

Assessment Service Bulletin Number I

Batería IV Woodcock-Muñoz Technical Abstract

Erica M. LaForte, PhD Barbara J. Wendling, MA Nancy Mather, PhD Fredrick A. Schrank, PhD, ABPP Kevin S. McGrew, PhD

This Assessment Service Bulletin provides a summary of the procedures followed in developing and validating the Batería IV Woodcock-Muñoz (Batería IV; Woodcock, Alvarado, Schrank, McGrew, Mather, & Muñoz-Sandoval, 2019a) as a comprehensive measure of individuals' cognitive abilities and academic achievement in Spanish. Throughout the development and design of the Batería IV, the test standards outlined in the Standards for Educational and Psychological Testing (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014) were followed carefully. Information in this bulletin is abstracted from the Woodcock-Johnson® IV Technical Manual (McGrew, LaForte, & Schrank, 2014) and the Batería IV examiner's manuals and is intended as an overview to highlight important aspects of the test design, reliability, and validity of the Batería IV. Readers who are interested in more detailed information should consult the Woodcock-Johnson IV Technical Manual and the Batería IV examiner's manuals.



Copyright © 2019 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 Insights unless such copying is expressly permitted by federal copyright law. Permission is hereby granted to individuals to photocopy entire pages from this publication for personal use and not for resale. Requests for information on other matters regarding duplication of this work should be addressed to Riverside Insights, Attention: Permissions, One Pierce Place, Suite 900W, Itasca, Illinois 60143.

Published in Itasca, Illinois

Manufactured in the United States

Batería III Woodcock-Muñoz, WJ III, WJ-R, and Woodcock-Johnson are registered trademarks of Riverside Assessments, LLC.

The Batería IV logo and WJ IV are trademarks of Riverside Assessments, LLC.

Reference Citation

To cite this document, use:

LaForte, E. M., Wendling, B. J., Mather, N., Schrank, F. A., & McGrew, K. S. (2019). Batería IV Woodcock-Muñoz *Technical Abstract* (Batería IV Assessment Service Bulletin No. 1). Itasca, IL: Riverside Insights.

For technical information, please visit www.wj-iv.com or call Riverside Insights Customer Service at 800.323.9540.

Batería IV Woodcock-Muñoz Technical Abstract

The Batería IV Woodcock-Muñoz (Batería IV; Woodcock, Alvarado, Schrank, McGrew, Mather, & Muñoz-Sandoval, 2019a) is a comprehensive, Spanish-language psychoeducational assessment system that includes two test batteries: the Batería IV Woodcock-Muñoz: Pruebas de habilidades cognitivas (Batería IV COG; Woodcock, Alvarado, Schrank, McGrew, Mather, & Muñoz-Sandoval, 2019b) and the Batería IV Woodcock-Muñoz: Pruebas de aprovechamiento (Batería IV APROV; Woodcock, Alvarado, Schrank, Mather, McGrew, & Muñoz-Sandoval, 2019). Tests included in the Batería IV are either adaptations or translations of tests from the Woodcock-Johnson[®] IV (WJ IV[™]; Schrank, McGrew, & Mather, 2014a). The Batería IV batteries can be used in conjunction with the Woodcock-Johnson IV Tests of Oral Language (WJ IV OL; Schrank, Mather, & McGrew, 2014b) to form a broad-based assessment of cognitive abilities, achievement, and comparative oral language abilities.

Some of the Batería IV tests can be used with Spanish-speaking individuals as young as 24 months, but the majority of the Batería IV tests are best suited for use with individuals from 5 to 95 years of age. Spanish-language calibration data, based on a sample of 601 native Spanish speakers, are equated to the large, nationally representative WJ IV norming sample of 7,416 individuals ranging from 2 to 90+ years of age. The Spanish data were used to calibrate the new test items and to equate the items to the scales underlying the WJ IV tests. The equating procedure produces a psychometrically sound interpretive model that allows an examiner to describe an individual's performance on the Spanish tests in terms of comparable ability in English.

Multiple goals guided the WJ IV revision blueprint that underlies the Batería IV. First, this comprehensive assessment system was designed to be on the cutting edge of practice. It facilitates exploration of individual strengths and weaknesses across cognitive, linguistic, and academic abilities; complements response to intervention (RTI) models; and reframes variations and ability/achievement comparisons. Second, the blueprint pushed the tests beyond Cattell-Horn-Carroll (CHC) theory as it was conceived in the Woodcock-Johnson III (WJ III®; Woodcock, McGrew, & Mather, 2001, 2007) and the Batería III Woodcock-Muñoz® (Batería III; Muñoz-Sandoval, Woodcock, McGrew, & Mather, 2005a, 2007a). Whereas the third editions of these tests focused primarily on broad CHC abilities, the fourth editions focus on the most important broad and narrow CHC abilities for describing cognitive performance and understanding the nature of learning problems (McGrew, 2012; McGrew & Wendling, 2010; Schneider & McGrew, 2012, 2018). Some of the tests and clusters emphasize narrow CHC abilities, and others reflect the importance of cognitive complexity through the influence of two or more narrow abilities on task requirements. Finally, additional revision goals addressed ease and flexibility of use. New features allow novice examiners to use the tests with confidence while providing experienced examiners with a rich array of interpretive options to customize and enhance their evaluations. The structure of the WJ IV and Batería IV systems also facilitates examiner use by creating comprehensive cognitive, achievement, and oral language batteries that can be used in conjunction with one another or as standalone batteries.

Theoretical Foundation of the Batería IV

The WJ IV and Batería IV represent the cutting edge of assessment based on the CHC theory of cognitive abilities, sometimes referred to as CHC theory version 2 (McGrew, LaForte, & Schrank, 2014; Schneider & McGrew, 2012, 2018).¹ As noted above, these tests emphasize the principal broad and narrow CHC abilities underlying cognitive performance and learning problems (McGrew, 2012; McGrew & Wendling, 2010; Schneider & McGrew, 2012, 2018).

The Batería IV COG has seven CHC factors. Two of the CHC factors, fluid reasoning (*Gf*) and comprehension-knowledge (*Gc*), can be traced to Cattell (1941, 1943, 1950) and his work on *Gf-Gc*, or fluid and crystallized intelligence. Later, Horn (1965) identified short-term memory (*Gsm*), long-term retrieval (*Glr*), processing speed (*Gs*), and visual-spatial thinking (*Gv*) as distinct abilities. Auditory processing (*Ga*) was identified by Horn and Stankov (1982). The CHC abilities have been refined and integrated by Woodcock (McArdle & Woodcock, 1998; Woodcock, 1988, 1990, 1993, 1994, 1998) and McGrew (1997, 2005, 2009) and recently revised by Schneider and McGrew (2012, 2018).

The Batería IV APROV contains tests that tap two other identified cognitive abilities: quantitative knowledge (Gq; identified by Horn, 1988, 1989) and reading-writing ability (Grw; identified by Carroll and Maxwell [1979] and Woodcock [1998]). The Batería IV APROV also includes additional measures of comprehension-knowledge (Gc), long-term storage and retrieval (Glr), and auditory processing (Ga). The Batería IV APROV tests were chosen to measure the major aspects of academic achievement, including reading, mathematics, and written language.

The interpretive model for the Batería IV reflects the most contemporary specification of CHC theory at the time of publication. Analysis of the *Woodcock-Johnson Psychoeducational Battery–Revised* (WJ-R[®]; Woodcock & Johnson, 1989), WJ III, and WJ IV norming samples provided three large, multi-ability data sets that were used to either confirm or revise initial construct specifications for the WJ IV and Batería IV. Support for changes to the interpretive constructs was gleaned from other sources of neuroscience research. Following are the most significant changes to the Batería IV blueprint resulting from theoretical revisions since the publication of the Batería III in 2005.

- The Batería III broad cognitive ability *short-term memory* (*Gsm*) is defined as *short-term working memory* (*Gwm*) in the Batería IV. This change reflects an update to the understanding of the importance of working memory in the active manipulation of information. Working memory can invoke other cognitive functions, including visualization as well as memory and retrieval processes, to effect goal attainment.
- Two tests measuring auditory processing (*Ga*)—*Procesamiento fonético* (Phonological Processing) and *Repetición de palabras sin sentido* (Nonword Repetition)—were added to the Batería IV COG battery. *Ga* abilities are now recognized as playing a pivotal scaffolding role in the development of language and general cognitive abilities.

¹Readers interested in learning about the recent revisions to CHC theory should consult Schneider and McGrew (2012, 2018). For a detailed description of all CHC broad and narrow abilities, see Appendix A of the WJ IV Technical Manual (McGrew et al., 2014).

• The Batería IV APROV *Fluidez en la lectura* (Reading Fluency) cluster contains two tests that conform to contemporary concepts of fluent reading performance: *Lectura oral* (Oral Reading) and *Fluidez en lectura de frases* (Sentence Reading Fluency). Together, these two tests provide a broad, cognitively complex reading fluency cluster intended to reduce the error inherent in generalizing fluent reading ability from a single, narrow aspect of performance, such as automaticity in word recognition.

Battery Organization

The Batería IV is organized into a cognitive battery and an achievement battery. Each battery contains one Test Book easel, a package of Test Records and examinee Response Booklets, an examiner's manual (specific to that battery), scoring guides, and an optional carrying case. Purchase of the Batería IV allows users to access the online scoring and reporting program. The online scoring and reporting program link provides access to the downloadable audio files for use with the applicable cognitive tests as well as the PDF file for the *Woodcock-Johnson IV Technical Manual* (McGrew et al., 2014).

The most educationally and diagnostically useful WJ IV tests were selected for inclusion in the Batería IV. Additional interpretive features are available when users combine the Batería IV with selected tests from the WJ IV OL and/or the *Batería III Woodcock-Muñoz: Pruebas de habilidades cognitivas* (Batería III COG; Muñoz-Sandoval, Woodcock, McGrew, & Mather, 2005b, 2007b).

Batería IV COG

The Batería IV COG includes all ten tests from the *Woodcock-Johnson IV Tests of Cognitive Abilities* (WJ IV COG; Schrank, McGrew, & Mather, 2014b) Standard Battery, plus the Number-Pattern Matching, Nonword Repetition, and Pair Cancellation tests from the WJ IV COG Extended Battery and the Rapid Picture Naming test from the WJ IV OL battery. Table 1 lists each test in the Batería IV COG and the corresponding WJ IV test number and name.

| Batería IV COG Test Number and Name | WJ IV Test Number and Name |
|--|---------------------------------------|
| Prueba 1: Vocabulario oral | Test 1: Oral Vocabulary |
| Prueba 2: Series numéricas | Test 2: Number Series |
| Prueba 3: Atención verbal | Test 3: Verbal Attention |
| Prueba 4: Pareo de letras idénticas | Test 4: Letter-Pattern Matching |
| Prueba 5: Procesamiento fonético | Test 5: Phonological Processing |
| Prueba 6: Rememoración de cuentos | Test 6: Story Recall |
| Prueba 7: Visualización | Test 7: Visualization |
| Prueba 8: Información general | Test 8: General Information |
| Prueba 9: Formación de conceptos | Test 9: Concept Formation |
| Prueba 10: Inversión de números | Test 10: Numbers Reversed |
| Prueba 11: Pareo de números idénticos | Test 11: Number-Pattern Matching |
| Prueba 12: Repetición de palabras sin sentido | Test 12: Nonword Repetition |
| Prueba 13: Cancelación de pares | Test 17: Pair Cancellation |
| Prueba 14: Rapidez en la identificación de dibujos | WJ IV OL Test 4: Rapid Picture Naming |

Table 1. Batería IV COG Tests Table 2 contains the Batería IV COG Selective Testing Table, illustrating the scope of interpretive information via combinations of tests that form various clusters. Note that several clusters require the administration of tests from either the WJ IV OL or the Batería III COG batteries.



Table 2. Batería IV COG Selective Testing Table

Tests required to create the cluster listed.

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

Batería IV APROV

The Batería IV APROV includes all 11 tests from the Woodcock-Johnson IV Tests of Achievement (WJ IV ACH; Schrank, Mather, & McGrew, 2014a) Standard Battery, plus the Reading Recall and Number Matrices tests from the WJ IV ACH Extended Battery. Table 3 lists each test in the Batería IV APROV and the corresponding WJ IV test number and name.

| Batería IV APROV Test Number and Name | WJ IV Test Number and Name |
|---|------------------------------------|
| Prueba 1: Identificación de letras y palabras | Test 1: Letter-Word Identification |
| Prueba 2: Problemas aplicados | Test 2: Applied Problems |
| Prueba 3: Ortografía | Test 3: Spelling |
| Prueba 4: Comprensión de textos | Test 4: Passage Comprehension |
| Prueba 5: Cálculo | Test 5: Calculation |
| Prueba 6: Expresión de lenguaje escrito* | Test 6: Writing Samples |
| Prueba 7: Análisis de palabras | Test 7: Word Attack |
| Prueba 8: Lectura oral | Test 8: Oral Reading |
| Prueba 9: Fluidez en lectura de frases | Test 9: Sentence Reading Fluency |
| Prueba 10: Fluidez en datos matemáticos | Test 10: Math Facts Fluency |
| Prueba 11: Fluidez en escritura de frases | Test 11: Sentence Writing Fluency |
| Prueba 12: Rememoración de lectura | Test 12: Reading Recall |
| Prueba 13: Números matrices | Test 13: Number Matrices |

*The WJ IV Writing Samples test was replaced with Written Language Expression (Expression de lenguaje escrito) in Batería IV. The two tests contain very similar item types; however, Written Language Expression scoring is simpler and does not require the examiner to use a separate scoring guide.

Table 4 contains the Batería IV APROV Selective Testing Table, illustrating the scope of interpretive information via combinations of tests that form various clusters.

| Table 4. | | | | | | | | I | Rea | ding | ļ | | Mat | hem | atics | : | Wri | iting | | Cros | s-D | oma | in Clusters |
|--|------|--------|----------------------------|------|---------|---------------|-------------------|-------------------|-----------------|-------------|----------|-------------------|------------------------|----------------------|-----------------|------------------------|------------------|-----------------|----------------|--------------------|-----------------|----------------------|------------------------|
| Batería IV APROV Selectiv Testing Table | ſĊ | | | Read | Brossis | Basic Feading | Read: Reading CL. | Reading Comprehis | Reading Fluency | Mathon Rate | Broad A. | Math Contennation | Math Provingition Suit | Written Land Solving | Broad Written I | Write Writing Cumulage | Acade Expression | Academic Skills | Acadon Fluenou | Acadon Application | Phones Anowleds | Brief Ica - Grapheme | I OL Broad Achievement |
| | | ACH 1 | Letter-Word Identification | | | | | | | | | | | | | | | | | | | | |
| | | ACH 2 | Applied Problems | | | | | | | | | | | | | | | | | | | | |
| | | ACH 3 | Spelling | | | | | | | | | | | | | | | | | | | | |
| | ery | ACH 4 | Passage Comprehension | | | | | | | | | | | | | | | | | | | • | |
| | Batt | ACH 5 | Calculation | | | | | | | | | | | | | | | | | | | • | |
| | E I | ACH 6 | Writing Samples | | | | | | | | | | | | | | | | | | | • | |
| | nda | ACH 7 | Word Attack | | | | | | | | | | | | | | | | | | | | |
| | Sta | ACH 8 | Oral Reading | | | | | | | | | | | | | | | | | | | | |
| | | ACH 9 | Sentence Reading Fluency | | | | | | | | | | | | | | | | | | | • | |
| | | ACH 10 | Math Facts Fluency | | | | | | | | | | | | | | | | | | | • | |
| | | ACH 11 | Sentence Writing Fluency | | | | | | | | | | | | | | | | | | | • | |
| - | | ACH 12 | Reading Recall | | | | | | | | | | | | | | | | | | | | |
| | | ACH 13 | Number Matrices | | | | | | | | | | • | | | | | | | | | | |
| | ery | ACH 14 | Editing | | | | | | | | | | | | | | | | | | | | |
| | Batt | ACH 15 | Word Reading Fluency | | | | | | | | | | | | | | | | | | | | |
| | ed F | ACH 16 | Spelling of Sounds | | | | | | | | | | | | | | | | | | | | |
| | bue | ACH 17 | Reading Vocabulary | | | | | | | | | | | | | | | | | | | | |
| | E I | 101110 | 0.1 | | 1 | | | | | | | | | | 1 | 1 | 1 | 1 | | | | _ | |

Table 3. Batería IV APROV Tests

Table 4.

Interpretation Features

The Batería IV online scoring and reporting program, included with the purchase of Batería IV Test Records, converts raw scores into derived scores. The range of interpretive information available for each test and cluster in the Batería IV includes information regarding testing behavior and examinee errors, developmental status, degree of proficiency, and comparison with grade or age peers. In contrast to that of many other test batteries, the interpretive design of the Batería IV enables the clinician to capitalize on the full range of criterion- and norm-referenced information. To aid examiners in the interpretation of an examinee's cognitive and achievement scores, the Batería IV reports can also include results from the WJ IV OL tests, when they are administered and committed to the online scoring and reporting program within 30 days of committing the Batería IV administration.

Language Background and Academic Language Exposure Information

The Batería IV Test Records include fields that are useful for gathering information about an examinee's language history. Although not required, when examiners enter this information into the online scoring and reporting program, it is included in the report and is important for ensuring accurate interpretation of the test scores.

The "Language Background Information" section allows examiners to document the language use and exposure of the examinee. In this section, the examiner is asked to indicate whether the examinee can be considered a native English speaker, a second-language learner of English, a native English speaker learning a foreign language or heritage language, or a simultaneous bilingual individual. Other information recorded in this section includes the examinee's native language; the language(s) spoken by others in the examinee's home; and the language(s) spoken by the examinee at home, with peers, and in the classroom.

The "Academic Language Exposure" section elicits information about current and prior language programs and the amount of time the examinee has spent in these programs, as well as information about academic language instruction outside of the United States. In the case of an examinee just entering a formal education setting or entering a new setting, there is also a space for the examiner to record information about the examinee's upcoming educational enrollment.

Comparative Language Index

A unique comparison procedure, the Comparative Language Index (CLI), is available when the parallel Spanish and English tests from the WJ IV OL have been administered. This comparison documents an individual's language proficiency in each language and helps determine which language is dominant. The parallel English and Spanish tests from the WJ IV OL battery are Test 1: Picture Vocabulary and *Prueba 10: Vocabulario sobre dibujos*, Test 2: Oral Comprehension and *Prueba 11: Comprensión oral*, and Test 6: Understanding Directions and *Prueba 12: Comprensión de indicaciones*. If all six tests are administered, three clusters are available for comparison: *Lenguaje oral* (Oral Language), *Amplio lenguaje oral* (Broad Oral Language), and *Comprensión auditiva* (Listening Comprehension). Examiners who wish to include the CLI information in the Batería IV report must administer the WJ IV OL tests and commit the results to the online scoring and reporting program within 30 days of committing the Batería IV administration.

CALP Levels

Cummins (1984) formalized a distinction between two types of language proficiency: basic interpersonal communication skill (BICS) and cognitive-academic language proficiency (CALP). BICS is defined as language proficiency in everyday communicative contexts, or those aspects of language proficiency that seem to be acquired naturally and without formal schooling. CALP is defined as language proficiency in academic situations, or those aspects of language proficiency that emerge and become distinctive with formal schooling. Classroom-appropriate academic proficiency is further defined by literacy skills involving conceptual-linguistic knowledge that occurs in a context of semantics, abstractions, and context-reduced linguistic forms. The online scoring and reporting program includes the option to report CALP levels to provide meaningful interpretations regarding the individual's language proficiency. If the option is selected, CALP levels can be reported for the *Comprensión-conocimiento* (*Gc*) cluster in the Batería IV APROV and the WJ IV OL that measure oral language, acquired knowledge, reading, and writing.

Test Design and Development

The Batería IV tests are parallel Spanish forms of the English tests from the WJ IV; both tests rely on the same set of norms to produce scores such as standard scores and percentile ranks. Using Rasch model equating procedures, all English WJ IV items and Spanish Batería IV items were calibrated onto the same scale, or "ruler." In the WJ IV and Batería IV this is called the *W* scale; examinees who take test forms constructed from the calibrated items are assigned *W* scores. The *W* score of an examinee who took the English form of a test from the WJ IV battery can be directly compared to the *W* score of another examinee who took the Spanish form of the same test from the Batería IV battery. These examinees' normative scores will reflect their relative standings in a distribution of their same-language peers of the same age or grade. This section describes the WJ IV and Batería IV, and the calibration study and subsequent equating of the Spanish items to the WJ IV *W* scale for each test.

Norming

The WJ IV norming study included data from 7,416 individuals from geographically diverse communities representing 46 states and the District of Columbia. The norming sample contained 664 children ages 2 through 5 years who were not enrolled in kindergarten; 3,891 examinees enrolled in kindergarten through grade 12; 775 college undergraduate and graduate students; and 2,086 adults (ages 18 and up) who were not enrolled in high school or college. Table 5 on page 8 displays the distribution of the WJ IV norming sample by age and grade. The higher density of examinees in the school-age population (kindergarten through grade 12) reflects the need to collect more concentrated data from examinees in this age range, where the abilities measured by the tests undergo the greatest rate of growth.

Table 5.

Distribution of the Norming Sample by Age and Grade

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

The norming sample was selected to be representative, within practical limits, of the U.S. population of individuals from ages 2 to 90+ years. Examinees were randomly selected within a stratified sampling design that controlled for several community and examinee variables. Table 6 contains the sampling variables and their distribution, both in the U.S. population according to the 2010 census projections and in the WJ IV norming sample, for the school-age sample. The *Woodcock-Johnson IV Technical Manual* (McGrew et al., 2014) provides similar information for the other major age groupings (preschool, college/university, and adult). Some variables were not relevant at all age levels of the norming sample. For example, occupational information was applied only to the adults in the sample, and type of college was applied only to the individuals enrolled in college.

Table 6.

Distribution of Sampling Variables in the U.S. Population and in the Norming Sample—Grades K Through 12

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

^a No reliable population information could be obtained.

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

^c AIANAT = American Indian or Alaska Native.

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

Data from the 7,416 norming study participants were summarized for each test and cluster. Individual examinee weights were applied during the norms construction process to ensure that the test, cluster, and difference score norms were based on a sample with characteristics proportional to the U.S. population distribution. The weight for each norming study participant was obtained by calculating the product of several partial weights, each corresponding to a demographic variable for the applicable sampling group (preschool, kindergarten through grade 12, college/university, or adult). For each demographic variable, if an examinee belonged to a category of the variable that was overrepresented in the norming study sample, the examinee's partial weight for that variable was less than 1.00. Likewise, if the examinee belonged to a category of the

variable that was underrepresented in the WJ IV norming study sample, the examinee's partial weight for that variable was greater than 1.00. Table 6 contains the partial weights assigned for each demographic variable value within the kindergarten through grade 12 sample of examinees. If demographic information was missing for an examinee on a particular variable, that examinee was assigned a null (1.00) partial weight for that variable. A partial weight of 1.00 is considered null because when it is multiplied with the other partial weights to compute a total norming study participant weight, a value of 1.00 has no effect on the overall weight. For some variables (indicated with superscript *b* in Table 6), null partial weights of 1.00 also were assigned to cells for which reliable population information could not be obtained or for which the sample counts were so low that they inappropriately skewed examinees' overall weights (e.g., the "Other, Not Hispanic" values of the race/ethnicity variable).

Calculation of Cluster Scores

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

Habilidad intelectual general (General Intellectual Ability; GIA) Cluster

The GIA is a general intelligence (*g*) score; it represents the first principal component obtained from principal component analysis (PCA). In PCA, the optimal weighted combination of tests that account for the largest proportion of the variance in a collection of tests is extracted as the first component (similar to a factor in factor analysis). Table 7 presents the GIA average smoothed *g* weights by 25 technical age groups.² A review of Table 7 reveals that the weights for the individual tests fluctuate little as a function of age.

Table 7.

Habilidad intelectual general Average (Smoothed) g Weights by Technical Age Groups

| | CHC | | | | | AGE | | | | |
|-------------------------------------|--------|------|------|------|------|------|------|------|------|------|
| WJ IV COG Test | Domain | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Prueba 1: Vocabulario oral | Gc | 0.16 | 0.16 | 0.16 | 0.17 | 0.17 | 0.17 | 0.18 | 0.18 | 0.18 |
| Prueba 2: Series numéricas | Gf | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 |
| Prueba 3: Atención verbal | Gwm | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 | 0.14 | 0.14 | 0.14 | 0.14 |
| Prueba 4: Pareo de letras idénticas | Gs | 0.17 | 0.16 | 0.16 | 0.15 | 0.14 | 0.12 | 0.11 | 0.11 | 0.10 |
| Prueba 5: Procesamiento fonético | Ga | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.17 |
| Prueba 6: Rememoración de cuentos | Glr | 0.11 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 |
| Prueba 7: Visualización | Gv | 0.07 | 0.07 | 0.07 | 0.08 | 0.08 | 0.09 | 0.10 | 0.10 | 0.11 |

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

Table 7. (cont.)Habilidad intelectualgeneral Average (Smoothed)g Weights by Technical AgeGroups

| | СНС | | | | | AGE | | | | |
|-------------------------------------|--------|------|------|------|------|------|------|------|------|------|
| WJ IV COG Test | Domain | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| Prueba 1: Vocabulario oral | Gc | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 |
| Prueba 2: Series numéricas | Gf | 0.18 | 0.18 | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 |
| Prueba 3: Atención verbal | Gwm | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 |
| Prueba 4: Pareo de letras idénticas | Gs | 0.10 | 0.10 | 0.10 | 0.10 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 |
| Prueba 5: Procesamiento fonético | Ga | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 |
| Prueba 6: Rememoración de cuentos | Glr | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 |
| Prueba 7: Visualización | Gv | 0.11 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 |

| | СНС | | AGE | | | | | | | |
|-------------------------------------|--------|-------|-------|-------|-------|-------|-------|------|--------|--|
| WJ IV COG Test | Domain | 20–29 | 30–39 | 40–49 | 50–59 | 60–69 | 70–79 | 80+ | Median | |
| Prueba 1: Vocabulario oral | Gc | 0.18 | 0.18 | 0.18 | 0.18 | 0.17 | 0.16 | 0.16 | 0.18 | |
| Prueba 2: Series numéricas | Gf | 0.16 | 0.16 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.17 | |
| Prueba 3: Atención verbal | Gwm | 0.13 | 0.14 | 0.14 | 0.14 | 0.15 | 0.15 | 0.15 | 0.14 | |
| Prueba 4: Pareo de letras idénticas | Gs | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | |
| Prueba 5: Procesamiento fonético | Ga | 0.17 | 0.17 | 0.17 | 0.18 | 0.18 | 0.18 | 0.18 | 0.17 | |
| Prueba 6: Rememoración de cuentos | Glr | 0.11 | 0.11 | 0.12 | 0.12 | 0.12 | 0.12 | 0.13 | 0.12 | |
| Prueba 7: Visualización | Gv | 0.13 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | |

Scholastic Aptitude Clusters

A major function of the Batería IV is to provide statements regarding a person's predicted performance in different achievement domains and to make comparisons between predicted and actual achievement within these domains. The four Batería IV Scholastic Aptitude (SAPT) cluster scores (two for the reading domain, and one each for the math and writing domains) are designed to provide optimal and efficient prediction of expected achievement in each domain. Each SAPT cluster score is based on a combination of four tests, each from a different CHC domain, that together produce the strongest and most efficient prediction for the selected achievement area. Although SAPTs were included in some prior editions of the battery, the Batería IV SAPTs represent an advance over those from the earlier batteries because they were constructed from statistical prediction combined with research and theoretical considerations. The Batería IV SAPTs also differ by academic areas, providing the best prediction of achievement skills. These advances were incorporated into the Batería IV SAPT clusters based on research indicating that predictor tasks vary within broad achievement areas and that these predictors change developmentally (McGrew, 2012; McGrew & Wendling, 2010; Schneider & McGrew, 2012, 2018).

Calculation of Norms

The development of test norms and derived scores requires the establishment of the normative (average) score for each measure for individuals at each specific age (age

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

Bootstrap Resampling Procedures

The innovative bootstrap resampling procedures (Efron & Tibshirani, 1993), first implemented and described for the *Woodcock-Johnson III Normative Update* (McGrew, Dailey, & Schrank, 2007), were used to calculate the WJ IV/Batería IV norms. The use of bootstrap resampling procedures allows for the incorporation of estimates of uncertainty and potential bias (in the sample data) in the calculation of the norms. When compared to more traditional norm-development procedures (such as those used in the Woodcock-Johnson [Woodcock & Johnson, 1977] and Batería [Woodcock, 1981]; WJ-R [Woodcock & Johnson, 1989] and *Batería Woodcock-Muñoz–Revisada* [Batería-R; Woodcock & Muñoz-Sandoval, 1996]; WJ III [Woodcock et al., 2001, 2007] and Batería III [Muñoz-Sandoval et al., 2005a, 2007a]; and most other individually administered cognitive ability, oral language, and academic achievement batteries), the bootstrap-based procedures used to calculate the WJ IV/Batería IV norms produce more precise estimates of an examinee's ability.

Difference Score Norms

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

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

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

One benefit of co-norming the COG and APROV batteries together is that it allows computation of actual differences between predictor and criterion variables for each individual in the norming sample, which can then be used to model these differences in the population. In the Batería IV, these types of difference scores take two forms: *variations* and *comparisons*. All Batería IV variation and comparison procedures are based on a common statistical model. What distinguishes variations from comparisons is the score that is used as the predictor in the model. While variations rely on a predictor score

that is an average of the scores from a pool of tests that excludes the criterion measure, comparisons rely on a single predictor, such as the GIA or *Gf-Gc combinado* cluster score. The scoring algorithms for these Batería IV difference-based variations and comparisons were constructed using a regression-based procedure similar to that used to calculate discrepancy scores in the WJ-R, Batería-R, WJ III, and Batería III batteries.

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

Translation and Adaptation Procedures

As mentioned earlier, all of the Batería IV tests are either translations or adaptations of the parallel tests from the WJ IV. Tests that are direct translations contain the same items as the WJ IV forms of the tests; for these tests, only the item instructions were translated into Spanish. Batería IV COG Prueba 2: Series numéricas is an example of a translated test. In this test, the stimulus material is exactly the same on the WJ IV and the Batería IV; the instructions are precisely parallel but are in different languages. In contrast, some tests could not be translated directly and needed to be adapted for use with Spanishspeaking individuals. A test is considered an adaptation when the measured construct is the same in English and Spanish, but the items were changed or adapted to be appropriate for Spanish-speaking examinees. For example, in Batería IV APROV Prueba 3: Ortografía, most Batería IV items are different from the WJ IV items, but the test measures the same broad and narrow abilities using the same procedure. Table 8 contains a list of the Batería IV tests and indicates whether each test was translated or adapted. In general, most of the visual processing, fluid reasoning, processing speed, and quantitative ability tests were translated, whereas the comprehension-knowledge, auditory, long-term storage and retrieval, reading, and writing tests required adaptation.

| Test Name | Translated | Adapted |
|-------------------------------------|------------|---------|
| Pruebas de habilidades cognitivas | | |
| Prueba 1: Vocabulario oral | | |
| Prueba 2: Series numéricas | | |
| Prueba 3: Atención verbal | | |
| Prueba 4: Pareo de letras idénticas | | |
| Prueba 5: Procesamiento fonético | | |
| Prueba 6: Rememoración de cuentos | | |
| Prueba 7: Visualización | | |
| Prueba 8: Información general | | |
| Prueba 9: Formación de conceptos | | |
| Prueba 10: Inversión de números | | |

Table 8.

Translated and Adapted Tests of the Batería IV **Table 8.** (cont.) Translated and Adapted Tests of the Batería IV

| Test Name | Translated | Adapted |
|--|------------|----------|
| Prueba 11: Pareo de números idénticos | | |
| Prueba 12: Repetición de palabras sin sentido | | |
| Prueba 13: Cancelación de pares | | |
| Prueba 14: Rapidez en la identificación de dibujos | 1 | |
| Pruebas de aprovechamiento | | |
| Prueba 1: Identificación de letras y palabras | | |
| Prueba 2: Problemas aplicados | | |
| Prueba 3: Ortografía | | |
| Prueba 4: Comprensión de textos | | |
| Prueba 5: Cálculo | | |
| Prueba 6: Expresión de lenguaje escrito | | 2 |
| Prueba 7: Análisis de palabras | | |
| Prueba 8: Lectura oral | 1 3 | |
| Prueba 9: Fluidez en lectura de frases | | |
| Prueba 10: Fluidez en datos matemáticos | | |
| Prueba 11: Fluidez en escritura de frases | | |
| Prueba 12: Rememoración de lectura | | |
| Prueba 13: Números matrices | | |

¹ This test is a direct translation of the WJ IV test with the exception of Item 104, which was changed from a football to a soccer ball in the Batería IV form.

² The WJ IV Writing Samples test was replaced with Written Language Expression (*Expresión de lenguaje escrito*) in Batería IV. The two tests contain very similar item types; however, Written Language Expression scoring is simpler and does not require the examiner to use a separate scoring guide.

³ Two items in the English form of this test did not translate well into Spanish; therefore, these items are slightly different in Batería IV.

The Batería IV test translation and adaptation work was performed by, or under the direction and supervision of, Dr. Criselda Alvarado. Some of the tests included in the Batería IV were translated or adapted during the development of the earlier editions of the Batería; other tests were new in the WJ IV and were translated into Spanish for the first time during the Batería IV development. An example of one such test is APROV *Prueba 8: Lectura oral.* For some adapted tests, Dr. Alvarado and her project team wrote new items to augment the existing Spanish item pools so that the Batería IV tests would contain new content and would be relevant for a wide range of Spanish-speaking examinees representing different linguistic and cultural backgrounds. For instance, for COG *Prueba 6: Rememoración de cuentos*, Dr. Alvarado and her team wrote 10 new stories containing a total of 113 new test items.

Calibration Study

Several tests required calibration, either because they were new tests in the Batería IV or because they contained new items. A calibration study was conducted that included six Batería IV tests: COG Prueba 5A: Procesamiento fonético – Acceso de palabras, COG Prueba 5C: Procesamiento fonético – Sustitución, COG Prueba 6: Rememoración de cuentos, COG Prueba 12: Repetición de palabras sin sentido, APROV Prueba 8: Lectura oral, and APROV Prueba 12: Rememoración de lectura. The primary goals of the study were to determine the difficulty levels of the new Spanish items and to equate those items to the scales underlying the English forms of the tests.³

³ Several additional adapted tests were not included in the calibration study because adequate item data from prior Spanish calibration studies existed to support construction of Batería IV test forms. These tests included *Prueba 1: Vocabulario oral* and *Prueba 8: Información general* in the COG battery and *Prueba 1: Identificación de letras y palabras*, *Prueba 2: Problemas aplicados*, *Prueba 3: Ortografía*, *Prueba 4: Comprensión de textos*, *Prueba 6: Expresión de lenguaje escrito*, and *Prueba 7: Análisis de palabras* in the APROV battery. The extant data from these earlier studies were used to equate the Spanish items for these tests to the scale underlying the English WJ IV tests, following the procedures described under "Calibration and Equating of Items" below. In addition, the extant item data for Batería III COG *Prueba 2: Aprendizaje visual-auditivo* and *Prueba 13: Reconocimiento de dibujos* were used to equate these two Batería III tests to the scales underlying the WJ IV forms of these tests so that the tests (and clusters that utilize the tests) can be scored with WJ IV/Batería IV norms.

Construction of Calibration Forms

For APROV *Prueba 8: Lectura oral*, the calibration form was almost a direct translation of Form *C* of the English WJ IV Oral Reading test. The only exceptions were two sentences that did not translate accurately into Spanish, for which comparable Spanish substitutions were made. For all other tests, the calibration forms contained a set of new Spanish items plus a set of "linking" items that were direct translations (or reasonable conceptual links) to items in the English forms of the tests. Linking items were distributed across the difficulty range of each test and served as statistical anchors in the Spanish-to-English equating process. The percentage of linking items on each calibration test form ranged from 25% for APROV *Prueba 12: Rememoración de lectura* to 43% for COG *Prueba 5A: Procesamiento fonético – Acceso de palabras*.

With the exception of APROV *Prueba 8: Lectura oral*, the calibration form of each test was approximately 10 to 15% longer than the targeted length for a published form of the test, to allow for flexibility to select the best-performing items after the calibration study. Traditional basal and ceiling rules and cutoff rules were used during administration of the calibration forms to minimize testing time, but the rules were set conservatively to ensure that every examinee in the study encountered all appropriately targeted items. Every examinee in the calibration study was administered all six tests.

Calibration Study Data Collection

The Batería IV calibration study was conducted between December 2017 and April 2018. In this study, the six Batería IV tests were administered to a sample of 601 native Spanish-speaking examinees between the ages of 2 and 81 years. Table 9 presents the distribution of the calibration sample by age group.

| Age | Number of Examinees | Percentage of Calibration Study Sample |
|-------|------------------------|--|
| 2–5 | 101 | 16.8 |
| 6—8 | 101 | 16.8 |
| 9–13 | 99 | 16.5 |
| 14–19 | 99 | 16.5 |
| 20–39 | 103 | 17.1 |
| 40+ | 98 | 16.3 |
| Total | 601 | 100.0 |

The calibration study examinees were selected from all regions of the United States. The sample was chosen to ensure a broad representation of sex, parent or examinee education level, and country of Hispanic origin/nativity. Table 10 on page 16 contains the distribution of these sampling variables in the calibration study.

| Tavic J. |
|--------------------------------|
| Distribution of the Batería IV |
| Calibration Sample by Age |
| Group |
| |

Tabla 0

Table 10.

Distribution of Sampling Variables in the Batería IV Calibration Study

| Sampling Variable | Number in Calibration Study | Percentage of Calibration Study Sample |
|---|--------------------------------|--|
| Sex | - | |
| Male | 246 | 40.9 |
| Female | 355 | 59.1 |
| Parent ¹ or Examinee Education | | |
| High School or Less | 301 | 50.1 |
| > High School | 300 | 49.9 |
| Geographic Location | | |
| Arizona | 3 | 0.5 |
| California | 22 | 3.7 |
| Connecticut | 1 | 0.2 |
| Florida | 5 | 0.8 |
| Illinois | 99 | 16.5 |
| New Jersey | 5 | 0.8 |
| New York | 7 | 1.2 |
| Tennessee | 67 | 11.2 |
| Texas | 375 | 62.4 |
| Virginia | 17 | 2.8 |
| Hispanic Origin | | |
| Cuban | 9 | 1.5 |
| Dominican | 5 | 0.8 |
| Guatemalan | 12 | 2.0 |
| Mexican | 421 | 70.1 |
| Puerto Rican | 33 | 5.5 |
| Salvadoran | 17 | 2.8 |
| Other/Mixed | 104 | 17.3 |

¹ Parent education is reported for examinees who are less than 18 years old.

Calibration and Equating of Items

At the completion of the data collection, data were analyzed using the Rasch model. The item data were freely calibrated and item W difficulties were estimated. Item difficulties for the Spanish form of each test were then linked to the W scale underlying each corresponding English test through the Rasch equating procedures (Wright & Stone, 1979)⁴ described below:

1. *Identify stable common linking items.* For each test, the separate Spanish and English item difficulties for the common items were cross-plotted. Extreme outliers, identified using a linear regression procedure, revealed some items with very different relative W-difficulty estimates in Spanish and English. These outlier items were removed from the common item linking set.

⁴ Wolfe (2004) terms this type of equating the "equating constants" method, while Linacre (2012) refers to it as the "Fahrenheit-Celsius" method. This method differs from the Rasch common-item-anchor equating design employed in the WJ IV norming (and described in the WJ IV Technical Manual) in that the item difficulty parameters for each data set are estimated separately, and the difficulty measures from one set of items are then transformed onto the other scale outside of the estimation process.

2. Apply the scale transformation equation. For each test, item W-difficulty means (Ms) and standard deviations (*SDs*) were computed for the subsets of common items from the Spanish and English item pools. Spanish item W-difficulty values were then adjusted to the scale of the English item pools using the following unit transformation equation:

$$D_{e'} = \frac{SD_e}{SD_s} (D_s - M_s) + M_e \,, \tag{1}$$

where $D_{e'}$ is the item difficulty of any Spanish item transformed onto the English item difficulty scale, SD_e is the standard deviation of the English common-item difficulties, SD_s is the standard deviation of the Spanish common-item difficulties, D_s is the difficulty of the Spanish item to be transformed, M_s is the mean of the Spanish common-item difficulties, and M_e is the mean of the English common-item difficulties. Application of this transformation equation placed the Spanish items onto the scale of the WJ IV English item pools.

Review of Item Statistics

In addition to content and bias considerations, the authors relied on both classical and Rasch-based statistical information to guide the item selection. In general, items under the following conditions were flagged and removed from consideration:

- 1. *Items with point-measure correlations less than .20.* The item point-measure correlation is the correlation between each examinee's W score and his or her score (1/0) on the item. This statistic provides insight into how well each item discriminates between low- and high-ability examinees. Items with point-measure correlations less than .20 may not discriminate well or may be measuring something other than what is intended by the other items in the scale.
- 2. *Rasch mean-square fit statistics greater than* 1.30. Rasch fit statistics describe the difference between an item's expected scores (i.e., under the Rasch model) and its observed scores (i.e., in the data). Mean-square fit statistics have an expected value of 1.0; values greater than 1.3 indicate that there may be more "noise" than useful measurement in the data. Low fit values (< 0.7) indicate that the item responses are more predictable than expected; this condition may reduce the statistical information in each item response but does not degrade measurement to the extent that values greater than 1.3 do.

Item Bias Analysis

Bias in item difficulty is often referred to as differential item functioning, or DIF. DIF occurs when an item is more difficult for a particular subgroup of examinees, even when the overall ability of those examinees is the same as that of other groups. For the Batería IV calibration items, gender DIF was evaluated during item calibration using the Rasch iterative-logit method within the WINSTEPS software (Linacre, 2012). In this method, item difficulty calibrations, and their associated standard errors, are estimated for each item and each subgroup individually, while all other item difficulty estimates (and examinee ability estimates) are held constant. The difference between the subgroup item difficulty estimates for each item, or the DIF contrast, was then evaluated using Welch's *t* statistic for the difference between two means (Linacre, 2012). Items were flagged if the DIF contrast between males and females was greater than or equal to 5.82 W points.⁵

⁵A DIF contrast with a W-point difference greater than or equal to 5.82 W points (i.e., 0.64 × 9.1024 W points, which is the value of 1 Rasch logit) corresponds to the commonly used Educational Testing Service (ETS) "C" classification for moderate to severe DIF (Linacre, 2012; Zieky, 1993).

Items were also flagged if significant (p < .05) Rasch-Welch *t*-test⁶ or Mantel-Haenszel DIF⁷ statistics were reported. Items were flagged regardless of the direction of the apparent bias. The percentages of flagged items with both DIF contrast greater than or equal to 5.82 points and significance at the p < .05 level are reported for the Batería IV calibration study tests in Table 11.

| | Total Number of | Percentage of Items More Difficult for | |
|---|----------------------|---|---------|
| Test Name | Calibration Items | Males | Females |
| COG Prueba 5A: Procesamiento fonético – Acceso de palabras | 27 | 3.7 | 0.0 |
| COG Prueba 5C: Procesamiento fonético – Sustitución | 23 | 0.0 | 4.3 |
| COG Prueba 6: Rememoración de cuentos | 159 | 3.1 | 7.5 |
| COG Prueba 12: Repetición de palabras sin sentido | 51 | 0.0 | 5.9 |
| APROV Prueba 8: Lectura oral | 27 | 0.0 | 0.0 |
| APROV Prueba 12: Rememoración de lectura | 143 | 0.7 | 3.5 |

Table 11.

Percentage of Batería IV Calibration Items Flagged for Potential Gender DIF

Assembly and Evaluation of Final Test Forms

After all test items had been placed onto the underlying W scales, the authors, with assistance from several native Spanish-speaking education and language professionals, chose items for the publication forms of the Batería IV. During the assembly of these forms, the authors followed some general principles of test construction. For instance, a common goal across all Batería IV tests is that items be evenly distributed across the W-score range of the test, with approximately three to four items per 10 W points of difficulty. Item content was chosen that would be current and relevant to as large an audience as possible, including individuals from a variety of Spanish-speaking countries. Finally, care was taken to ensure that no item cued the correct response to any other item in the same test.

Reliability

Reliability refers to the *precision* of a test score. High reliability indicates that an individual's measure on a test would be unlikely to change if he or she were retested under similar conditions. Reliability is a necessary, but not sufficient, condition for validity. Although high reliability does not necessarily imply that a test score is valid for a specific purpose, reliability is an important element of the overall validity argument for a test. The reliability coefficient can be thought of as an index of the precision with which relative standing or position in a group is measured.

⁶ In a test of 20 items, one would expect one item to exhibit significant DIF by chance (p < .05, the Type I error rate). Several authors (Linacre, 2012; Wolfe et al., 2006) suggest the use of the Bonferroni correction to adjust for Type I error when performing multiple statistical tests. Because the purpose of this DIF analysis was exploratory-items exhibiting significant DIF contrast were not rejected outright but rather were flagged for further review-no correction was applied in these analyses. The numbers of pairwise t tests in the analysis of DIF for each test suggests that some unbiased items were likely flagged; however, this potential overidentification was deemed acceptable for the purposes of this DIF study. ⁷ The Mantel-Haenszel procedure is a statistical approach that utilizes a contingency table to test the significance of score differences between a referent and a focal

group across an ability continuum.

Test Reliabilities

For the six tests that were included in the Batería IV calibration study, reliability coefficients were calculated using item-level data from the calibration study. For all other tests in the Batería IV battery, reliability coefficients were calculated using item-level data from the norming study.

For most nontimed, or nonspeeded, tests, internal consistency reliabilities were calculated using the split-half procedure. Raw scores were computed based on the oddand even-numbered items, and correlations were computed between these sets of scores. However, the split-half procedure is inappropriate⁸ for tests containing multiple-point items (e.g., APROV *Prueba 8: Lectura oral*). Therefore, the reliabilities for these tests were calculated using information provided by the Rasch model.

Test-retest reliabilities are provided for the Batería IV speeded tests (e.g., COG *Prueba 4: Pareo de letras idénticas*, COG *Prueba 5B: Procesamiento fonético – Fluidez de palabras*). These reliability coefficients were computed from a test-retest study that was conducted during the WJ IV norming for all speeded tests. Examinees in three separate age groups were administered the norming form of each speeded test, followed by a second administration of the same form of the test 1 day later. The retest interval in this study was intentionally short to minimize changes in test scores due to changes in the examinee's state or latent trait. Correlations between the first and second administrations were computed, and a correction was applied for restriction of range in the study samples (Sackett & Yang, 2000).

For the tests with subtests (COG Prueba 1: Vocabulario oral, COG Prueba 5: Procesamiento fonético, COG Prueba 7: Visualización, and COG Prueba 8: Información general), test reliabilities were computed using Mosier's (1943) formula for reliability of composite scores. Details of the procedures for computing reliabilities are included in Chapter 4 of the WJ IV Technical Manual.

All reliability coefficients were corrected for published test length using the Spearman-Brown correction formula. Table 12 presents the median reliability coefficients (*r*) and standard errors of measurement (*SEM*) in standard score (SS) units for the nonspeeded tests included in the Batería IV. Table 13 on page 20 presents the median reliabilities for the speeded tests from the speeded test-retest study.

| Test | - | | Test | | |
|--|--------------------|---------------------------|--|------------------------|---------------------------|
| Pruebas de habilidades cognitivas | Median <i>r</i> 11 | Median <i>SEM</i> (SS) | Pruebas de aprovechamiento | Median r ₁₁ | Median <i>SEM</i> (SS) |
| Prueba 1: Vocabulario oral | 0.89 | 4.97 | Prueba 1: Identificación de letras y palabras | 0.94 | 3.78 |
| Prueba 2: Series numéricas | 0.91 | 4.64 | Prueba 2: Problemas aplicados | 0.91 | 4.27 |
| Prueba 3: Atención verbal | 0.86 | 5.70 | Prueba 3: Ortografía | 0.92 | 4.13 |
| Prueba 5: Procesamiento fonético | 0.85 | 6.00 | Prueba 4: Comprensión de textos | 0.89 | 5.00 |
| Prueba 6: Rememoración de cuentos | 0.95 | 3.90 | Prueba 5: Cálculo | 0.93 | 3.86 |
| Prueba 7: Visualización | 0.85 | 5.81 | Prueba 6: Expresión de lenguaje escrito | 0.79 | 4.74 |
| Prueba 8: Información general | 0.88 | 5.20 | Prueba 7: Análisis de palabras | 0.91 | 4.75 |
| Prueba 9: Formación de conceptos | 0.93 | 4.04 | Prueba 8: Lectura oral | 0.9 | 3.00 |
| Prueba 10: Inversión de números | 0.88 | 5.15 | Prueba 12: Rememoración de lectura | 0.97 | 4.34 |
| Prueba 12: Repetición de palabras sin sentido | 0.91 | 4.55 | Prueba 13: Números matrices | 0.92 | 4.31 |

⁸ Internal consistency reliability methods, such as the split-half procedure, assume that the average correlation between items within a test is the same as the average correlation between items from the hypothetical alternative forms created by splitting the test into two smaller tests (e.g., odd and even items). This assumption is violated when tests contain items that produce a different range of scores for each item (as in the Batería IV tests with multiple-point item scoring). In this case, splitting the test in half may produce tests that are no longer equivalent; the items on one half of the test may have a higher maximum possible total score than the items on the other half. More detailed information about the Rasch reliability calculations is included in Chapter 4 of the WJ Technical Manual (McGrew et al., 2014).

| Median Reliability |
|----------------------|
| Coefficients and |
| Standard Errors |
| of Measurement |
| in Standard Score |
| Units for Batería IV |
| Nonspeeded Tests |

Table 12.

Table 13.

Median Test-Retest Reliability Coefficients for Batería IV Speeded Tests

| Test | Median r ₁₂ |
|--|------------------------|
| COG Prueba 4: Pareo de letras idénticas | 0.91 |
| COG Prueba 11: Pareo de números idénticos | 0.85 |
| COG Prueba 13: Cancelación de pares | 0.89 |
| COG Prueba 14: Rapidez en la identificación de dibujos | 0.90 |
| APROV Prueba 9: Fluidez en lectura de frases | 0.93 |
| APROV Prueba 10: Fluidez en datos matemáticos | 0.95 |
| APROV Prueba 11: Fluidez en escritura de frases | 0.83 |

Cluster Reliabilities

Cluster reliabilities were also computed using Mosier's (1943) formula for composite reliability. Table 14 presents the median cluster reliabilities and standard errors of measurement in standard score units for all Batería IV clusters.

Table 14.

Median Reliability Coefficients and Standard Errors of Measurement in Standard Score Units for Batería IV Clusters

| Cluster | | | Cluster | | |
|--|------------------------|---------------------------|--|------------------------|---------------------------|
| Pruebas de habilidades cognitivas | Median r ₁₁ | Median <i>SEM</i> (SS) | Pruebas de aprovechamiento | Median r ₁₁ | Median <i>SEM</i> (SS) |
| Habilidad intelectual general | 0.97 | 2.60 | Lectura | 0.95 | 3.35 |
| Habilidad intelectual breve | 0.94 | 3.67 | Lectura amplia | 0.96 | 2.80 |
| Gf-Gc combinado | 0.95 | 3.00 | Destrezas básicas en lectura | 0.95 | 3.35 |
| Comprensión-conocimiento (Gc) | 0.93 | 3.97 | Comprensión de lectura | 0.94 | 3.87 |
| Comprensión-conocimiento – Extendida | 0.95 | 3.67 | Fluidez en la lectura | 0.95 | 3.00 |
| Razonamiento fluido (Gf) | 0.94 | 3.67 | Matemáticas | 0.96 | 3.00 |
| Memoria de trabajo a corto plazo (Gwm) | 0.91 | 4.50 | Matemáticas amplias | 0.97 | 2.60 |
| Velocidad de procesamiento cognitivo (Gs) | 0.94 | 3.67 | Destrezas en cálculos matemáticos | 0.97 | 2.60 |
| Procesamiento auditivo (Ga) | 0.92 | 4.24 | Resolución de problemas matemáticos | 0.95 | 3.35 |
| Destreza numérica | 0.90 | 4.74 | Lenguaje escrito | 0.92 | 3.67 |
| Rapidez perceptual | 0.93 | 3.97 | Lenguaje escrito amplio | 0.94 | 3.35 |
| Vocabulario | 0.93 | 4.97 | Expresión escrita | 0.88 | 4.24 |
| Eficiencia cognitiva | 0.93 | 3.97 | Destrezas académicas | 0.95 | 2.60 |
| Eficiencia cognitiva—Extendida | 0.95 | 3.35 | Fluidez académica | 0.97 | 2.60 |
| Aptitud de lectura 1 | 0.93 | 4.97 | Aplicaciones académicas | 0.95 | 3.00 |
| Aptitud de lectura 2 | 0.93 | 4.86 | Aprovechamiento breve | 0.96 | 2.60 |
| Aptitud matemática | 0.94 | 4.97 | Aprovechamiento amplio | 0.98 | 1.50 |
| Aptitud de escritura | 0.92 | 4.86 | | | |

Evidence to Support the Use and Interpretation of Batería IV Scores

The WJ IV Technical Manual (McGrew et al., 2014) outlines several propositions for the use and interpretation of the WJ IV and Batería IV scores and provides support for each proposition within a framework consistent with that outlined in the *Standards for Educational and Psychological Testing* (AERA et al., 2014). This section summarizes the evidence relevant to test content, process, and construct coverage and provides evidence for the developmental patterns of the Batería IV ability clusters. Readers interested in validity evidence to support the internal structure of the battery should consult the WJ IV Technical Manual.

Representativeness of Content, Process, and Construct Coverage

The Batería IV includes tests measuring a complex set of unique abilities, defined by CHC theory, constituting cognitive ability and academic achievement. Evidence to support this proposition, often termed *content validity evidence* or *substantive validity evidence*, for the Batería IV test scores is provided via the specification of test and cluster content according to contemporary CHC theory and research.⁹ This aspect of the validity argument builds upon the theories contained in the three prior editions of the Woodcock-Johnson and Batería batteries. The Batería IV test design blueprint pushes the design of tests "beyond CHC theory" (McGrew, 2012; Schneider & McGrew, 2012, 2018) as CHC theory was defined in the WJ III and Batería III. In addition to its reliance on contemporary CHC theory as the basis for the overarching test battery design blueprint, the Batería IV plan was influenced by the incorporation of contemporary findings from neurocognitive, neuropsychological, and developmental research.

The distinction between broad and narrow abilities is an important concept in CHC theory. As in the Batería III, most of the Batería IV tests were designed to measure one narrow ability. This CHC-based test design approach, first operationalized in the WJ III and Batería III, focuses on increasing CHC construct representation and decreasing construct-irrelevant variance in tests (Benson, 1998; McGrew & Flanagan, 1998; Messick, 1995). To increase breadth, clusters were constructed to subsume two or more qualitatively different narrow abilities. The principle of cluster interpretation was adopted to improve the content validity of measures for broad abilities such as reading, fluid reasoning, and general intelligence.

The five broad CHC clusters in the Batería IV COG include Razonamiento fluido (Fluid Reasoning; *Gf*), *Comprensión-conocimiento* (Comprehension-Knowledge; *Gc*), *Memoria de trabajo a corto plazo* (Short-Term Working Memory; *Gwm*), *Velocidad de procesamiento cognitivo* (Cognitive Processing Speed; *Gs*), and *Procesamiento auditivo* (Auditory Processing; *Ga*). By administering two additional tests from the *Batería III Pruebas de habilidades cognitivas*, users can also obtain the *Almacenamiento visual* (Visual Processing; *Gv*) clusters. Two-test narrow CHC ability clusters are available for *Destreza numérica* (Number Facility) and *Rapidez perceptual* (Perceptual Speed). Cognitive efficiency, which represents the amalgam of processing speed (*Gs*) and short-term working memory (*Gwm*), is represented by two- and four-test clusters (*Eficiencia cognitiva* and *Eficiencia cognitiva – Extendida*).

The Batería IV APROV contains 13 tests that tap two other identified CHC cognitive abilities—quantitative knowledge (Gq) and reading and writing ability (Grw). The

⁹ Refer to Chapter 1 and Appendix A of the WJ IV Technical Manual (McGrew et al., 2014) for a description of contemporary CHC theory and the tests and clusters contained in the Batería IV, which, as noted above, are parallel forms of the WJ IV tests and clusters.

Batería IV APROV also includes additional measures of comprehension-knowledge (Gc), long-term storage and retrieval (*Glr*), and auditory processing (*Ga*). The 13 Batería IV APROV tests were developed to measure the major aspects of academic achievement in Spanish, including reading, mathematics, and written language. The specification of item content in these tests was based primarily on the goal of providing a broad sampling of achievement areas rather than an in-depth assessment of a relatively narrow area.

Table 15 and Table 16 (on page 24) provide descriptions of the broad and narrow constructs measured by the Batería IV COG and Batería IV APROV tests, respectively, as well as stimulus and response characteristics, task requirements, and inferred cognitive processes.

| Cognitive Test | Primary Broad CHC Ability Narrow Ability | Stimuli | Task Requirements | Cognitive Processes | Response |
|---|---|---------------------------------|--|--|--|
| 1: Vocabulario oral A: Sinónimos B: Antónimos | Comprehension- Knowledge (<i>Gc</i>) <i>Lexical knowledge</i> (VL) <i>Language development</i> (LD) | Auditory (words) | Listening to a word and providing a synonym; listening to a word and providing an antonym | Semantic activation, access, and matching | Oral (words) |
| 2: Series numéricas | Fluid Reasoning (<i>Gt</i>) <i>Quantitative reasoning</i> (RQ) <i>Induction</i> (I) | Visual (numeric) | Determining a numerical sequence | Representation and manipulation of points on a mental number line; identifying and applying an underlying rule/ principle to complete a numerical sequence | Oral (numbers) |
| 3: Atención verbal | Short-Term Working Memory (<i>Gwm</i>) <i>Working memory capacity</i> (WM) <i>Attentional control</i> (AC) | Auditory (words, numbers) | Listening to a series of numbers and animals intermingled and answering a specific question regarding the sequence | Controlled executive function; working memory capacity; recoding of acoustic, verbalized stimuli held in immediate awareness; selective auditory attention; attentional control | Oral (words) |
| 4: Pareo de letras idénticas | Processing Speed (<i>Gs</i>) <i>Perceptual speed</i> (P) | Visual (letters) | Rapidly locating and circling identical letters or letter patterns | Speeded visual perception and matching; visual discrimination; orthographic processing; divided attention | Motoric (circling) |
| 5: Procesamiento fonético A: Acceso de palabras B: Fluidez de palabras C: Sustitución | Auditory Processing (Ga) Phonetic coding (PC) Word fluency (GIr-FW) Speed of lexical access (GIr-LA) | Auditory (words) | Providing a word with a specific phonic element; naming as many words as possible that begin with a specified sound; substituting part of a word to make a new word | Semantic activation, access; speed of lexical access | Oral (words) |
| 6: Rememoración de cuentos | Long-Term Storage and Retrieval (<i>GIr</i>) <i>Meaningful memory</i> (MM) <i>Listening ability</i> (<i>Gc</i> -LS) | Auditory (text) | Listening to and recalling details of stories | Construction of propositional representations and recoding | Oral (passages) |
| 7: Visualización A: Relaciones espaciales B: Rotación de bloques | Visual Processing (<i>Gv</i>) <i>Visualization</i> (Vz) | Visual (shapes, designs) | Identifying two- dimensional pieces that form a shape; identifying two three- dimensional rotated block patterns that match a target | Visual feature detection; manipulation (mental rotation) of visual images in space; matching | Oral (letters) or Motoric (pointing) |

Table 15.

Batería IV COG Test Content, Process, and Construct Descriptions **Table 15.** (cont.)Batería IV COG TestContent, Process, andConstruct Descriptions

| | | | · · · · · · · · · · · · · · · · · · · | | |
|---|--|---------------------------------|---|--|---------------------------------|
| Cognitive Test | Primary Broad CHC Ability Narrow Ability | Stimuli | Task Requirements | Cognitive Processes | Response |
| 8: Información general A: Dónde B: Qué | Comprehension- Knowledge (<i>Gc</i>) <i>General (verbal)</i> <i>information</i> (K0) | Auditory (questions) | Identifying where an object is found and what people typically do with an object | Semantic activation and access to declarative generic knowledge | Oral (phrases, sentences) |
| 9: Formación de conceptos | Fluid Reasoning (<i>Gf</i>) <i>Induction</i> (I) | Visual (drawings) | ldentifying, categorizing, and determining rules | Rule-based categorization; rule switching; induction/ inference | Oral (words) |
| 10: Inversión de números | Short-Term Working Memory (<i>Gwm</i>) <i>Working memory capacity</i> (WM) <i>Attentional control</i> (AC) | Auditory (numbers) | Listening to and recalling a sequence of digits in reversed order | Span of apprehension and recoding in working memory; working memory capacity, attentional capacity | Oral (numbers) |
| 11: Pareo de números idénticos | Processing Speed (<i>Gs</i>) <i>Perceptual speed</i> (P) | Visual (numbers) | Rapidly locating and circling identical numerals from a defined set | Speeded visual perception and matching; visual discrimination; divided attention | Motoric (circling) |
| 12: Repetición de palabras sin sentido | Auditory Processing (<i>Ga</i>) Phonetic coding (PC) Memory for sound patterns (UM) Memory span (Gwm-MS) | Auditory (nonsense words) | Listening to a nonsense word and repeating it exactly | Analysis of a sequence of acoustic phonological elements in immediate awareness; efficiency of the phonological loop | Oral (words) |
| 13: Cancelación de pares | Processing Speed (<i>Gs</i>) Perceptual speed (P) Spatial scanning (Gv-SS) Attentional control (Gwm-AC) | Visual (drawings) | Rapidly locating and marking a repeated pattern | Executive processing; attentional control; inhibition and interference control; sustained attention | Motoric (circling) |
| 14: Rapidez en la identificación de dibujos | Processing Speed (<i>Gs</i>) <i>Speed of lexical access</i> (LA) | Visual (pictures) | Recognizing objects, then retrieving and articulating their names rapidly | Speed/fluency of retrieval and oral production of recognized objects; speeded serial naming; rapid object recognition | Oral (words) |

Table 16.

Batería IV APROV Test Content, Process, and Construct Descriptions

| Achievement Test | Primary Broad CHC Ability Narrow Ability | Stimuli | Task Requirements | Cognitive Processes | Response |
|---|--|--|--|---|----------------------------------|
| 1: Identificación de letras y palabras | Reading & Writing Ability (<i>Grw</i>) <i>Reading decoding</i> (RD) | Visual (text) | Identifying printed letters and words | Feature detection and analysis (for letters) and recognition of visual word forms from a phonological lexicon; access of pronunciations associated with visual word forms | Oral (letter names, words) |
| 2: Problemas aplicados | Quantitative Knowledge (Gq) Mathematical achievement (A3) Fluid Reasoning (Gf) Quantitative reasoning (RQ) | Auditory (questions) Visual (numeric, text) | Performing math calculations in response to orally presented problems | Construction of mental models via language comprehension, application of calculation and/or quantitative reasoning; formation of insight | Oral (numbers, words) |
| 3: Ortografía | Reading & Writing Ability (<i>Grw</i>) <i>Spelling ability</i> (SG) | Auditory (words) | Spelling orally presented words | Access to and application of knowledge of orthography of word forms by mapping whole- word phonology onto whole-word orthography, by translating phonological segments into graphemic units, or by activating spellings of words from the semantic lexicon | Motoric (writing) |
| 4: Comprensión de textos | Reading & Writing Ability (<i>Grw</i>) <i>Reading comprehension</i> (RC) | Visual (text) | Identifying a missing key word that makes sense in the context of a written passage | Construction of propositional representations; integration of syntactic and semantic properties of printed words and sentences into a representation of the whole passage | Oral (words) |
| 5: Cálculo | Quantitative Knowledge (Gq) <i>Mathematical</i> <i>achievement</i> (A3) | Visual (numeric) | Performing various mathematical calculations | Access to and application of knowledge of numbers and calculation procedures; verbal associations between numbers represented as strings of words | Motoric (writing) |
| 6: Expresión de lenguaje escrito | Reading & Writing Ability (<i>Grw</i>) <i>Writing ability</i> (WA) | Auditory (text) Visual (text) | Writing meaningful sentences for a given purpose | Retrieval of word meanings; application of psycholinguistic rules of case, grammar, and syntax; planning and construction of bridging inferences in immediate awareness (auditory and/ or visual buffer) | Motoric (writing) |
| 7: Análisis de palabras | Reading & Writing Ability (Grw) Reading decoding (RD) Auditory Processing (Ga) Phonetic coding (PC) | Visual (word) | Reading phonically regular nonwords | Grapheme-to-phoneme translation and accessing pronunciations of pseudowords not contained in the mental lexicon | Oral (words) |
| 8: Lectura oral | Reading & Writing Ability (Grw) Reading comprehension (RC) Verbal (print) language comprehension (V) | Visual (text) | Reading sentences orally with accuracy and fluency | Integration of orthographic, phonological, and semantic processes; articulatory planning and motor execution | Oral (sentences) |

Table 16. (cont.)Batería IV APROV TestContent, Process, andConstruct Descriptions

| Achievement Test | Primary Broad CHC Ability Narrow Ability | Stimuli | Task Requirements | Cognitive Processes | Response |
|--|--|---------------------------------|--|--|-----------------------|
| 9: Fluidez en lectura de frases | Reading & Writing Ability (Grw) Reading comprehension (RC) Reading speed (RS) Processing Speed (GS) | Visual (text) | Reading printed statements rapidly and responding true or false (yes or no) | Speeded semantic decision making requiring reading ability and generic knowledge | Motoric (circling) |
| 10: Fluidez en datos matemáticos | Quantitative Knowledge (Gq) Mathematical achievement (A3) Processing Speed (Gs) Number facility (N) | Visual (numeric) | Adding, subtracting, and multiplying rapidly | Speeded access to and application of digit-symbol arithmetic procedures | Motoric (writing) |
| 11: Fluidez en escritura de frases | Reading & Writing Ability (Grw) Writing ability (WA) Writing speed (WS) Processing Speed (GS) | Visual (words with pictures) | Formulating and writing simple sentences rapidly | Speeded formation of constituent sentence structures requiring fluent access to semantic and syntactic knowledge | Motoric (writing) |
| 12: Rememoración de lectura | Reading & Writing Ability (Grw) Reading comprehension (RC) Long-Term Storage and Retrieval (GIr) Meaningful memory (MM) | Visual (text) | Reading and recalling details of stories | Construction of propositional representations and recoding | Oral (passages) |
| 13: Números matrices | Fluid Reasoning (<i>Gf</i>) <i>Quantitative reasoning</i> (RQ) | Visual (numeric) | Determining a two- dimensional numerical pattern | Access to verbal- visual numeric codes; transcoding verbal and/ or visual representations of numeric information into analogical representations; determining the relationship between/ among numbers on the first part of the structure and mapping (projecting) the structure to complete the analogy | Oral (numbers) |

Developmental Patterns of Batería IV Ability Clusters

The Batería IV tests and clusters display average score changes consistent with the developmental growth and decline of cognitive abilities and achievement across the life span. Divergent growth curves provide evidence for the existence of distinct, unique abilities (Carroll, 1993). Figures 1 through 3 on pages 26 and 27 present examples of growth curves, or "difference curves," from ages 6 to 90 years for several Batería IV COG and Batería IV APROV clusters.¹⁰ The difference curves illustrate that the unique abilities measured by the Batería IV follow different developmental courses or trajectories over the age span from childhood to geriatric levels. The examples were constructed using age 6 years, 0 months (6-0) as a starting point and subtracting the norm-based REF W score for age 6-0 for each cluster from all other REF Ws for that cluster through age 90. This procedure produced difference curves all starting with an assigned common origin of zero.

¹⁰ These curves are based on cross-sectional data collection and are not to be confused with growth curves based on longitudinal data collection designs.



*Requires one test from the Batería III Pruebas de habilidades cognitivas.



Plot of Batería IV COG GIA and four narrow and other clinical cluster W-score difference curves by age.



Figure 3.

Plot of 11 Batería IV APROV cluster W-score difference curves by age.



Figure 1 presents difference curves for the GIA (*g*), seven CHC cognitive factor clusters, and the *Gf-Gc combinado* (*Gf-Gc* Composite). The patterns of growth and decline of the seven Batería IV CHC cognitive factor clusters differ markedly, providing evidence to support the existence of distinct abilities. Figure 2 demonstrates the existence of distinct patterns of growth and decline among the abilities measured by the Batería IV COG narrow ability factors and clinical clusters of *Vocabulario* (Vocabulary; VL/LD), *Eficiencia cognitiva* (Cognitive Efficiency; *Gs* + *Gwm*), *Destreza numérica* (Number Facility; N), and *Rapidez perceptual* (Perceptual Speed; P). The GIA (g) curve is included for comparison purposes.

Growth curves for the 11 broad and narrow Batería IV APROV clusters are presented in Figure 3. A number of points regarding achievement cluster measures are apparent in Figure 3. First, a majority of the Batería IV APROV clusters show rapid acceleration of growth from age 6 through approximately 15 years. Second, the majority of achievement levels peak at a much higher point relative to their origin (in this case, 6 years) than the cognitive abilities do. Third, most achievement levels do not demonstrate as much absolute decline across the age span as the cognitive abilities do; the achievement skills are generally maintained at higher levels into the older age ranges. These three features distinguish the achievement cluster curves from most cognitive growth curves.

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

Summary

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

References

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice*, 17(1), 10–22.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytical studies*. New York, NY: Cambridge University Press.
- Carroll, J. B., & Maxwell, S. E. (1979). Individual differences in cognitive abilities. *Annual Review of Psychology*, *30*, 603–640.
- Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin*, *38*, 592.
- Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, 40, 153–193.
- Cattell, R. B. (1950). *Personality: A systematic theoretical and factoral study*. New York, NY: McGraw-Hill.
- Cummins, J. (1984). Bilingualism and special education: Issues in assessment and pedagogy. Austin, TX: Pro-Ed.
- Efron, B., & Tibshirani, R. (1993). *An introduction to bootstrap*. New York, NY: Chapman and Hall.
- Horn, J. L. (1965). Fluid and crystallized intelligence. (Unpublished doctoral dissertation). University of Illinois, Urbana-Champaign, IL.
- Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate psychology* (2nd ed., pp. 645–865). New York, NY: Academic Press.
- Horn, J. L. (1989). Models for intelligence. In R. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 29–73). Urbana, IL: University of Illinois Press.
- Horn, J. L., & Stankov, L. (1982). Auditory and visual factors of intelligence. *Intelligence*, 6, 165–185.
- Linacre, J. M. (2012). WINSTEPS (Version 3.74.0) [Computer software]. Chicago, IL: Winsteps.com.
- McArdle, J. J., & Woodcock, R. W. (1998). *Human cognitive abilities in theory and practice*. Mahwah, NJ: Lawrence Erlbaum.
- 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. 151–179). New York, NY: Guilford Press.
- McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 136–181). New York, NY: Guilford Press.

- McGrew, K. S. (2009). Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1–10.
- McGrew, K. S. (2012). *Implications of 20 years of CHC cognitive-achievement research: Back to the future and beyond CHC.* Paper presented at the Richard Woodcock Institute, Tufts University, Medford, MA.
- McGrew, K. S., Dailey, D., & Schrank, F. A. (2007). Woodcock-Johnson III/Woodcock-Johnson III Normative Update score differences: What the user can expect and why (Woodcock-Johnson III Assessment Service Bulletin No. 9). Rolling Meadows, IL: Riverside Publishing.
- McGrew, K. S., & Flanagan, D. P. (1998). The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment. Boston, MA: Allyn & Bacon.
- McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). Technical Manual. *Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside Publishing.
- McGrew, K. S., & Wendling, B. J. (2010). Cattell-Horn-Carroll cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools*, 47, 651–675.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses in performances as scientific inquiry into score meaning. *American Psychologist*, *50*, 741–749.
- Mosier, C. I. (1943). On the reliability of a weighted composite. *Psychometrika*, 8, 161–168.
- Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005a, 2007a). Batería III Woodcock-Muñoz. Rolling Meadows, IL: Riverside Publishing.
- Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005b, 2007b). Batería III Woodcock-Muñoz: Pruebas de habilidades cognitivas. Rolling Meadows, IL: Riverside Publishing.
- Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology*, 85, 112–118.
- Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence.
 In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). New York, NY: Guilford Press.
- 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. 73–164). New York, NY: Guilford Press.
- Schrank, F. A., Mather, N., & McGrew, K. S. (2014a). Woodcock-Johnson IV Tests of Achievement. Rolling Meadows, IL: Riverside Publishing.
- Schrank, F. A., Mather, N., & McGrew, K. S. (2014b). Woodcock-Johnson IV Tests of Oral Language. Rolling Meadows, IL: Riverside Publishing.
- Schrank, F. A., McGrew, K. S., & Mather, N. (2014a). Woodcock-Johnson IV. Rolling Meadows, IL: Riverside Publishing.
- Schrank, F. A., McGrew, K. S., & Mather, N. (2014b). Woodcock-Johnson IV Tests of Cognitive Abilities. Rolling Meadows, IL: Riverside Publishing.

- Wolfe, E. W. (2004). Equating and item banking with the Rasch model. In E. V. Smith, Jr. & R. M. Smith (Eds.), *Introduction to Rasch measurement*. Maple Grove, MN: JAM Press.
- Wolfe, F., MacIntosh, R., Kreiner, S., Lange, R., Graves, R., & Linacre, J. M. (2006). Multiple significance tests. *Rasch Measurement Transactions*, *19*, 1044.
- Woodcock, R. W. (1981). Batería Woodcock psico-educativa en español. Chicago, IL: Riverside Publishing.
- Woodcock, R. W. (1988, August). Factor structure of the tests of cognitive ability from the 1977 and 1989 Woodcock-Johnson. Paper presented at the Australian Council on Educational Research Seminar on Intelligence, Melbourne, Australia.
- Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment*, 8, 231–258.
- Woodcock, R. W. (1993). An information processing view of *Gf-Gc* theory. *Journal of Psychoeducational Assessment*. Monograph Series: Advances in Psychoeducational Assessment: Woodcock-Johnson Psycho-Educational Battery–Revised, 80–102.
- Woodcock, R. W. (1994). Measures of the abilities of *Gf-Gc* theory. In R. Sternberg (Ed.), *Encyclopedia of intelligence* (pp. 452–456). New York, NY: Macmillan.
- Woodcock, R. W. (1998). Extending *Gf-Gc* theory into practice. In J. J. McArdle & R.
 W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 137–156).
 Mahwah, NJ: Lawrence Erlbaum.
- Woodcock, R. W., Alvarado, C. G., Schrank, F. A., Mather, N., McGrew, K. S., & Muñoz-Sandoval, A. F. (2019). Batería IV Woodcock-Muñoz: Pruebas de aprovechamiento. Itasca, IL: Riverside Insights.
- Woodcock, R. W., Alvarado, C. G., Schrank, F. A., McGrew, K. S., Mather, N., & Muñoz-Sandoval, A. F. (2019a). Batería IV Woodcock-Muñoz. Itasca, IL: Riverside Insights.
- Woodcock, R. W., Alvarado, C. G., Schrank, F. A., McGrew, K. S., Mather, N., & Muñoz-Sandoval, A. F. (2019b). Batería IV Woodcock-Muñoz: Pruebas de habilidades cognitivas. Itasca, IL: Riverside Insights.
- Woodcock, R. W., & Johnson, M. B. (1977). Woodcock-Johnson Psycho-Educational Battery. Chicago, IL: Riverside Publishing.
- Woodcock, R. W., & Johnson, M. B. (1989). Woodcock-Johnson Psycho-Educational Battery–Revised. Chicago, IL: Riverside Publishing.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001, 2007). Woodcock-Johnson III. Rolling Meadows, IL: Riverside Publishing.
- Woodcock, R. W., & Muñoz-Sandoval, A. F. (1996). Batería Woodcock-Muñoz–Revisada. Chicago, IL: Riverside Publishing.
- Wright, B. D., & Stone, M. H. (1979). Best test design. Chicago, IL: MESA Press.
- Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 337–347). Hillsdale, NJ: Lawrence Erlbaum.



800.323.9540 www.wj-iv.com