment of multiple areas of functioning. Another strength of the WJ IV ACH is that test materials are well structured and the repeated presentation of administration procedures in the Test Books and Test Record booklet is helpful. The addition of Qualitative Observation Checklists is welcome and may encourage examiners to be more thoughtful about collecting observation data throughout testing. Finally, for evaluations conducted with school-age children, the WJ IV ACH tests and clusters are now aligned with all of the reading, writing, and math categories listed in the specific learning disability definition of the Individuals With Disabilities Education Act (2004).

The WJ IV ACH also has some weaknesses. Although the authors state that it may be useful for instructional planning, the WJ IV ACH provides only a broad sampling of achievement areas, and the sample of skills is too limited for comprehensive instructional planning. Moreover, although the WJ IV ACH yields up to 22 cluster scores, it is important to note that most of the 20 WJ IV ACH tests are utilized in calculating multiple clusters; this results in especially high correlations between clusters in related areas (e.g., reading) and suggests some redundancy between clusters. In addition, although the WJ IV ACH has been normed on children as young as 2 years of age, many of the tests have inadequate floors for children in early childhood; examiners should consider giving alternate tests when working with very young children.

The WJ IV ACH also has changes that examiners have to consider. First, those who have used the WJ III ACH may be surprised to find that the tests of oral language abilities have been removed from the WJ IV ACH. This change may be particularly relevant to those working in school settings where administering tests of oral language abilities are considered a typical part of completing a comprehensive evaluation. Second, the complete shift to an online scoring and data management system, while having benefits, may also cause concerns about privacy and confidentiality.

Despite these relatively minor limitations, the WJ IV ACH is a strong test and meets its stated purpose. If used appropriately, and as a complement to other forms of psychological and educational data, it can certainly assist with diagnosis of specific disabilities and can serve as a general evaluation tool to guide more narrow evaluations that can better inform intervention and educational planning.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### References

Bradley-Johnson, S., Morgan, S. K., & Nutkins, C. (2004). Book review: The Woodcock-Johnson Tests of Achievement—Third edition. *Journal of Psychoeducational Assessment, 22*, 261-274.

- Carrow-Woolfolk, E. (1996). *Oral and Written Language Scales: Written expression*. Torrance, CA: Western Psychological Services.
- Individuals With Disabilities Education Act, 20 U.S.C. § 1400 (2004).
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Test of Educational Achievement* (2nd ed.). San Antonio, TX: Pearson.
- Krasa, N. (2007). Is the Woodcock-Johnson III a test for all seasons? Ceiling and item gradient considerations in its use with older students. *Journal of Psychoeducational Assessment*, 25, 3-16.
- McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical manual. Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: University of Chicago Press.
- Schneider, J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99-144). New York, NY: Guilford Press.
- Schrank, F. A., Mather, N., & McGrew, K. S. (2014a). *Woodcock-Johnson IV Tests of Achievement*. Rolling Meadows, IL: Riverside.
- Schrank, F. A., Mather, N., & McGrew, K. S. (2014b). *Woodcock-Johnson IV Tests of Oral Language*. Rolling Meadows, IL: Riverside.
- Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside.
- Wechsler, D. (2009). *Wechsler Individual Achievement Test* (3rd ed.). San Antonio, TX: Pearson.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Rolling Meadows, IL: Riverside.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago, IL: Mesa Press.
- Ysseldyke, J., & Nelson, P. (2012). Assessment in special and inclusive education. In J. Banks (Ed.), *Encyclopedia of diversity in education* (pp. 165-168). Thousand Oaks, CA: Sage.