Woodcock Johnson® IV

ECAD™

Assessment Service Bulletin Number 4

Woodcock-Johnson® IV Tests of Early Cognitive and Academic Development: Overview and Technical Abstract

Erica M. LaForte, PhD

Kevin S. McGrew, PhD

Fredrick A. Schrank, PhD, ABPP

This bulletin provides an overview of the Woodcock-Johnson IV Tests of Early Cognitive and Academic Development *(ECAD™; Schrank, McGrew, & Mather, 2015) and highlights important technical aspects of the ECAD that establish its validity for measuring emergent cognitive and expressive language abilities and early academic skills. Information in this bulletin is abstracted from the* Woodcock-Johnson IV Tests of Early Cognitive and Academic Development Comprehensive Manual *(Wendling, Mather, LaForte, McGrew, & Schrank, 2015). Additionally, because the ECAD was developed and normed simultaneously with the* Woodcock-Johnson IV *(WJ IV™; Schrank, McGrew, & Mather, 2014a), much of the technical information relevant to the WJ IV also applies to the ECAD tests. Readers who are interested in more detailed information should consult the* Woodcock-Johnson IV Technical Manual *(McGrew, LaForte, & Schrank, 2014).*

Houghton Mifflin Harcourt

Printed in the United States of America.

Reference Citation

- To cite this document, use:

  LaForte, E. M., McGrew, K.S., & Schrank, F. A. (2015). *Woodcock-Johnson IV Tests of Early Cognitive and Academic Development: Overview and Technical Abstract* (Woodcock-Johnson IV Assessment Service Bulletin No. 4). Rolling Meadows, IL: Riverside.

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

Woodcock-Johnson® IV Tests of Early Cognitive and Academic Development: Overview and Technical Abstract

The Woodcock-Johnson IV Tests of Early Cognitive and Academic Development (ECAD™; Schrank, McGrew, & Mather, 2015) is a special-purpose battery of individually administered tests, contained within a single test easel, developed primarily for use with children ages 2 years 6 months through 7. The ECAD also is appropriate for use with children ages 8 and 9 who have a cognitive developmental delay. The tests were designed to be interesting and attractive to young children and easy to administer by a wide variety of assessment professionals.

The ECAD provides developmentally appropriate measures for identification of emergent cognitive abilities and early academic skills; these measures are particularly appropriate for determining the presence and severity of cognitive delay and for identifying relative strengths and weaknesses among different cognitive abilities that may be relevant to early interventions. The associated WJ IV online scoring and reporting program (Schrank & Dailey, 2014, 2015) quickly calculates and reports all derived scores, including a developmental age profile for displaying ECAD scores and communicating results to teachers and parents. This program includes options for assisting examiners in identifying the presence and severity of a cognitive developmental delay.

Organization of the WJ IV Tests of Early Cognitive and Academic Development

The 10 ECAD tests are contained in one easel-style test book. Table 1 lists the tests included in the ECAD. Special notations following four tests indicate two tests that are administered using an audio recording (🎧), one test that is timed (🕒), and one test that requires the Response Worksheet (📖). Table 1 presents the tests by order of appearance in the ECAD Test Book. The second column in the table indicates the broad cognitive abilities that are measured by each test. Because cognitive abilities emerge and become more differentiated over time, some tests measure aspects of more than one broad ability.

Table 1.
Test Names and Broad Cognitive Abilities Measured in the ECAD

Test Name	Broad Cognitive Abilities Measured
Test 1: Memory for Names	Long-Term Retrieval (*Glr*)
Test 2: Sound Blending	Auditory Processing (*Ga*)
Test 3: Picture Vocabulary	Comprehension-Knowledge (*Gc*)
Test 4: Verbal Analogies	Fluid Reasoning and Comprehension-Knowledge (*Gf* & *Gc*)
Test 5: Visual Closure	Visual Processing (*Gv*)
Test 6: Sentence Repetition	Short-Term Working Memory and Comprehension-Knowledge (*Gwm* & *Gc*)
Test 7: Rapid Picture Naming	Cognitive Processing Speed and Long-Term Retrieval (*Gs* & *Glr*)
Test 8: Letter-Word Identification	Reading (*Grw*)
Test 9: Number Sense	Mathematics (*Gq*)
Test 10: Writing	Written Language (*Grw*)

Table 2 illustrates the three clusters, or groupings of tests, that are available in the ECAD. These clusters are useful for describing a child's cognitive and expressive language performance levels and overall level of early academic skills.

Table 2.
ECAD Clusters and Required Tests

Cluster	Tests Required for Cluster
General Intellectual Ability–Early Development (GIA-EDev)	Test 1: Memory for Names Test 2: Sound Blending Test 3: Picture Vocabulary Test 4: Verbal Analogies Test 5: Visual Closure Test 6: Sentence Repetition Test 7: Rapid Picture Naming
Early Academic Skills	Test 8: Letter-Word Identification Test 9: Number Sense Test 10: Writing
Expressive Language	Test 3: Picture Vocabulary Test 6: Sentence Repetition

Relationship of the ECAD to the *Woodcock-Johnson IV*

The ECAD is part of the *Woodcock-Johnson IV* (WJ IV™; Schrank, McGrew, & Mather, 2014a) family of assessment instruments that also includes the *Woodcock-Johnson IV Tests of Cognitive Abilities* (WJ IV COG; Schrank, McGrew, & Mather, 2014b), the *Woodcock-Johnson IV Tests of Oral Language* (WJ IV OL; Schrank, Mather, & McGrew, 2014b), and the *Woodcock-Johnson IV Tests of Achievement* (WJ IV ACH; Schrank, Mather, & McGrew, 2014a).

Among the 10 ECAD tests, 4 are unique to the ECAD (Test 1: Memory for Names, Test 4: Verbal Analogies, Test 5: Visual Closure, and Test 9: Number Sense). Four of the tests are alternate forms of tests included the WJ IV OL (Test 2: Sound Blending, Test 3: Picture Vocabulary, Test 6: Sentence Repetition, and Test 7: Rapid Picture Naming), and two tests are alternate forms of tests included in the WJ IV ACH (Test 8: Letter-Word Identification and Test 10: Writing, which is called "Spelling" in the WJ IV ACH).

The ECAD was co-normed with the WJ IV COG, WJ IV OL, and WJ IV ACH. All of the WJ IV batteries share a common interpretive model and utilize the same *W*-score metric, which is useful in situations where a child is tested at the preschool level with the ECAD and later in the school years with the WJ IV COG, WJ IV OL, and/or WJ ACH. Finally, because norming data for the ECAD and WJ IV tests were collected from the same sample of examinees, the test authors were able to include the ECAD tests in the comprehensive validity analyses that were conducted during the WJ IV development.

Relationship of the ECAD to the CHC Theory of Cognitive Abilities

The Cattell-Horn-Carroll (CHC) theory of cognitive abilities (McGrew, 2005, 2009; Schneider & McGrew, 2012; McGrew et al., 2014) provides the theoretical basis for the ECAD. CHC theory is the most widely accepted and empirically validated theory of cognitive abilities (Alfonso, Flanagan, & Radwan, 2005; Schneider & Newman, 2015). CHC theory posits that cognitive ability is not unidimensional but instead is composed of multiple broad and narrow cognitive abilities (Carroll, 1993).

Two of the broad CHC abilities, Fluid Reasoning (*Gf*) and Comprehension-Knowledge (*Gc*), can be traced to Cattell (1941, 1943, 1950), who called these abilities 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). Quantitative Knowledge (*Gq*) was identified by Horn (1988), and Reading-Writing ability (*Grw*) was identified by Carroll and Maxwell (1979) and later by Woodcock (1998). The CHC abilities and initialisms have subsequently been refined and updated by Woodcock (McArdle & Woodcock, 1998; Woodcock, 1990, 1993, 1994, 1998) and McGrew (1997, 2005, 2009) and most recently by Schneider and McGrew (2012) and McGrew, LaForte, and Schrank (2014). Additional information on CHC theory can be found in the *Woodcock-Johnson IV Tests of Early Cognitive and Academic Development Comprehensive Manual* (Wendling, Mather, LaForte, McGrew, & Schrank, 2015), the WJ IV COG Examiner's Manual (Mather & Wendling, 2014), and the *Woodcock-Johnson IV Technical Manual* (McGrew, LaForte, & Schrank, 2014).

In the ECAD, the use of CHC theory as the interpretive model is particularly helpful for identifying a delay in one or more aspects of cognitive functioning because emergent cognitive abilities may be malleable and subject to improvement through intervention. Identification of delays in any of the cognitive abilities posited by CHC theory and measured by the ECAD may be the first step in developing an individualized program such as an Individual Family Service Plan (IFSP) or an Individualized Education Program (IEP), or in recommending accommodations or curricular adjustments for a child. Early identification and intervention are keys to improved educational outcomes.

Use of the ECAD for Determination of Developmental Delay

The ECAD is particularly appropriate for determining the presence and severity of a developmental delay in cognitive ability and for evaluating a child's preacademic skills. The various ECAD scoring options allow practitioners to evaluate children's cognitive abilities and preacademic skills using both criterion-referenced and norm-referenced interpretations, allowing flexibility to meet most local jurisdictional requirements for assessment of developmental delay. This section describes the criterion-referenced proficiency information that is included in the ECAD interpretive model, as well as three norm-referenced scores that sometimes are used for determining developmental delay: Months Delay, Percentage Delay, and *SD* Delay.

ECAD Levels of Cognitive and Academic Development

The ECAD interpretive model includes a set of criterion-referenced labels describing a child's cognitive and academic skill development levels that are based directly on proficiency with the measured tasks. The Rasch-based (Rasch, 1960/1980; Wright & Stone, 1979) *W* scale (Woodcock & Dahl, 1971) underlying the ECAD provides the basis for these proficiency labels. When items and examinees are calibrated with the Rasch model, the resulting item difficulty estimates and examinee ability estimates are placed onto the same scale; the same set of numbers can then be used to describe an examinee's ability and the difficulty of the task represented by any item on the test. Additionally, the Rasch-based *W* scale has implications for an examinee's probability of success on test items based on those items' relative distances to the examinee's ability along the scale. Because the *W* scale is an equal-interval scale, the same distance between two points has the same interpretation at any ability level measured by the test. If an examinee's ability estimate (W_A) is higher than the difficulty (W_D) of an administered item (i.e., when W_{A-D} is positive), the examinee will have a greater than 50% chance of success on the item. Likewise, if an examinee's ability estimate is lower than the difficulty of an administered item (i.e., when W_{A-D} is negative), the examinee will have a less than 50% chance of success on the item. Table 3 describes the probability of examinee success for several key values of W_{A-D}.

Table 3.
Probability of Success for Several Key Values of W_{A-D}

Examinee Ability Minus Task Difficulty (W_{A-D})	Probability of Success (P_{A-D})	Examinee Ability Minus Task Difficulty (W_{A-D})	Probability of Success (P_{A-D})
+50	.996	0	.500
+45	.993	–5	.366
+40	.988	–10	.250
+35	.979	–15	.161
+30	.964	–20	.100
+25	.940	–25	.060
+20	.900	–30	.036
+15	.839	–35	.021
+10	.750	–40	.012
+5	.634	–45	.007
0	.500	–50	.004

The probability relationships that exist for W_{A-D} also apply to other situations in which a difference score can be determined. For example, an examinee's ability measure (W_A) can be compared to the average ability of the examinee's same-age peers from the norming sample (REF *W*) to obtain the *W* difference score (*W* DIFF). The *W* DIFF allows meaningful applications for interpreting a child's proficiency with cognitive and academic tasks (Jaffe, 2009). The Relative Proficiency Index (RPI) is one application of the *W* DIFF. The RPI is represented as a fraction, with the child's expected level of success as the numerator and the 90% criterion as the denominator. For example, an RPI of 60/90 on Test 2: Sound Blending suggests that the child being evaluated would be about 60% successful on sound blending tasks that typical children of the same age would perform with 90% success.

Table 4 contains several *W* DIFF value ranges and their corresponding RPI ranges, criterion-referenced developmental levels, and instructional implications. Two important blended-category developmental levels (*Age-Appropriate to Advanced* and *Mildly Delayed to Age-Appropriate*) draw attention to task performance that is at or near a critical change in interpretation. In the Test 2: Sound Blending example provided earlier, this child's sound blending ability is mildly delayed; age-level sound blending tasks would be very difficult for him or her. Systematically applied phonological awareness intervention, including targeted practice in blending sounds into words, may reduce or even eliminate the mild developmental delay identified by the Sound Blending test.

Table 4.
ECAD W *Difference Scores and Corresponding RPIs, Developmental Levels, and Instructional Implications*

W Difference Range	Reported RPI	Developmental Level	Instructional Implication (i.e., age-level tasks will be…)
+31 and above	100/90	Very Advanced	Extremely easy
+14 to +30	98/90 to 100/90	Advanced	Very easy
+7 to +13	95/90 to 98/90	Age-Appropriate to Advanced	Easy
–6 to +6	82/90 to 95/90	Age-Appropriate	Manageable
–13 to –7	67/90 to 82/90	Mildly Delayed to Age-Appropriate	Difficult
–30 to –14	24/90 to 67/90	Mildly Delayed	Very difficult
–50 to –31	3/90 to 24/90	Moderately Delayed	Extremely difficult
–51 and below	0/90 to 3/90	Extremely Delayed	Nearly impossible

The ECAD's recommended interpretation of developmental levels can be compared and contrasted with three other methods that sometimes are used for determining developmental delay: Months Delay, Percentage Delay, and *SD* Delay. Any of these three scores can be derived from the ECAD test results and can be selected for inclusion in a score report in the Woodcock-Johnson online scoring and reporting program (Schrank & Dailey, 2014, 2015). However, as described below, each of these scores has limitations when compared with the ECAD developmental levels.

Months Delay

Some agencies or jurisdictions allow a determination of delay to be calculated based on the actual number of months of delay a child exhibits in one or more developmental areas, entailing the use of age equivalent (AE) scores. The age equivalent is the age in the norming sample at which the median score of children in the sample is the same as the examinee's score. If a child's AE score is lower than his or her actual chronological age, then the months delay is calculated as the AE score (in months) minus the child's chronological age (in months). For example, a 47-month-old child who receives an AE of 39 months on Test 2: Sound Blending would receive a Months Delay score of 8 (39 months minus 47 months = −8 months). It is important to recognize that the interpretation of Months Delay scores is not comparable at all points along the scale. At ages where rapid growth occurs in the underlying ability or skill, an 8-month delay will represent a much larger change in the actual underlying ability than at ages where growth occurs more slowly. Users should evaluate Months Delay scores along with other types of scores, such as the ECAD criterion-referenced developmental levels or norm-referenced percentile ranks.

Percentage Delay

Percentage delay is the number of months delay expressed as a percentage of the child's chronological age. Some agencies or jurisdictions require a determination of delay to be made based on percentage delay in one or more developmental areas. If the child's AE score is lower than his or her actual chronological age, then the percentage delay is calculated as:

$$\frac{Age\ (in\ months) - AE\ score\ (in\ months)}{Age\ (in\ months)} \times 100.$$

In the example above, the 47-month-old child who received an AE score of 39 months on Test 2: Sound Blending would have a 17% delay when compared to the median score of his or her same-aged peers in the norming sample. Because percentage delay is not comparable across the age span (e.g., a 20% delay at age 36 months represents about a 7-month delay, while a 20% delay at age 5 represents a 12-month delay), it is important that examiners interpret Percentage Delay scores along with other types of scores, such as the ECAD criterion-referenced developmental levels or ECAD norm-referenced percentile ranks.

SD Delay

Specific levels of standard deviations (*SD*s) of delay (e.g., 1.5 *SD*s, 2.0 *SD*s) are required by some agencies or jurisdictions for determination of delay in one or more developmental areas. *SD* delays are based on z scores, with negative z scores indicating scores that fall below the average score of same-age children in the norming sample. For example, a child who receives a z score of −1.3 has a score that is 1.3 *SD*s below the mean score of his or her same-age peers in the norming sample. *SD*s vary greatly across the age span for any given ability; at ages where there is more variability in examinee ability (such as during the early school years), *SD*s will be larger. For this reason, an *SD* Delay score at one age for a given ability can represent more or less delay in terms of actual development than the same *SD* Delay score at another age for the same ability. For example, a 42-month-old child with a delay of 1 *SD* on Test 8: Letter-Word Identification has a *W* score that is 25 points below the median *W* score for children his or her age in the norming sample. However, a 66-month-old child with a delay of 1 *SD* on the same test has a *W* score that is 32 points below the median *W* score for children of the same age. As with the Months Delay and Percentage Delay scores, users should evaluate *SD* Delay scores in the context of other, criterion-referenced scores, such as the ECAD developmental levels, to better understand the child's current functioning.

Norming Study

The ECAD tests were included in the battery of tests that were administered to 7,416 individuals during the norming study for the WJ IV (Schrank, McGrew, & Mather, 2014a). The WJ IV/ECAD norming sample included individuals ages 2 through 90+. The norm tables for all WJ IV and ECAD tests were constructed from this sample of examinees. The subset of the norms for ages 2 years 6 months through 9 are applicable for the ECAD tests. The normative information for ages 2 years 0 months through 2 years 5 months and for age 10 is also described in this section to help ECAD users evaluate the technical characteristics of the test at the extreme ends of the recommended age range. The norm development processes relevant to the ECAD are summarized below. More in-depth information about the WJ IV/ECAD norming study can be found in the WJ IV Technical Manual (McGrew et al., 2014).

The WJ IV/ECAD norming study was conducted between December 2009 and January 2012. During this 25-month period, data were collected from individuals from geographically diverse communities representing 46 U.S. states and the District of Columbia. Of the 7,416 examinees, 2,378 were between the ages of 2 and 10. Table 5 displays the distribution of this subset of the norming sample by age and grade.

Table 5.
Distribution of the WJ IV/ECAD 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	318
7	310	5	138
8	336	6	1
9	306		
10	314		
Total	2,378	**Total**	1,714

The norming sample was selected to be representative, within practical limits, of the U.S. population. 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/ECAD norming sample, for the preschool-age sample. The ECAD Comprehensive Manual (Wendling et al., 2015) provides similar information for the portion of the ECAD norming sample in grades Kindergarten and above. The demographic characteristics of the sample were carefully tracked during the norming study to ensure that the distributions of each demographic characteristic in the sample matched those in the U.S. population as closely as possible.

Table 6.
Distribution of Sampling Variables in the U.S. Population and in the WJ IV/ECAD Norming Sample–Preschool

Sampling Variable	Percentage in U.S. Population	Number Obtained	Percentage in Norm Sample	Partial Examinee Weight
Census Region				
Northeast	15.6	113	17.0	0.914
Midwest	21.5	167	25.2	0.854
South	37.2	219	33.0	1.127
West	25.8	165	24.8	1.038
Community Type				
Metropolitan	83.7	551	83.0	1.008
Micropolitan	10.0	86	13.0	0.774
Rural	6.3	27	4.1	1.552
Sex				
Male	51.1	332	50.0	1.022
Female	48.9	332	50.0	0.970
Country of Birth				
United States	98.7	661	99.5	0.992
Other	1.3	3	0.5	2.809
Race/Ethnicity				
White, Not Hispanic	63.7	427	64.3	0.975
Black, Not Hispanic	12.5	96	14.5	0.852
AIANAT, Not Hispanic	—[a]	—	—	—
ASIPAC, Not Hispanic	5.2	23	3.5	1.483
Other, Not Hispanic	—[a]	2	0.3	1.000[b]
White, Hispanic	16.6	102	15.4	1.067
Black, Hispanic	0.7	4	0.6	1.138
AIANAT, Hispanic	0.3	1	0.2	1.894
ASIPAC, Hispanic	0.2	1	0.2	1.131
Other, Hispanic	—[a]	8	1.2	1.000[b]
Parent Education				
< High School	13.9	94	14.2	0.984
High School	23.6	181	27.3	0.865
> High School	62.5	389	58.6	1.067

[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.

Gathering accurate data for this many tests on a large nationally representative sample presented a number of logistical design constraints. A balance was needed between the competing goals of reasonable testing time for each norming study participant (to minimize examinee response burden) and completeness of data for all tests (National Research Council, 2013; Thomas, Raghunathan, Schenker, Katzoff, & Johnson, 2006). To overcome these challenges, a planned incomplete data collection design was used in the norming study.

The WJ IV/ECAD norming study was based on a *multiple matrix sampling* (MMS) design, in which subsets of a larger complete set of test items or tests are administered to different random subsamples of the total norming sample. Planned incomplete (missing) data collection methods (Graham, Taylor, Olchowski, & Cumsille, 2006; McArdle, 1994; McKnight, McKnight, Sidani, & Figueredo, 2007; Rhemtulla & Little, 2012; Schafer, 1997; Wolf, 2006) have been developed as a statistically sound technique for gathering data in large studies where design constraints preclude complete data collection. WJ IV/ECAD norming study participants were each randomly assigned to one block of tests, each of which contained between 15 and 19 tests. Best practice approaches to generating plausible *W* scores for tests not taken by norming study participants were then utilized to generate a "complete record" for all norming study participants. Details regarding the block design criteria, study constraints, and data imputation methods can be found in the WJ IV Technical Manual (McGrew et al., 2014).

Individual examinee weights were applied during the norms construction process to ensure that the norms were based on a sample with characteristics proportional to the U.S. population distribution. The population and sample percentages in Table 6 show a close match for most sampling variables; however, to ensure that the ECAD norms are representative of the exact demographic distributions in the U.S. population, examinee weighting was used. 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. For each demographic variable, if an examinee belonged to a category of the variable that was overrepresented in the WJ IV/ECAD 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 Preschool sample of examinees.

Calculation of ECAD Cluster Scores

The ECAD Expressive Language and Early Academic Skills clusters are based on the arithmetic average of the *W* scores of the tests that contribute to the cluster score. The ECAD General Intellectual Ability–Early Development (GIA-EDev) cluster score is a differentially weighted score. Principal component analysis was used to obtain differential *g* weights across the age range of the ECAD for the tests that contribute to the GIA-EDev score. The differential weights are used by the Woodcock-Johnson online scoring and reporting program (Schrank & Dailey, 2014, 2015) to calculate GIA-EDev scores for individual examinees. Table 7 presents the average GIA-EDev weights by the ECAD ages. A review of the weights in Table 7 reveals that the weights for the individual tests fluctuate little as a function of age.

Table 7.
General Intellectual Ability–Early Development (GIA-EDev) Average (Smoothed) g *Weights by Age*

Test	CHC Domain	AGE								Median
		2	3	4	5	6	7	8	9	
Test 1: Memory for Names	*Glr*	0.12	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
Test 2: Sound Blending	*Ga*	0.09	0.10	0.12	0.13	0.13	0.13	0.11	0.10	0.11
Test 3: Picture Vocabulary	*Gc*	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.18	0.17
Test 4: Verbal Analogies	*Gf*	0.21	0.21	0.21	0.21	0.21	0.21	0.22	0.22	0.21
Test 5: Visual Closure	*Gv*	0.15	0.15	0.14	0.14	0.13	0.13	0.13	0.13	0.13
Test 6: Sentence Repetition	*Gwm*	0.13	0.13	0.13	0.13	0.14	0.14	0.14	0.14	0.14
Test 7: Rapid Picture Naming	*Gs*	0.12	0.12	0.12	0.12	0.12	0.11	0.11	0.12	0.12

Calculation of Norms

As described in the *Woodcock-Johnson IV Technical Manual* (McGrew et al., 2014), 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 where normative interpretations are intended. In the WJ family of instruments, this normative score is called the Reference *W* score (REF W). When plotted as a function of chronological age, the REF *W*s assume the characteristics of developmental growth curves. These test and cluster REF *W* curves are visual-graphic representations of the average performance of norming study participants at every age for the effective use of the specific measure.

The REF *W* curves serve as the foundation for the AE scores, RPIs, and developmental level features in the ECAD. In addition, when the *SD*s of the scores are plotted as a function of age, the resultant curves represent the *SD* values that, when combined with the REF *W* values, provide the foundation for the calculation of all other norm-referenced score metrics (e.g., standard scores and percentile ranks).

Bootstrap resampling procedures (Efron & Tibshirani, 1993; McGrew, Dailey, & Schrank, 2007[1]) were used to calculate the WJ IV and ECAD 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 with more traditional norm development procedures (such as those used in the *Woodcock-Johnson Psycho-Educational Battery* (WJ; Woodcock & Johnson, 1977), *Woodcock-Johnson Psycho-Educational Battery–Revised* (WJ-R®; Woodcock & Johnson, 1989), *Woodcock-Johnson III* (WJ III®; Woodcock, McGrew, & Mather, 2001), and most other individually administered cognitive, language, and achievement batteries), the bootstrap-based procedures used to calculate the WJ IV and ECAD 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 ECAD and WJ IV batteries, which can then 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 follow:

1. To determine whether an examinee's relative standing in a group on an individual test or cluster (e.g., ECAD Test 3: Picture Vocabulary) is statistically significantly different from the examinee's relative standing in the same group on another individual test or cluster (e.g., ECAD Test 2: Sound Blending).
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.

[1] A copy of this document can be obtained at the following URL: http://www.riversidepublishing.com/products/wjIIIComplete/pdf/WJIII_ASB9.pdf

Actual differences between predictor and criterion variables for each individual in the norming sample can be used to model these differences in the population. In the ECAD, this type of difference score takes two forms: *variations* and *comparisons*. The ECAD variation and comparison procedures are based on a common statistical model.[2] What distinguishes variations and comparisons from each other is the score that is used as the predictor in the model. While the intra-cognitive and early intra-achievement variations rely on a predictor score that is an average of the (noncriterion) scores from a pool of tests that excludes the criterion measure, the ability/achievement comparison relies on a single predictor, the GIA-EDev score.

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

Reliability

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

ECAD Test Reliabilities

Reliability statistics were calculated for all ECAD tests across the age ranges of intended use. Reliability calculations for tests included data from all norming examinees tested at each technical age level. Internal-consistency reliabilities for all tests except Test 7: Rapid Picture Naming were calculated using the split-half procedure. Raw scores were computed for the norming examinees based on the odd and even items in these tests. Correlations were then computed between the two item sets. The basal-ceiling rules used during the norming study were stringent enough that the probability of an examinee failing an item below the basal or passing an item above the ceiling was very low. Therefore, all item responses below an examinee's basal level were assumed to be correct, and all responses above the examinee's ceiling level were assumed to be incorrect. These coefficients were then corrected for published test length using the Spearman-Brown correction formula. The reliabilities for Test 7: Rapid Picture Naming were calculated using information provided by the Rasch model.

[2] Information about the construction of the difference score norms for the ECAD and the WJ IV can be found in Chapter 6 of the ECAD Comprehensive Manual (Wendling et al., 2015) and in Chapter 3 of the WJ IV Technical Manual (McGrew et al., 2014), respectively.

Table 8 reports the median test reliability coefficients (r_{11}) and the standard errors of measurement in standard score units (*SEM* SS) for ages 2 through 7 for all ECAD tests, obtained using the procedures described above. Of the 10 median test reliability coefficients reported in Table 8, all are .79 or higher, and nine are .80 or higher. Although these are strong reliabilities for individual tests, the ECAD cluster scores are recommended for making important decisions about an individual due to the higher reliabilities of those scores.

Table 8.
ECAD Median Test and Cluster Reliability Coefficients and Standard Errors of Measurement in Standard Score Units

Test or Cluster	Median r_{11}	Median *SEM* (SS)
Tests		
Test 1: Memory for Names	0.97	2.60
Test 2: Sound Blending	0.84	6.00
Test 3: Picture Vocabulary	0.84	5.66
Test 4: Verbal Analogies	0.82	6.45
Test 5: Visual Closure	0.81	6.60
Test 6: Sentence Repetition	0.91	4.48
Test 7: Rapid Picture Naming	0.89	5.03
Test 8: Letter-Word Identification	0.96	3.00
Test 9: Number Sense	0.79	6.87
Test 10: Writing	0.91	4.50
Clusters		
General Intellectual Ability–Early Development	0.96	2.98
Early Academic Skills	0.96	3.00
Expressive Language	0.93	3.99

ECAD Cluster Reliabilities

The reliability coefficients for the ECAD clusters were computed using Mosier's (1943) formula. Table 8 reports the median reliability coefficients and *SEM* SSs for the three ECAD clusters for ages 2 through 7. All three median cluster reliabilities are .93 or higher.

Validity Evidence for the ECAD

This section contains several types of evidence to support the proposition that the ECAD scores are valid for describing a child's cognitive abilities, expressive language ability, and early academic skills. The evidence is presented in a framework consistent with that outlined in the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association, & National Council on Measurement in Education, 2014).

Representativeness of the ECAD Test Content, Process, and Construct Coverage

The ECAD provides a set of developmentally appropriate tests for identification of emergent cognitive and expressive language abilities and early academic skills. Evidence to support this proposition, often termed *content validity* or *substantive validity* evidence, for the ECAD test scores is provided via the specification of test and cluster content according to contemporary CHC research and theory.[3] This aspect of the ECAD validity argument builds upon the theories contained in the four editions of the Woodcock-Johnson tests. The WJ-R, WJ III, and WJ IV were based on successive revisions to the Cattell-Horn Extended *Gf-Gc* and Cattell-Horn-Carroll (CHC) theories of cognitive abilities (McGrew, 2005, 2009; Schneider & McGrew, 2012; McGrew et al., 2014).

CHC Theory Content Coverage

The distinction between broad and narrow abilities is an important concept in CHC theory. Each ECAD test was chosen to represent a broad CHC factor (see Table 1). This CHC-based test design approach, operationalized in the WJ III and WJ IV, focuses on increasing CHC construct representation and decreasing construct-irrelevant variance in tests (Benson, 1998; McGrew & Flanagan, 1998; Messick, 1995). To increase breadth, the three ECAD clusters were constructed to subsume two or more qualitatively different narrow abilities. The principle of cluster interpretation improves the content validity of measures for broad abilities such as general intelligence, expressive language, and academic skills.

Construct, Process, and Content Coverage

Table 9 provides further descriptions of the broad and narrow constructs measured by the ECAD, as well as stimulus and response characteristics, task requirements, and inferred cognitive processes.

Table 9.
ECAD Test Content, Process, and Construct Descriptions

Test	**Primary Broad CHC Ability** *Narrow Ability*	Stimuli	Task Requirements	Cognitive Processes	Response
Test 1: Memory for Names	Long-Term Retrieval (*Glr*) *Associative memory*	Auditory-visual (names, pictures)	Learning and recalling names	Associative encoding via directed spotlight attention, storage, and retrieval	Motoric (pointing)
Test 2: Sound Blending	Auditory Processing (*Ga*) *Phonetic coding* (PC)	Auditory (phonemes)	Synthesizing language sounds (phonemes) to say a word	Synthesis of acoustic, phonological elements in immediate awareness; matching the sequence of elements to stored lexical entries; lexical activation and access	Oral (words)
Test 3: Picture Vocabulary	Comprehension-Knowledge (*Gc*) *Lexical knowledge* (VL) *Language development* (LD)	Visual (pictures)	Identifying objects	Object recognition; lexical access and retrieval	Oral (words)

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

Table 9. *(cont.)*
ECAD Test Content, Process, and Construct Descriptions

Test	**Primary Broad CHC Ability** *Narrow Ability*	Stimuli	Task Requirements	Cognitive Processes	Response
Test 4: Verbal Analogies	Fluid Reasoning (*Gf*) *General sequential reasoning* (RG) *Induction* (I) Comprehension-Knowledge (*Gc*) *Language development* (LD) *General (verbal) information* (KO)	Auditory	Deducing a relationship between the two words in the first part of an analogy; applying the inference to a third word to complete the analogy with a fourth word	Deduction of the structure for the first part of the analogy and then mapping (or projecting) that structure onto the second part of the analogy	Oral (word)
Test 5: Visual Closure	Visual-Spatial Thinking (*Gv*) *Closure speed*	Visual (pictures)	Identifying an object from an incomplete or masked visual representation	Object identification from a limited set of component geons	Oral (word)
Test 6: Sentence Repetition	Short-Term Working Memory (*Gwm*) *Memory span* (MS) Comprehension-Knowledge (*Gc*) *Listening ability* (LS)	Auditory (words, sentences)	Listening to and repeating words, phrases, or sentences in the correct sequence	Formation of echoic memories aided by a semantic, meaning-based code	Oral (words, sentences)
Test 7: Rapid Picture Naming	Cognitive Processing Speed (*Gs*) Long-Term Retrieval (*Glr*) *Naming facility* (NA) *Speed of lexical access* (LA)	Visual (pictures)	Recognizing objects, then retrieving and articulating their names rapidly	Speed/fluency of retrieval and oral production of recognized objects; speeded serial naming; rapid object recognition	Oral (words)
Test 8: Letter-Word Identification	Reading & Writing Ability (*Grw*) *Reading decoding* (RD)	Visual (text)	Identifying printed letters and words	Feature detection and analysis (for letters) and recognition of visual word forms from a phonological lexicon; access of pronunciations associated with visual word forms	Oral (letter names, words)
Test 9: Number Sense	Quantitative Knowledge (*Gq*) *Mathematics achievement* (A3) Fluid Reasoning (*Gf*) *Quantitative reasoning* (RQ)	Visual (pictures, numbers) Auditory (words, sentences)	Mental arithmetic	Number recognition; spatial and size orientation; counting; number line estimation; number sequencing; magnitude representation; inductive reasoning	Oral (words, numbers)
Test 10: Writing	Reading & Writing Ability (*Grw*) *Spelling ability* (SG)	Auditory (words)	Spelling orally presented words	Access to and application of knowledge of orthography or 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)

Developmental Patterns of ECAD Ability Clusters

ECAD tests and clusters display average score changes consistent with developmental growth of cognitive and achievement abilities across the age span. Divergent growth curves provide evidence for the existence of distinct, unique abilities (Carroll, 1993). Figures 1 and 2 present examples of growth curves, or "difference curves," for the ECAD cognitive tests and clusters and ECAD achievement and expressive language tests and clusters, respectively. Age 10 is included in the figures to show the developmental trajectories of these abilities in the norming sample outside the age range of the ECAD. The growth curves illustrate that the unique abilities measured by the ECAD tests follow different developmental courses or trajectories over the childhood age span. The examples were constructed using age 2 years 0 months (2-0) as a starting point and subtracting the norm-based Reference W score (REF W) for age 2-0 for each test from all other REF Ws for that test to age 10. This procedure produced growth curves all having an assigned common origin of zero.

Figure 1.
Plot of ECAD cognitive test and GIA-EDev cluster W-score difference curves by age.

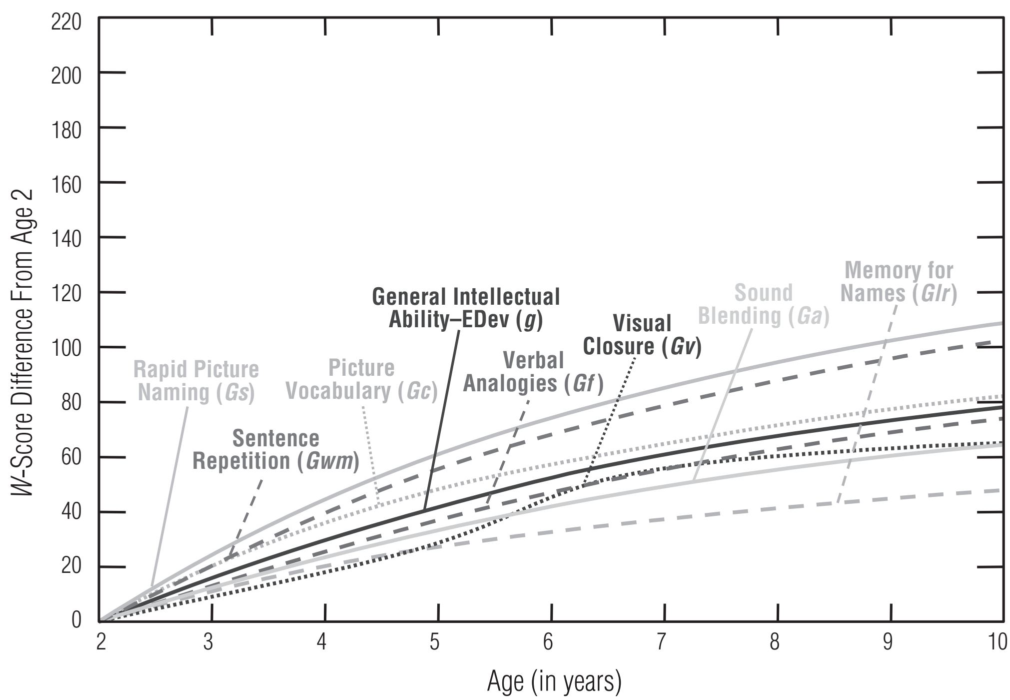
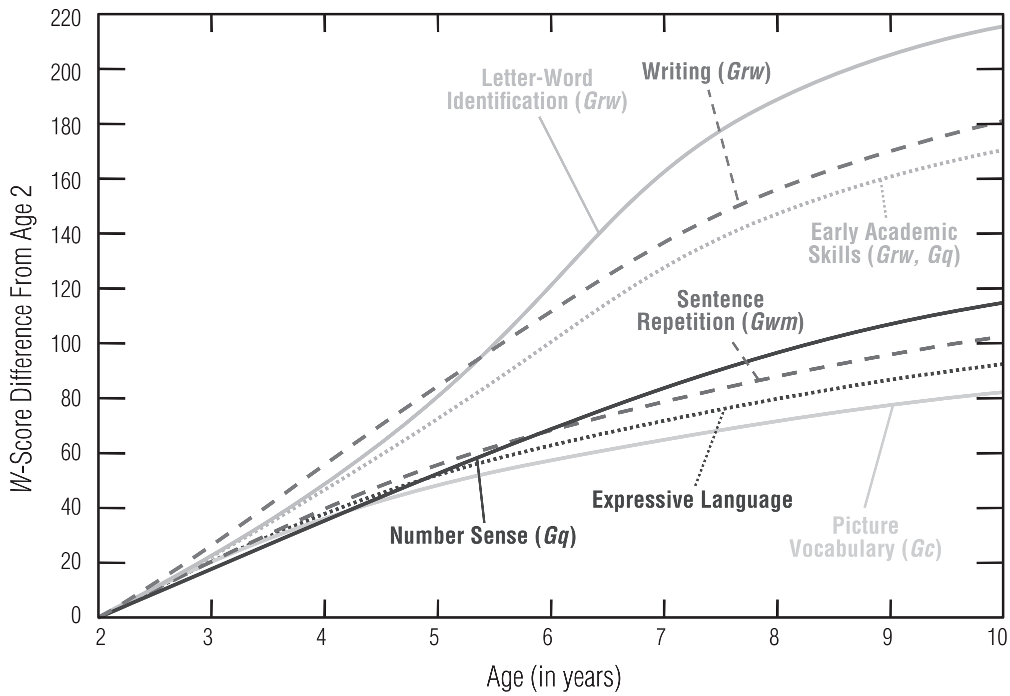


Figure 2.
Plot of ECAD achievement and expressive language tests and Early Academic Skills and Expressive Language cluster W*-score difference curves by age.*



A number of conclusions are apparent when the plots in both Figures 1 and 2 are evaluated together. First, all the cognitive, expressive language, and achievement abilities measured by the ECAD tests and clusters demonstrate steady asymptotic growth throughout the entire age range of the ECAD. Second, as expected given the inclusion of one test from each of the seven CHC cognitive domains, the GIA-EDev (g) cluster growth curve falls approximately in the middle of the other cognitive test curves presented in Figure 1. The Early Academic Skills cluster in Figure 2 also is approximately in the middle of its three component tests (Letter-Word Identification, Writing, and Number Sense). The same conclusion is apparent for the Expressive Language cluster in Figure 2, which consists of the Sentence Repetition and Picture Vocabulary tests. Third, the three ECAD achievement tests and one achievement (Early Academic Skills) cluster show more rapid acceleration of growth from ages 2 through 10 (Figure 2) than do the ECAD cognitive measures (Figure 1), particularly the two measures of Grw (Letter-Word Identification and Writing). The ECAD cognitive measures (Figure 1) show approximately 50 to 110 *W* points of growth over the 2- to 10-year age range, whereas the ECAD achievement measures (Figure 2) show approximately 120 to 220 *W* points of growth. Fourth, in the cognitive domain (Figure 1), the Rapid Picture Naming and Sentence Repetition tests display the most rapid rate of growth, and the Memory for Names test displays the slowest rate of growth. The remaining cognitive tests (Picture Vocabulary, Verbal Analogies, Visual Closure, and Sound Blending) are generally similar in their rate of growth from ages 2 to 10. Finally, within the achievement domain (Figure 2), the three tests show noticeably different rates of growth. Letter-Word Identification shows a very rapid rate of growth, followed by the Writing test. Number Sense shows a more moderate rate of growth and is differentiated from Letter-Word Identification and Writing.

The existence of unique developmental patterns for most of the ECAD tests and clusters, within and across 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 ECAD scores for measuring children's cognitive and expressive language abilities and early academic skills.

Internal Structure of the ECAD

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

ECAD Norming Sample Test and Cluster Intercorrelations

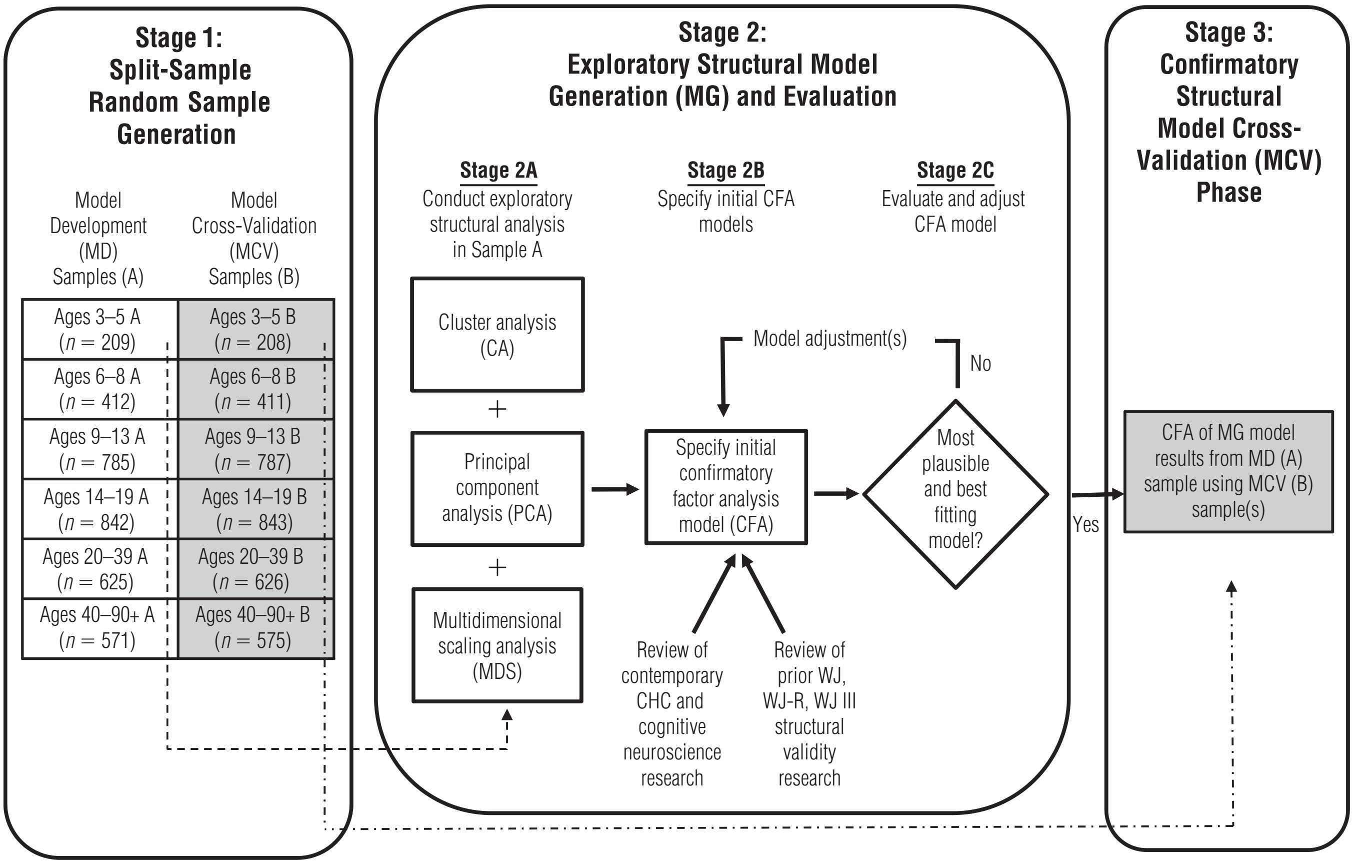
The direction and magnitude of correlations among test and cluster scores can provide evidence that the scores conform to theoretical expectations about the underlying constructs (AERA et al., 2014; Campbell & Fiske, 1959). The test and cluster intercorrelations for the ECAD provide empirical support for several inferences about the relations between ECAD test scores.[4] The correlations are generally lower among the CHC domain cognitive tests (ranging from .09 to .61) than among the achievement tests (ranging from .54 to .85), providing evidence that the ECAD cognitive tests measure cognitive abilities that are distinct from each other and from other measures of academic achievement. The achievement test correlations with the Early Academic Skills cluster are generally high (.72 to .95), lending support for the Early Academic Skills cluster score as a strong measure of early achievement. The correlations between the GIA-EDev and Early Academic Skills clusters (.66 to .67) are similar to correlations commonly found between cognitive and achievement measures. These values also indicate that the global ECAD cognitive and achievement clusters, although significantly correlated, do measure different constructs (i.e., .66 to .67 correlations represent approximately 44% shared score variance).

Three-Stage Structural Validity Analysis

Co-norming the ECAD with the WJ IV allowed the ECAD tests to be examined in the context of the larger structural validity analysis completed for the WJ IV. A systematic exploratory, model generation and a cross-validation structural validity strategy were applied to the WJ IV and ECAD norming data. This split-sample, multiple-stage, exploratory-confirmatory approach is the most thorough scientific approach to the examination of the structural validity of any contemporary battery of cognitive, oral language, and achievement tests. This three-stage process is portrayed in Figure 3. A summary of the process is provided here, with emphasis placed on the interpretation of the results for the ECAD tests; however, readers are encouraged to consult Chapter 5 of the WJ IV Technical Manual (McGrew et al, 2014) for further details.

[4] Complete correlation matrices for all tests and clusters are reported in Chapter 6 of the ECAD Comprehensive Manual (Wendling et al., 2015) for the relevant ECAD age group samples.

Figure 3.
Three-stage structural validity procedures for the WJ IV and ECAD.



As illustrated in Figure 3, the norming sample was divided into six age-differentiated groups. Each sample was randomly split into separate model development (MD; sample A) and model cross-validation (MCV; sample B) samples of approximately equal size (see Stage 1 in Figure 3). Each of the six MD samples was analyzed with three different exploratory multivariate methods—cluster analysis (CA), exploratory principal components analysis (PCA), and multidimensional scaling (MDS) analysis (see Stage 2A in Figure 3). The use of three methodological lenses allows for the detailed exploration of the relations among the complete collection of WJ IV tests. The next step was the specification of the initial model-generating (MG) confirmatory factor analysis (CFA) models based on the integration of the CA, PCA, and MDS results from Stage 2A. A comprehensive review of contemporary CHC and neuroscience research and structural research on all three prior editions of the Woodcock-Johnson tests was integrated with the exploratory results from Stage 2A to specify the initial WJ IV MG CFA models (see Stage 2B in Figure 3).

Two models were found to be most plausible. The *broad CHC factor top-down* model was specified to best represent the broad CHC constructs outlined in contemporary CHC theory. The *broad plus narrow CHC factor bottom-up* model focused on specifying and evaluating plausible narrow and broad CHC factors. In both models, all model parameters for the exemplar age group MD sample were positive, significant ($p < .05$), and meaningful. In Stage 3 of Figure 3, the two models were taken "as is" and cross-validated with the exemplar age group MCV sample. The WJ IV CFA models were evaluated for overall statistical model fit and for size, statistical significance, and interpretability of all model parameter estimates (Brown, 2006). The *broad CHC factor top-down* model is the preferred model per the *parsimony principle* (also known as Occam's razor), which states that "given two models with similar fit to the data, the simpler model is preferred" (Kline, 2011, p. 102). Users should refer to the ECAD Comprehensive Manual (Wendling et al., 2015) for complete results of the evaluation of both CFA models.

Relationship of ECAD Scores to Other Measures

A variety of studies were conducted that examined the relations between ECAD scores and a number of external criterion variables. The types of external validity evidence reported include correlations of the ECAD tests with other measures of cognitive ability, oral language, academic achievement, and early development.

Correlations for the ECAD With Other Measures of Cognitive Ability

The ECAD scores were examined in three studies that included the following external measures: the *Wechsler Preschool and Primary Scale of Intelligence™–Third Edition* (WPPSI™-III; Wechsler, 2002), the *Wechsler Preschool and Primary Scale of Intelligence–Fourth Edition* (WPPSI-IV; Wechsler, 2012), and the *Differential Ability Scales®–Second Edition* (DAS-II®; Elliott, 2007). Each of these external measures is an individually administered assessment of cognitive abilities. The WPPSI-III and DAS-II studies were conducted at the same time as the WJ IV/ECAD norming study and utilized the norming forms of the ECAD tests. The WPPSI-IV study was conducted in 2014; children in the WPPSI-IV study were administered the publication forms of the ECAD tests. Table 10 presents correlations for the ECAD GIA-EDev (*g*) and Expressive Language cluster scores with the composite measures of general intelligence (*g*) from the external batteries.

Table 10.
Correlations for Select ECAD Measures and Other Measures of Cognitive Abilities

Other Measure	*N*	ECAD GIA-EDev (*g*)	ECAD Expressive Language Cluster
Wechsler Preschool and Primary Scale of Intelligence–Third Edition (WPPSI-III)[a]	99	0.75	0.72
Wechsler Preschool and Primary Scale of Intelligence–Fourth Edition (WPPSI-IV)[a]	100	0.78	0.62
Differential Ability Scales–Second Edition (DAS-II)[b]	49	0.87	0.71

Note. Correlations were corrected for the variability of the ECAD norming sample.
[a] The measure reported is the Full-Scale IQ (*g*) score.
[b] The measure reported is the General Conceptual Ability (*g*) score.

The correlations in Table 10 provide support for the ECAD GIA-EDev cluster score as a measure of general intelligence and for the ECAD Expressive Language cluster score as a measure of early language development. The relatively lower correlation between the ECAD Expressive Language cluster and the WPPSI-IV general intelligence measure (compared to the WPPSI-III) is likely due to the reduced language demands of the

WPSSI-IV when compared to the WPPSI-III. These results are discussed further in the ECAD Comprehensive Manual (Wendling et al., 2015).

Correlations of the ECAD With Other Measures of Language

The ECAD scores were examined in several studies that included the publication forms of the ECAD tests along with the following external measures: the *Clinical Evaluation of Language Fundamentals®–Fourth Edition* (CELF®-4; Semel, Wiig, & Secord, 2003), the *Peabody Picture Vocabulary Test–Fourth Edition* (PPVT™-4; Dunn & Dunn, 2007), the *Comprehensive Assessment of Spoken Language* (CASL™; Carrow-Woolfolk, 1999), and the *Oral and Written Language Scales: Listening Comprehension/Oral Expression* (OWLS™; Carrow-Woolfolk, 1995). The CELF-4, CASL, and OWLS are individually administered multidimensional batteries of different aspects of oral language ability. The PPVT-4 is an individually administered measure of expressive vocabulary and word retrieval. Additionally, the co-norming of the ECAD and the WJ IV provided the unique opportunity to examine the relationship between the ECAD language measures and the WJ IV Tests of Oral Language (WJ IV OL) battery clusters using data from the norming study. Children in the WJ IV OL studies were administered the norming forms of the WJ IV and ECAD tests.

Table 11 displays correlations for the ECAD Expressive Language cluster with other measures of language. The moderate to strong correlations in Table 11 provide support for the ECAD Expressive Language cluster score as a measure of early language development. It should be noted that the ECAD Expressive Language cluster correlations with the WJ IV Oral Language cluster are spuriously high due to the shared content of the two clusters. The ECAD Expressive Language cluster contains the early development form of the WJ IV Picture Vocabulary test; the Picture Vocabulary test comprises one half of the WJ IV Oral Language cluster.

Table 11.
Correlations for the ECAD Expressive Language Cluster and Other Measures of Language

Other Measure	Age Range	*N*	ECAD Expressive Language Cluster
Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4)[a]	5–8	50	0.82
Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4)	5–8	50	0.79
Comprehensive Assessment of Spoken Language (CASL)[b]	3–6	50	0.52
Oral and Written Language Scales: Listening Comprehension/Oral Expression (OWLS)[c]	3–6	50	0.50
WJ IV Oral Language Battery[d]	3–5	631	0.77
WJ IV Oral Language Battery[d]	6–8	954	0.80
WJ IV Oral Language Battery[d]	9–10	620	0.81

Note. Correlations with the CELF-4, PPVT-4, CASL, and OWLS were corrected for the variability of the ECAD norming sample.

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

[b] The measure reported is the Core Composite.

[c] The measure reported is the Oral Composite.

[d] The measure reported is the Oral Language cluster.

Correlations for the ECAD With Other Measures of Achievement

The co-norming of the ECAD with the WJ IV provided a unique opportunity to investigate the relationship between the ECAD tests and clusters and the WJ IV measures of achievement. Data from children in the norming study sample who were administered norming versions of the ECAD tests and the WJ ACH tests were included in three separate age-specific analyses to investigate these relationships. Table 12 contains correlations between the ECAD achievement tests and Early Academic Skills cluster and the WJ IV domain-specific achievement clusters (Reading, Mathematics, and Written Language), Brief Achievement cluster, and Broad Reading cluster for three ECAD age ranges. The shading in Table 12 indicates shared content between the ECAD test/cluster and the WJ IV ACH cluster. Light shading indicates one test in common (for instance, the WJ IV Reading cluster is comprised of the Letter-Word Identification and Passage Comprehension tests), whereas dark shading indicates two tests in common (for instance, the ECAD Early Academic Skills cluster and the WJ IV Brief Achievement cluster both contain the Letter-Word Identification and Writing [called "Spelling" in the WJ IV] tests). The correlations in these shaded cells should be interpreted with caution, as they are spuriously inflated by the shared content between the two batteries.

Table 12.
Correlations for Select WJ IV and ECAD Achievement Measures

		ECAD Measure			
Age Range	**WJ IV Cluster**	**Test 8: Letter-Word Identification**	**Test 9: Number Sense**	**Test 10: Writing**	**Early Academic Skills Cluster**
3–5[a]	Reading	0.88	0.55	0.63	0.85
	Mathematics[d]	—	—	—	—
	Written Language	0.69	0.57	0.79	0.83
	Brief Achievement	0.90	0.71	0.86	0.98
	Broad Achievement[d]	—	—	—	—
6–8[b]	Reading	0.96	0.60	0.83	0.93
	Mathematics	0.70	0.79	0.71	0.81
	Written Language	0.87	0.60	0.89	0.91
	Brief Achievement	0.94	0.69	0.93	0.98
	Broad Achievement	0.89	0.67	0.87	0.94
9–10[c]	Reading	0.95	0.56	0.82	0.92
	Mathematics	0.69	0.73	0.69	0.78
	Written Language	0.83	0.52	0.90	0.89
	Brief Achievement	0.94	0.66	0.93	0.98
	Broad Achievement	0.87	0.66	0.88	0.93

Note. Cells shaded light gray indicate one test containing shared content in the ECAD/WJ IV correlation. Cells shaded dark gray indicate two tests containing shared content in the ECAD/WJ IV correlation.
[a] Sample sizes for correlations range from 591 to 631.
[b] Sample sizes for correlations range from 884 to 954.
[c] Sample size is 620 for all correlations.
[d] Sample sizes for the WJ IV Mathematics and Broad Achievement clusters were too small to interpret correlations in this age range.

Across all three age ranges in Table 12, domain-specific ECAD tests correlate highest with their corresponding WJ IV curriculum-area cluster. For example, at ages 6 through 8, the ECAD Number Sense test correlates highest with the WJ IV Mathematics cluster (.79) and lowest with the WJ IV Reading (.60) and Written Language (.60) clusters. This provides support for the ECAD achievement tests as valid measures of children's early academic skills within the reading, mathematics, and written language curriculum areas. The ECAD Early Academic Skills cluster correlates strongly with all the curriculum-specific (.78 to .93) as well as cross-domain (.93 to .98) WJ IV ACH clusters in Table 12, supporting the use of the ECAD Early Academic Skills cluster as a measure of overall early academic skills development.

Correlations for the ECAD With Early Childhood Developmental Measures

One study examined the relationship between select ECAD test and cluster scores and scores from the *Battelle Developmental Inventory, Second Edition* (BDI-2™; Newborg, 2005). The BDI-2 is a norm-referenced assessment of developmental skills in children from birth through age 7. Table 13 contains correlations between the GIA-EDev and Expressive Language clusters from the ECAD and the BDI-2 Cognitive and Communication Domain composite scores. The correlations in Table 13 are moderate to strong, suggesting that the ECAD tests and clusters are tapping into abilities that are similar to those measured by the BDI-2 Cognitive and Communication domains. The highest correlations in Table 13 are for the GIA-EDev cluster score and the BDI-2 Communication (.75) and Cognitive (.79) Domain scores, providing evidence to support the use of the ECAD GIA-EDev cluster as a valid measure of cognitive ability in developmental assessments with young children.

Table 13.
Correlations for Select ECAD Cluster Scores and BDI-2 Domain Scores

ECAD Measures	BDI-2 Domain Scores	
	Communication	Cognitive
Cluster		
General Intellectual Ability–Early Development	0.75	0.79
Expressive Language	0.74	0.72

Note. Sample size is 98 for all correlations. Correlations were corrected for the variability of the ECAD norming sample.

Performance of Clinical Samples on the ECAD Measures

The correlations between the ECAD scores and scores from external measures of similar constructs provide one form of validity evidence; the relationship between ECAD scores and clinical group status (e.g., children with developmental delay or autism) provides another form of test-criterion validity evidence. Selective tests were administered to children with cognitive delay, children with speech and/or language delays, and children with autism. The ECAD clinical validity study participants were drawn from a variety of educational and clinical settings. Patterns of ECAD cluster scores are as expected for the three groups, with the children in the Cognitive Delay group having the lowest scores for all three clusters (M = 73.8, SD = 19.3 for GIA-EDev; M = 79.4, SD = 21.4 for Expressive Language; and M = 77.2, SD = 15.6 for Early Academic Skills). Mean ECAD cluster scores for children in the Speech/Language Delay and Autism groups fall in the low-average range (88.0 to 91.8). The complete results of the clinical validity studies are reported in the ECAD Comprehensive Manual (Wendling et al., 2015).

Summary

The *Woodcock-Johnson® IV Tests of Early Cognitive and Academic Development* (ECAD; Schrank et al., 2015) primarily is intended to be used for evaluating early cognitive and expressive language abilities and early academic skills for children ages 2 years 6 months through 7. The ECAD also may be used for children ages 8 and 9 who have cognitive delays or intellectual disabilities.

Like the other WJ IV batteries, the ECAD is based on contemporary CHC theory. The tests are designed to identify any relative strengths and weaknesses among the child's abilities, as defined by CHC theory, so that early learning needs can be identified and targeted for improved educational outcomes.

The scaling procedures used for development and interpretation of the ECAD measures are particularly appropriate for determining the presence and severity of any developmental delay in cognitive ability and early academic skills. Test information provides developmental levels for several different cognitive abilities in addition to a global cognitive, or intellectual ability, score. Developmental levels for expressive language ability and early academic skills also are available on the ECAD.

The procedures followed in developing and standardizing the ECAD have produced an instrument that can be used with confidence in a variety of settings. This bulletin provided an overview of the technical information about the development and norming of the ECAD battery, following the standards proposed in the *Standards for Educational and Psychological Testing* (AERA et al., 2014). The ECAD Comprehensive Manual (Wendling et al., 2015) provides users with a comprehensive resource for evaluating the validity of the scores and interpretations from the ECAD battery for measuring children's cognitive and expressive language abilities and early academic skills. Interested users should consult the ECAD Comprehensive Manual for more in-depth details about the technical characteristics of the test.

References

Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). The impact of the Cattell-Horn-Carroll theory on test development and interpretation of cognitive and academic abilities. In D. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed.; pp. 185–202). New York, NY: Guilford Press.

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

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

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

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

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, United Kingdom: Cambridge University Press.

Carroll, J. B., & Maxwell, S. E. (1979). Individual differences in cognitive abilities. *Annual Review of Psychology, 30*, 603–640.

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

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

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

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

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

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

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

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

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

Horn, J. L. (1965). *Fluid and crystallized intelligence* (Unpublished doctoral dissertation). University of Illinois, Urbana–Champaign.

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. & Stankov, L. (1982). Auditory and visual factors of intelligence. *Intelligence, 6*, 165–185.

Jaffe, L. E. (2009). *Development, interpretation, and application of the* W *score and the relative proficiency index* (Woodcock-Johnson III Assessment Service Bulletin No. 11). Rolling Meadows, IL: Riverside.

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

Mather, N., & Wendling, B. J. (2014). Examiner's Manual. *Woodcock-Johnson IV Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside.

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

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. 151–180). New York, NY: Guilford Press.

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

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

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

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

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

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

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

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

National Research Council (2013). *Nonresponse in social science surveys: A research agenda*. Washington, DC: The National Academies Press.

Newborg, J. (2005). *Battelle Developmental Inventory* (2nd ed.). Rolling Meadows, IL: Riverside.

Rasch, G. (1980). *Probabalistic models for some intelligence and attainment tests*. Chicago, IL: University of Chicago Press. (Original work published 1960)

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

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

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

Schneider, W. J., & Newman, D. A. (2015). Intelligence is multidimensional: Theoretical review and implications of specific cognitive abilities. *Human Resource Management Review, 25*, 12–27.

Schrank, F. A., & Dailey, D. (2014, 2015). *Woodcock-Johnson Online Scoring and Reporting* [Online format]. Rolling Meadows, IL: Riverside.

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

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

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

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

Schrank, F. A., McGrew, K. S., & Mather N. (2015). *Woodcock-Johnson IV Tests of Early Cognitive and Academic Development*. Rolling Meadows, IL: Riverside.

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

Thomas, N., Raghunathan, T. E., Schenker, N., Katzoff, M. J., & Johnson, C. L. (2006). An evaluation of matrix sampling methods using data from the National Health and Nutrition Survey. *Survey Methodology, 32*(2), 217–231.

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

Wechsler, D. W. (2012). *Wechsler Preschool and Primary Scale of Intelligence* (4th ed.). San Antonio, TX: Pearson.

Wendling, B. J., Mather, N., LaForte, E. M., McGrew, K. S., & Schrank, F. A. (2015). Comprehensive Manual. *Woodcock-Johnson IV Tests of Early Cognitive and Academic Development*. Rolling Meadows, IL: Riverside.

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

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. (1994). Norms. In R. J. Sternberg (Ed.), *The encyclopedia of intelligence.* 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., & Dahl, M. N. (1971). *A common scale for the measurement of person ability and test item difficulty* (AGS Paper No. 10). San Antonio, TX: Pearson.

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

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

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

Wright, B. D., & Stone, M. S. (1979). *Best test design: Rasch measurement*. Chicago, IL: MESA Press.

Houghton
Mifflin
Harcourt™

3800 Golf Road, Suite 200
Rolling Meadows, IL 60008

800.323.9540
www.wj-iv.com