Woodcock-Johnson® III

Assessment Service Bulletin Number 11

Development, Interpretation, and Application of the *W* Score and the Relative Proficiency Index

Lynne E. Jaffe, PhD

The Woodcock-Johnson III (WJ III®) *(Woodcock, McGrew, & Mather, 2001a) provides a wide variety of score options for interpreting an individual's test performance. Many of these scores, such as standard scores (SS), percentile ranks (PR), age equivalents (AE), and grade equivalents (GE) are provided by most other educational and psychological tests. However, the WJ III tests are unique in providing two metrics that report the quality of an individual's performance: the* W *score and the relative proficiency index (RPI). The* W *score is the foundational metric—the score on which all of the other WJ III scores are based—and it is useful for measuring an individual's progress over time. The RPI is a measure of a person's proficiency in a skill, ability, or area of knowledge compared with average age or grade peers. Since the* W *score and the RPI are not available in most other assessments, many psychologists and diagnosticians may be unaware of the clinical utility of these metrics. The purpose of this bulletin is to familiarize users of the WJ III with the development, interpretation, and application of the* W *score and the RPI. Specifically, this bulletin describes the levels of interpretive information available in the WJ III, explains the special characteristics and usefulness of the* W *scale, and describes how the RPI fits into the hierarchy of information used to interpret test results. In addition, the bulletin explains the differences between the RPI and peer-comparison scores and the usefulness of the RPI in clarifying diagnostic profiles and designing interventions. Finally, it describes considerations for using the RPI in view of the Individuals with Disabilities Education Act (IDEA) 2004 and discusses the use of the RPI in clinical research.*

RIVERSIDE

HOUGHTON MIFFLIN HARCOURT

The author appreciates the generosity of Richard Woodcock, Fredrick Schrank, Nancy Mather, and Kevin McGrew for their information regarding the content of this paper and their permission to use material from their publications.

Reference Citation

- To cite this document, use:

  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 Publishing.

For technical information, please call 800.323.9540 or visit our website at www.woodcock-johnson.com

1 2 3 4 5 6 7 8 9 10-XBS-12 11 10 09

Development, Interpretation, and Application of the *W* Score and the Relative Proficiency Index

Levels of Interpretive Information

Four levels of interpretive information are provided by the *Woodcock-Johnson III* (WJ III) batteries (Mather & Woodcock, 2001; Woodcock, 1999), including qualitative information, level of development, degree of proficiency, and relative standing in a group.

The four levels of test information are cumulative; that is, each level provides different information about a person's test performance, and each successive level builds on information from the previous level. Information from one level is not interchangeable with information from another. For example, standard scores cannot be used in place of age or grade equivalents, or vice versa. Consequently, to interpret and describe a person's performance completely, information from all four levels must be considered.

Table 1 describes the hierarchy of interpretive information available from the WJ III batteries.

Table 1.
Hierarchy of Interpretive Information Available From the WJ III Batteries

Level	Type of Information	Basis	Information and Scores	Uses
1	Qualitative (Criterion Referenced)—Describes context or supports a clinical hypothesis	Observations during testing and analysis of responses	• Description of the individual's behavior during testing • Patterns of errors and correct responses within specific tasks • Strategies (correct or erroneous) used to perform specific tasks	• Consideration of the possible effect of the individual's behavior on the obtained test scores • Prediction of the individual's behavior and reactions in instructional situations • Analysis of an individual's strengths, misunderstandings, and limitations regarding specific academic skills, procedures, knowledge, and cognitive abilities • Instructional recommendations for specific skills

Table 1, continued

Hierarchy of Interpretive Information Available From the WJ III Batteries

Level	Type of Information	Basis	Information and Scores	Uses
2	Level of Development (Norm Referenced)—Indicates an individual's level of development, such as age or grade equivalents	Sum of item scores Age or grade level in the norming sample at which the median score is the same as the individual's score	• Raw score • [1]Test or cluster *W* score • Age equivalent (AE) • Grade equivalent (GE)	• Reporting an individual's general level of development in a skill, ability, or area of knowledge compared with others of the same age or in the same grade in the norming sample • Monitoring an individual's progress within a specific skill or ability • Basis for describing the implications of developmental strengths and weaknesses • Basis for initial recommendations regarding instructional level and materials • Placement decisions based on a criterion of significantly advanced or delayed development
3	Proficiency (Criterion Referenced)—Indicates the quality of performance on criterion tasks of a given difficulty level	Distance of an individual's score on the *W* scale from an age or grade reference point	• Quality of performance on assessed skills and abilities compared to that of age or grade peers in the norming sample • [1]Test or cluster *W* difference (*W* DIFF) • Relative proficiency index (RPI) • Cognitive-academic language proficiency (CALP) level • Instructional or developmental zone	• Degree of proficiency on tasks mastered by average age or grade peers • Developmental level at which the individual will perceive typical tasks to be easy, mildly challenging, or very difficult • Placement decisions based on a criterion of significantly strong or weak proficiency • Prediction of performance with similar task
4	Relative Standing in a Group (Norm Referenced)—Provides a basis for making peer comparisons (percentile ranks or standard scores)	Relative position (A transformation of a difference score, such as dividing it by the standard deviation of the reference group)	• Rank order • [1]Standard score (SS) (including *T* score, *z* score, NCE, discrepancy *SD* DIFF) • Percentile rank (PR) (including discrepancy PR)	• Statement of the relative (ordinal) position of an individual's score, based on the standard deviation (SD), within the range of scores obtained by age or grade peers in the norming sample • Placement decisions based on a criterion of significantly high or low standing in a group

Note. Adapted from *Examiner's Manual: Woodcock-Johnson III Tests of Cognitive Abilities* (p. 66) by N. Mather and R. W. Woodcock, 2001, Rolling Meadows, IL: Riverside Publishing. Copyright 2001 by The Riverside Publishing Company. Adapted with permission.

[1] Equal interval units are the preferred metric for statistical analyses.

Descriptions of Interpretive Levels

Text Box 1.
Examples of qualitative information.

- **Behavior:** If a child is inattentive during the Numbers Reversed test, the score might underestimate his or her actual working memory ability.
- **Response pattern:** An individual who has numerous misspellings that do not approximate phonetic spellings may have a specific weakness in phonological awareness and/or phonics.
- **Compensatory strategy:** A person who obtains an adequate score on the Calculation test but uses repeated addition rather than multiplication indicates a lack of knowledge of math facts and possibly the multiplication algorithm.

Level 1: Qualitative Information

Qualitative information is obtained through observation and analysis: observation of a person's behavior during testing and analysis of the task demands, the person's responses (correct and incorrect) to test items, and the strategies used to generate those responses. This type of information is critical to understanding and interpreting the scores obtained by an individual. Often a description of a person's behaviors during the test and the strategies he or she used to obtain a particular score are as important as the information provided by the score itself. Qualitative analysis is one of the cornerstones of proper individualized assessment and is an integral part of the interpretation and reporting of test results.

Level 2: Level of Development

The second level of information contains scores that report the level of development in the skill, ability, or area of knowledge (trait) measured. Typically, raw scores are converted into age equivalents (AE) and grade equivalents (GE). In the WJ III batteries, the raw score is converted into a *W* score, from which age and grade equivalents are derived.

Text Box 2.
Definition of trait.

Trait: Skill, ability, or area of knowledge

Level 3: Degree of Proficiency

This level provides information about a person's proficiency in specific tasks when compared to age or grade peers and serves as criterion-referenced scores. Although proficiency scores may be the most useful of the four levels of test information, most other assessment measures do not offer these scoring options. Level 3 supplies the relative proficiency index (RPI), instructional and developmental zones (which are derived from the RPI), and the cognitive-academic language proficiency (CALP).

Level 4: Relative Standing in a Group

Level 4 contains peer-comparison scores, the most commonly used scores in educational and clinical settings. These are norm-referenced scores, such as standard scores (SS) and percentile ranks (PR), that describe a person's relative standing, or rank order, in comparison to age or grade peers. In contrast to Level 3 scores, Level 4 scores indicate ordinal position in a group, not the quality of performance. The significance of this difference is discussed in the section titled, "Relative Proficiency Index."

The *W* Scale

The unit of the *W* scale, the *W* score, is the foundational metric for all derived scores (e.g., standard scores, percentile ranks, relative proficiency indexes) available for the WJ III assessments (McGrew & Woodcock, 2001). Developed by Richard Woodcock and Marshall Dahl (1971), the *W* scale is a mathematical transformation of the Rasch model of data analysis, which is based on item response theory. In modifying the parameters of the associated statistical software program, Woodcock and Dahl (1971) produced a measurement scale with special characteristics that have advantages for test developers and for evaluators in interpreting test results.

This section describes these characteristics, the procedure for representing levels of ability and item difficulty on the *W* scale for each test and cluster in the WJ III assessments, and the utility of the *W* scale, both for reporting an individual's growth in a measured trait and for predicting his or her chances of success on a specific task at any difficulty level.

Characteristics of the *W* Scale

Text Box 3.
Definition of an equal-interval scale.

Equal-interval scale: Any given interval on an equal-interval scale represents the same amount of difference in the skill or ability measured, regardless of the interval's location along the scale.

Particularly useful characteristics of the *W* scale include the following:

- The *W* scale is an equal-interval scale. On an equal-interval scale, a given interval, such as 3 points, represents the same difference (e.g., amount of growth) in the trait measured, regardless of where that interval is along the scale or what is being measured. This characteristic allows the differences between two sets of scores situated anywhere on the scale to be compared. A more familiar example of an equal-interval scale is a ruler; an interval of 3 inches represents the same difference in length between 1 and 4 inches as it does between 37 and 40 inches. Equal-interval scales are generally considered the most appropriate scales for statistical calculations (Woodcock, 1999).
- A person's ability level in a measured trait and the difficulty levels of all items are both represented on the same scale (the *W* scale), allowing them to be used as factors within the same mathematical calculations. This characteristic "provides a particularly convenient set of predictive relationships based on the difference between a person's ability and item difficulty along the scale" (Woodcock, 1978, p.77).

Development of Critical Values: The Reference *W* and *W* Difference

The norming of a test is actually done in stages. Using the Rasch model of measurement, the initial stage is *calibration* of the items. All items being considered for inclusion in a test are administered to a large group of respondents. The resultant data are evaluated by Rasch statistical software, which:

- identifies any items that are a poor fit (i.e., do not match the assessment intention) and require omission or revision.
- sorts the items by difficulty level. The more people who respond correctly to an item, taking into account their ability levels, the easier the item. The fewer people who respond correctly, the harder the item.
- assigns a value to each item representing its difficulty level. This value is termed the *W* difficulty. The *W* difficulty of an item determines its relative position on the *W* scale. The easier the item, the lower the *W* difficulty; the harder the item, the higher the *W* difficulty.
- generates a *W* score for every possible raw score for the test. Consequently, for any raw score a person might obtain, there is an associated *W* score that represents his or her ability level in the task. This *W* score is termed the *W* ability. The *W* scale for each test is centered on a value of 500, which is set to approximate the average performance of a typical child age 10–0 (when using age norms) or at the beginning of grade 5 (when using grade norms). The typical range of *W* abilities within a test is about 430 to 550, although it can be wider or narrower depending on the trait being measured (McGrew & Woodcock, 1989; R. W. Woodcock, personal communication, August 4, 2007).

Text Box 4.
Definition of W *difficulty.*

> ***W* difficulty:** The *W* difficulty of an item is indicated by its relative position on the *W* scale. Higher *W* difficulties are associated with more difficult items; lower *W* difficulties are associated with easier items.

Text Box 5.
Definition of W *ability.*

> ***W* ability:** The *W* score that represents the individual's level of ability on the task presented.

During the next stage of test development, when data have been collected from the entire norming sample, the *W* scores are *anchored*, or linked, to age and grade levels in increments of year and month. For each age and grade group in the norming sample, the median *W* ability value is identified. This corresponds to the difficulty level at which 50% of the group responded correctly and 50% responded incorrectly. Thus, the median *W* ability represents the average difficulty level that each group can manage. That value is designated the reference *W*, the criterion score against which the performance of a person within that group is measured (McGrew & Woodcock, 1989; R. W. Woodcock, personal communication, August 15, 2007). For example, in the Passage Comprehension test, 50% of children age 11 years, 6 months (11–6) obtained a *W* ability of 509 (raw score 33) or above, and 50% scored 509 or below. Accordingly, the *W* ability of 509 was established as the reference *W* for children age 11–6 on Passage Comprehension (based on *WJ III Normative Update Compuscore and Profiles Program*, Woodcock, Schrank, McGrew, & Mather, 2007).

Text Box 6.
Definition of reference W.

> **Reference *W*:** The median *W* score for any age or grade group on a specific test. The reference *W* is the criterion—the score against which an individual's *W* ability is compared. The reference *W* represents the difficulty level of a hypothetical item to which 50% of the age or grade group would respond correctly.

The difference between a person's obtained *W* ability and his or her peer group's reference *W* is termed the *W* difference. For example, a child age 11–6 who obtained a *W* ability of 519 on the Passage Comprehension test would have a *W* difference of +10

(519–509). The *W* difference is the value from which standard scores, percentile ranks, and relative proficiency indexes are derived (Woodcock, 1999).

Text Box 7.
Definition of W *difference.*

> ***W* difference (*W* DIFF):** The difference in *W* units between an individual's W ability and the reference *W* (median score of the individual's age or grade peers).

Utility of the *W* Scale for Reporting Growth

Because the *W* scale is an equal-interval scale, it is particularly useful for reporting an individual's growth in a skill, ability, or area of knowledge. An increase in a person's *W* ability represents actual growth in the trait measured. The *W* scale is constructed so that an increase of 10 *W* units represents the individual's ability to perform, with 75% success, tasks that he or she could previously perform with 50% success. This is true for any 10-point increase on the *W* scale, regardless of the ability being measured or the difficulty level of the task (e.g., a child's ability to identify letters or a college student's ability to solve higher-level math problems). Accordingly, if a person's proficiency in an ability increases from one testing to the next, his *W* ability will increase also, reflecting his progress (Woodcock & Dahl, 1971; Woodcock, 1999).

In contrast, peer-comparison scores do not show growth as clearly because they describe a person's ordinal position, or rank order, in a group rather than his or her proficiency. If a person improves in a trait at the same rate as his or her peers, regardless of the amount of improvement, the resultant score will be the same as it was on the previous testing (see Figure 1). Consequently, the *W* score is a more informative metric than peer-comparison scores for examining and reporting growth.

Michael was tested on the WJ III Passage Comprehension test at age 8 years, 3 months (8–3) and again at age 11 years, 2 months (11–2).

Although his *W* ability had increased by 26 points, his standard score (SS) remained the same.

	WJ III Passage Comprehension			
Age	*W* Ability	Increase	SS	PR
8–3	462		84	14
11–2	488	+26	84	14

Michael did, in fact, make progress in reading comprehension, but he did so at about the same rate as other students *at the 14th percentile of his age peers.* Thus, the normative scores do not indicate a change. It is the *W* score that demonstrates the amount of progress he has made. A 26-point increase indicates that the sentence difficulty level Michael could previously read and understand with 50% success, he can now read and understand with 94% success (see Table 2 on page 8).

Figure 1.
Growth illustrated by W *score versus peer-comparison score.*

Utility of the *W* Scale for Predicting Success on Items and Tasks at Varying Difficulty Levels

A significant advantage of having *W* ability and *W* difficulty on the same scale is that, given a person's ability level (*W* ability), there is a mathematical basis for predicting the individual's likelihood of success on an item at any other level of difficulty (Woodcock, 1999)—most importantly, at the average difficulty level for his or her peer group (the reference *W*). The degree of probability is a function of the size and direction (positive or negative) of the *W* difference. As shown in Table 2, when a person's *W* ability equals the *W* difficulty, the *W* difference is 0 (zero). Because the reference *W* for any age or grade group is the median difficulty level, the person has a 50% chance of responding correctly to any item of equal difficulty. It is likely, then, that the person will be able to handle similar tasks of equal difficulty as well as average age or grade peers could handle them. Accordingly, if the person's ability level is higher than the reference *W*, his or her odds of success increase, and his or her performance is likely to exceed that of peers. Conversely, if the individual's ability level is lower than the reference *W*, the odds of success decrease, and he or she is likely to have more difficulty than peers (Woodcock, 1978, 1999). The larger the *W* difference in a positive direction, the higher the probability of success; the larger the *W* difference in a negative direction, the lower the probability of success. As noted previously, during the norming process, reference Ws were established for every month of every age and grade level. Consequently, if a person's *W* ability is known, a mathematical prediction can be made regarding his or her degree of success at the level of task difficulty where age or grade peers will score 50%.

Based on this capability, the person's probability of success can be projected. In Table 2, the *W* difference of 0 (zero) is equated with a 50% probability of passing the criterion item. A *W* difference of +10 (*W* ability − reference *W*) indicates a 75% probability of success, while a *W* difference of −10 indicates a 25% probability of success (Woodcock, 1978; Woodcock & Dahl, 1971).

Table 2.
Probability of Success Given the Difference on the W *Scale Between Ability and Difficulty*

Ability Minus Difficulty (W_{A-D})	Probability of Success (*P*)	Ability Minus Difficulty (W_{A-D})	Probability of Success (*P*)
+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

Note. From *Development and Standardization of the Woodcock-Johnson Psycho-Educational Battery* (p. 79), by R. W. Woodcock, 1978, Rolling Meadows, IL: Riverside Publishing.

Relative Proficiency Index

A practical limitation with using 50% as the reference *W* is that, in education, 90%, not 50%, is more typically considered proficient. Consequently, in developing the relative proficiency index, the reference *W*—the criterion—is set at a difficulty level 20 *W* units easier than the median *W* (Woodcock, 1999). Accordingly, the relative proficiency index is a criterion-referenced score describing the probability of a person's success on a task similar to that used in the assessment, at the level of difficulty that 90% of average grade or age peers can manage.

The RPI is represented as a fraction, with the person's expected level of success as the numerator and the 90% criterion as the denominator. For example, an RPI of 60/90 suggests that the person would be about 60% successful on a task that typical peers would perform with 90% success.

Text Box 8.
Effect of changing reference W *from 50% to 90% proficiency criterion.*

Jennifer's scores exemplify the change in the reference *W* from a 50% to a 90% proficiency criterion. Jennifer, age 13-9, obtained a raw score of 11 on the WJ III Word Attack test. Using age norms, the Rasch program transformed her raw score to a *W* ability of 496. The reference *W* score on Word Attack for Jennifer's age group is 518, so the W difference is −22.

Using the original table (Table 2), Jennifer's probability of success is 8% at a difficulty level where her typical age peers' probability of success is 50%. Using the revised table (Table 3, now centered on 90%), and the same *W* difference of −22, Jennifer has a 45% chance of success at the difficulty level where her typical age peers would have a 90% chance of success. The *W* difference score is the same, but the prediction is more relevant to classroom expectations.

The advantage of the RPI over other types of criterion-referenced scores is that the criterion items and the person's proficiency level are derived from norms—real scores from real people. Table 3 shows the range of RPIs with their corresponding *W* differences.

Reporting RPIs Using Descriptive Labels

A useful feature of the RPI as presented in the WJ III is the choice of descriptive labels for different levels of proficiency, functioning, and development associated with an individual's *W* difference or RPI (see Table 4 on page 10). The examiner can select the type of descriptor typically used within his or her profession. In education, for example, "Proficiency" might be used to describe academic achievement, while "Development" might be used to describe cognitive and language abilities. "Implications" represents the individual's perceived level of difficulty or facility with the task (Schrank & Woodcock, 2002).

Table 3.
Relative Performance Indexes (RPI) Associated With W *Differences (*W *DIFF) Along the* W *Scale*

***W* DIFF**	**RPI**	***W* DIFF**	**RPI**	***W* DIFF**	**RPI**
29 and above	100[1]/90	−1	89/90	−36	15/90
28	99/90	−2	88/90	−37	13/90
27	99/90	−3	87/90	−38	12/90
26	99/90	−4	85/90	−39	11/90
25	99/90	−5	84/90	−40	10/90
24	99/90	−6	82/90	−41	9/90
23	99/90	−7	81/90	−42	8/90
22	99/90	−8	79/90	−43	7/90
21	99/90	−9	77/90	−44	7/90
20	99/90	−10	75/90	−45	6/90
19	98/90	−11	73/90	−46	5/90
18	98/90	−12	71/90	−47	5/90
17	98/90	−13	68/90	−48	4/90
16	98/90	−14	66/90	−49	4/90
15	98/90	−15	63/90	−50	4/90
14	98/90	−16	61/90	−51	3/90
13	97/90	−17	58/90	−52	3/90
12	97/90	−18	55/90	−53	3/90
11	97/90	−19	53/90	−54	2/90
10	96/90	−20	50/90	−55	2/90
9	96/90	−21	47/90	−56	2/90
8	96/90	−22	45/90	−57	2/90
7	95/90	−23	42/90	−58	2/90
6	95/90	−24	39/90	−59	1/90
5	94/90	−25	37/90	−60	1/90
4	93/90	−26	34/90	−61	1/90
3	93/90	−27	32/90	−62	1/90
2	92/90	−28	29/90	−63	1/90
1	91/90	−29	27/90	−64	1/90
0	90/90	−30	25/90	−65	1/90
		−31	23/90	−66	1/90
		−32	21/90	−67	1/90
		−33	19/90	−68	1/90
		−34	18/90	− 69 and	0[2]/90
		−35	16/90	below	

Note. Copyright (1999) From "What Can Rasch-Based Scores Convey About a Person's Test Performance?" by R. W. Woodcock, 1999, in *The New Rules of Measurement: What Every Psychologist and Educator Should Know* (pp. 105–127) by S. E. Embretson & S. L. Hershberger, (Eds.), Mahwah, NJ: Erlbaum. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.

[1] Approximate value (>99.5).

[2] Approximate value (<0.5).

Table 4.
Descriptive Labels and Implications Corresponding to W *Differences (*W *DIFF) and Relative Proficiency Indexes (RPI)*

***W* Difference**	**RPI**	**Proficiency**	**Functionality**	**Development**	**Implications for Age- or Grade-Level Tasks**
+31 and above	100/90	Very Advanced	Very Advanced	Very Advanced	Extremely Easy
+14 to +30	98/90 to 100/90	Advanced	Advanced	Advanced	Very Easy
+7 to +13	95/90 to 98/90	Average to Advanced	Within Normal Limits to Advanced	Age-Appropriate to Advanced	Easy
−6 to +6	82/90 to 95/90	Average	Within Normal Limits	Age-Appropriate	Manageable
−13 to −7	67/90 to 82/90	Limited to Average	Mildly Impaired to Within Normal Limits	Mildly Delayed to Age-Appropriate	Difficult
−30 to −14	24/90 to 67/90	Limited	Mildly Impaired	Mildly Delayed	Very Difficult
−50 to −31	3/90 to 24/90	Very Limited	Moderately Impaired	Moderately Delayed	Extremely Difficult
−51 and below	0/90 to 3/90	Negligible	Severely Impaired	Severely Delayed	Impossible

Note. From *Manual and Checklists. Report Writer for the WJ III* (p. 10), by F. A. Schrank and R. W. Woodcock, 2002, Rolling Meadows, IL: Riverside Publishing. Copyright 2002 by Riverside Publishing. Adapted with permission.

Sample Statements for Reporting RPIs

The following are examples of statements that might be used to describe an individual's RPIs (Mather & Jaffe, 2002, pp. 30–31). Specific wordings will vary depending on the achievement area or cognitive ability being addressed and the level of the RPI.

> Mark's level of proficiency on the Broad Mathematics cluster was limited (RPI 66/90). He is likely to find grade-level tasks requiring mathematics to be very difficult.
>
> Sam's RPI of 21/90 on the Phoneme/Grapheme cluster indicates that on similar tasks in which the average fourth-grade child would demonstrate 90% proficiency, Sam would demonstrate 21% proficiency. Sam's knowledge of phoneme-grapheme correspondence and spelling patterns is very limited. He is likely to find grade-level reading and spelling tasks extremely difficult.
>
> Although Nicholas's standard score on the Basic Reading Skills cluster is within the average range for seventh-graders overall, his RPI (45/90) indicates that he will have considerably more difficulty than most of his grade peers in tasks requiring basic reading skills.
>
> Bryn's RPI of 98/90 on Visual-Spatial Thinking signifies advanced development. When average age peers demonstrate 90% accuracy on similar tasks, Bryn's expected accuracy would be approximately 98%. She is likely to find visual-spatial tasks very easy.
>
> David's RPI of 54/90 in Auditory Memory Span represents a mild delay in the skills necessary for similar classroom tasks, such as repeating a set of instructions. His

expected success in doing so would be 54% compared with his classmates' 90% success. He is likely to find many tasks requiring rote memorization to be very difficult.

Jason's RPI on the Verbal Comprehension test was 75/90, suggesting that, compared with typical age peers, his proficiency is at the lower end of the instructional range.

Luz obtained a standard score of 87 in Passage Comprehension, which is within the low average range. Nevertheless, her RPI of 41/90 indicates that she is limited in her ability to comprehend brief passages, and this type of task will be very difficult for her.

Interpreting Instructional and Developmental Zones

Text Box 9.
Development of a ninth-grade student's instructional zone.

Trent, a young man in ninth grade, obtained a *W* ability of 453 on the WJ III Basic Reading Skills cluster. The *WJ III Normative Update Compuscore and Profiles Program* (Schrank & Woodcock, 2007) showed his instructional zone as GE 1.8 to 2.4, indicating that decoding words at the late first-grade level was likely to be easy for him and decoding words at the mid-second-grade level would be difficult.

Trent's instructional zone was generated as follows. A test item or similar task that would be very easy for Trent would have a *W* difficulty of 443, 10 points *below* his ability level. The group in the norming sample with a median score of 443 on the Basic Reading Skills cluster was in grade 1.8. Consequently, tasks appropriate for students in grade 1.8 would likely be at Trent's independent level, the lower end of his instructional zone.

Tasks expected to be very difficult for Trent (at a frustration level) would have a *W* difficulty of 463, 10 points *above* his ability level. The group with a median score of 463 was in grade 2.4. Consequently, tasks involving decoding and identifying words that are appropriate for students in grade 2.4 would likely be at Trent's frustration level, the upper limit of his instructional zone.

The instructional zones and developmental zones available on the WJ III Normative Update (NU) Compuscore and Profiles Program (Schrank & Woodcock, 2007) display RPI ranges in bands. Each zone encompasses a range from 10 *W* units below an individual's *W* ability (RPI 75/90) to 10 *W* units above an individual's ability (RPI 96/90) (see Table 3). The lower and higher limits of the zones are represented by age or grade equivalents for use in instructional planning. Tasks similar in difficulty level to items at the lower end of the range will be quite easy for the individual; those at the higher end will be quite difficult. Instructional zones are reported for academic achievement tests and clusters; developmental zones are reported for cognitive abilities tests and clusters.

Psychologists and diagnosticians will undoubtedly note that usefulness of the instructional zones for program planning depends on how closely the instructional materials used in their local schools are aligned with the ability levels of their students' age or grade peers nationally (those represented in the norming sample). Also, regardless of the curriculum, different publishers may use different methods to develop their instructional materials vis-à-vis grade levels. For example, third-grade reading material in one basal reading series may be generally easier or harder than third-grade reading material in another basal reading series. Furthermore, the readability level of either of these reading series may be quite different than that of a trade book that a publisher has labeled as having a third-grade readability level. When making instructional decisions, the psychologist or diagnostician must consider the alignment of the school's instructional materials with national standards. Qualitative analysis of the student's skills

and abilities can help to fine-tune the match between the instructional zone indicated by the WJ III, the specific materials used in the local school, and the actual functioning of the student.

The Difference Between RPI and Peer-Comparison Scores

A common misconception is that peer-comparison scores, such as standard scores or percentile ranks, indicate ability or achievement levels. In fact, this is not true. Rather, they merely show a person's rank order or "place in the line"—the position in which his or her score falls within the distribution of scores obtained by age or grade peers in the norming sample. In contrast, the RPI describes the person's level of proficiency in the skill, ability, or area of knowledge based on the probability of his or her success on a specific level of task difficulty.

The Apparent Contradiction Between an RPI and a Peer-Comparison Score

Occasionally, an evaluator will note a marked difference between the RPI and a standard score (SS) on a test or cluster. For example, on the Letter-Word Identification test (LWI), Leo, a sixth-grade student, obtained an RPI of 39/90 and a standard score of 86. Although these scores appear incongruous, actually, they are not because the RPI and the standard score are derived from separate and unrelated calculations. The derivation of each score is as follows:

RPI = 39/90: For grade 6.1, the reference *W* for LWI is 515; Leo's *W* ability was 491, producing a *W* difference of −24. Referring to Table 3, this *W* difference translates to an RPI of 39/90. Note that standard deviations (*SD*) are not listed in Table 3. The *SD* is not involved in this calculation.

SS = 86: This score is based on the relationship between the *W* difference and the SD. As noted, Leo's *W* difference is −24. The *SD* in *W* units for LWI at grade 6.1 is 25. Accordingly, Leo's *z* score is −0.96:

$$z = W \text{ difference} \div SD$$

The following algorithm uses the *z* score to generate a standard score with a mean (*M*) of 100 and a *SD* of 15:

$$M - (z \times 15) = \text{SS}$$

Using this equation, Leo's standard score is calculated: 100 – (0.96 × 15) = 85.6, which converts to a rounded standard score of 86.

The Effect of Standard Deviations on Peer-Comparison Scores

When the standard deviation of the *W* scores for a test within a specific age or grade group is relatively large (as it is in the example of Leo) or relatively small, it is more likely that there will be a marked difference between the RPI and the standard score. Because the standard score is tied to the standard deviation, standard scores vary accordingly, whereas the same *W* difference will always produce the same RPI (see Table 3). Consequently, "trying to compare the RPI to the standard score is like trying to compare apples to ostriches" (R. W. Woodcock, personal communication, June 4, 2007).

The lack of a constant relationship between the RPI and the standard deviation is illustrated in Figure 2, which shows Leo's scores on the LWI and Calculation tests. At grade 6.1, the mean (*M*) of both tests is the same (515), and, as it happened, Leo

obtained a *W* ability of 491 on both tests, resulting in a *W* difference of −24 and an RPI of 39/90 (see Table 3) for both tests. The standard deviations, however, are different for each test. As stated above, the *SD* of LWI is 25 *W* units—just slightly larger than Leo's *W* difference. Accordingly, his *W* ability falls just inside the lower limit of 1 *SD*. When converted into a standard score scale (as shown above), his standard score on LWI is 86. The standard deviation of Calculation is 16. With the same *W* difference (−24), Leo's *W* score now falls 8 *W* units outside the lower limit of 1 *SD*, resulting in a standard score of 78. Note that the only difference between Leo's standard score of 86 on LWI and 78 on Calculation is the size of the standard deviations at grade 6.1 for each test.

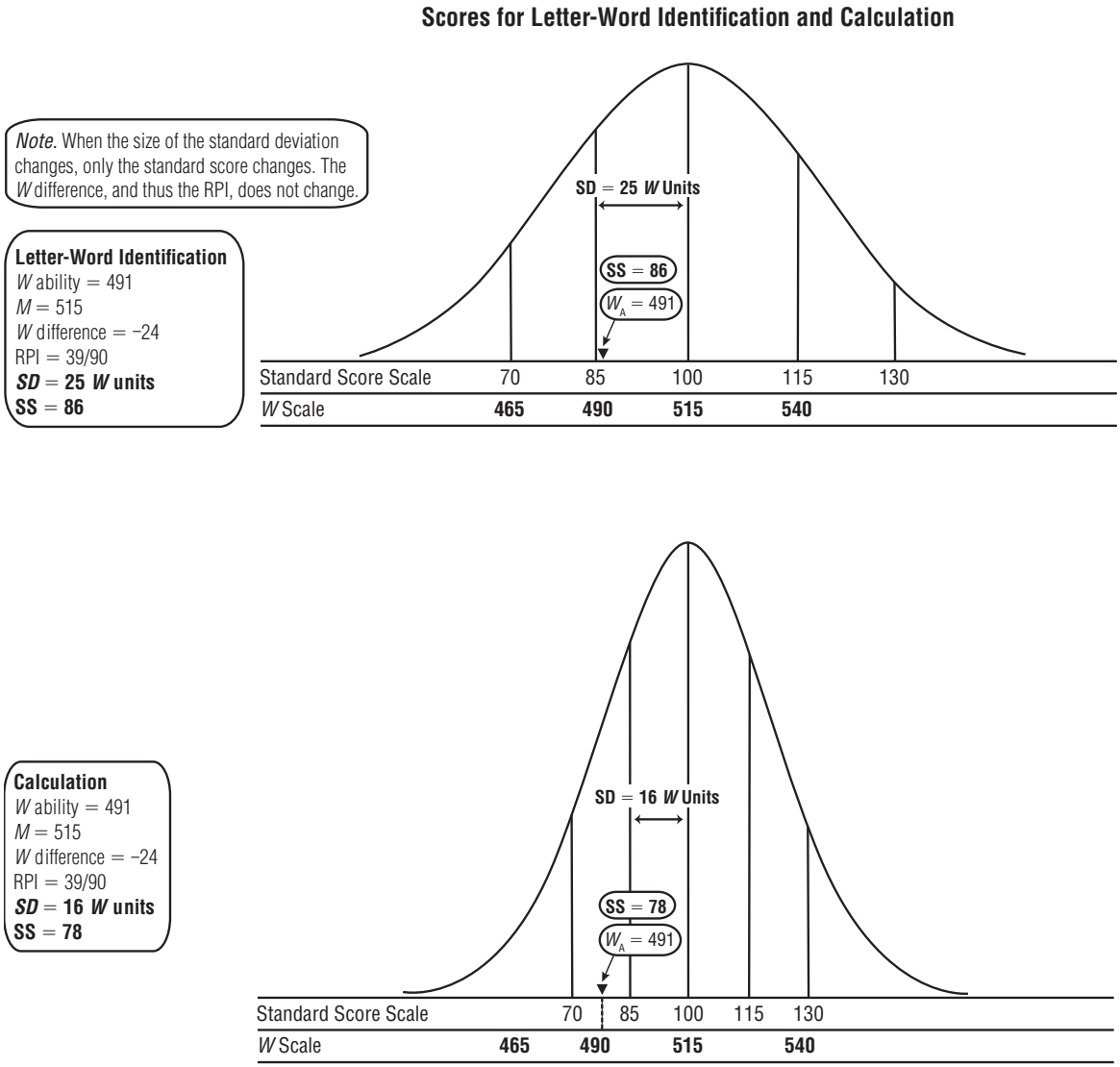

Figure 2.
*Comparison of a sixth-grade student's relative proficiency indexes (RPI), standard scores (SS), and standard deviations (*SD*) on two tests.*

Proficiency Versus Position

In evaluations, the relative proficiency index makes an important contribution that cannot be derived from peer-comparison scores. For example, if a child's standard score in a cognitive ability is not significantly low, the examiner may not consider a weakness in this ability as a possible explanation for the child's academic difficulty. The weakness in this cognitive ability may become obvious, however, when the RPI is considered. Moreover, the standard score may underestimate the child's academic weakness, whereas the RPI might more accurately reflect his or her level of performance in the classroom. Many psychologists have had the experience of telling a teacher that a student scored in the average or low average range on a test, only to have the teacher respond with disbelief because the score does not reflect the child's struggles in daily class work. Because of the difference between the information provided by peer-comparison scores and that provided by the RPI, the results of an evaluation might be misinterpreted if RPI scores are not considered. Woodcock (1999) illustrated the importance of recognizing the difference between *position* in a distribution and level of *competence* by explaining that people with visual or hearing problems are usually classified as having a disability and in need of services based on weaknesses in the quality of their performance in vision and hearing tasks, not because they fall below some point on a norm-referenced scale (p. 109).

The Case of Jeremy

An example of the importance of this distinction is the case of Jeremy, age 7–9, who is making very limited progress in reading and spelling. Despite individualized tutoring by the first-grade teacher, Jeremy has entered second grade unable to sound out three-letter words. His father reports that he reads with Jeremy every evening, going through the text word by word, reviewing the letter sounds, and helping Jeremy sound out the words. He comments, "We'll work on a word on one page, and I'll think he's got it, and then he doesn't recognize it later on the same page." Jeremy has recently begun claiming that he is sick in the mornings before school and pleads to stay home. Clearly, Jeremy is seriously delayed in his development of reading and spelling skills, a situation that is producing secondary emotional problems.

Jeremy's parents have requested an evaluation to determine his current achievement in reading and spelling and to ascertain the reasons for his difficulty in learning these skills. Jeremy is administered selected tests from the *Woodcock-Johnson III Tests of Cognitive Abilities* (Woodcock, McGrew, & Mather, 2001c), *Woodcock-Johnson III Diagnostic Supplement to the Tests of Cognitive Abilities* (Woodcock, McGrew, Mather, & Schrank, 2003), and *Woodcock-Johnson III Tests of Achievement* (Woodcock, McGrew, & Mather, 2001b). Table 5 shows Jeremy's standard scores and RPIs in reading and writing skills.

Table 5.
Jeremy's Standard Scores (SS) and Relative Proficiency Indexes (RPI) in Reading and Writing with Descriptive Labels

CLUSTER/Test	SS	RPI	Proficiency	Implications
BASIC READING SKILLS	77	9/90	Very Limited	Extremely Difficult
WRITTEN EXPRESSION	88	69/90	Limited to Average	Difficult
Passage Comprehension	77	22/90	Very Limited	Extremely Difficult
Spelling	78	23/90	Very Limited	Extremely Difficult

Analysis of Jeremy's responses during the evaluation indicates that when sounding out two- and three-letter words, he usually produces the correct sounds individually but is unable to retain the sounds and their sequence when blending them (e.g., *nap* becomes *pen*). He does not recognize common sight words (e.g., *here*, *they*), and his spelling is likely to represent correctly only the first and last sounds of a word (e.g., *kad* for *crawled*). Jeremy is occasionally correct in recalling orthographic patterns (e.g., the *ck* in *rock*), but he frequently produces inaccurate and unlikely letter combinations (e.g., *hasl* for *house*, *eher* for *here*). His written expression scores were reduced due to indecipherable spellings that obscured the meaning of his sentences.

Although Jeremy's standard scores in reading and spelling are low (SS 77, 78), they do not suggest the degree of weakness that his parents and teacher report or the difficulties he exhibited during testing. For example, even after looking at a word printed in a prompt for an item, he spelled it incorrectly. In contrast, the RPIs accurately reflect the real-world severity of his difficulties. When Jeremy's grade peers demonstrate 90% success on basic reading skills, Jeremy is predicted to be 9% successful. When his classmates understand 90% of what they read, Jeremy is expected to understand 22%. The descriptive labels of the RPIs (see Table 5) indicate that Jeremy's proficiency in basic reading skills and comprehension is very limited; the instructional implications indicate that grade-level reading tasks will be extremely difficult, which, in fact, they are. Decisions based solely on Jeremy's standard scores would underestimate the gravity of his need for specialized instruction in reading and spelling.

The evaluation further shows that Jeremy has many strengths in cognitive abilities and language. His standard scores and RPIs indicate he is in the average or high average range in logical reasoning for problem solving, visual-spatial thinking, memory for picture-word associations, general knowledge, and oral language. Test results also identify cognitive weaknesses, the importance of which is to ascertain the major factors contributing to Jeremy's academic difficulties and to inform the design of an effective instructional program for him.

Table 6 shows Jeremy's standard scores and RPIs for three WJ III cognitive ability clusters.

Table 6.
Jeremy's Standard Scores (SS) and Relative Proficiency Indexes (RPI) for Three WJ III Cognitive Ability Clusters with Descriptive Labels

CLUSTER/Test	SS	RPI	Proficiency	Implications
AUDITORY MEMORY SPAN	85	65/90	Limited	Very Difficult
SHORT-TERM MEMORY	88	66/90	Limited	Very Difficult
PERCEPTUAL SPEED	85	80/90	Limited to Average	Difficult

If Jeremy's standard scores are interpreted in a typical (and incorrect) manner as indications of ability level, it would appear that Jeremy has only a mild weakness in auditory memory and perceptual speed. After all, the lowest of his scores is well within the low average range and all are within a standard deviation of the mean. In many school districts that use a discrepancy procedure, these scores will not qualify him for special education services, and a pattern of strengths and weaknesses is not obvious. However, Jeremy's RPIs provide a more understandable explanation for the severity of his difficulty in learning reading and spelling skills (see Table 5). Regarding the ability to hold noncontextual information in short-term memory, Jeremy is predicted to retain 65% to 66% compared with his typical grade peers' ability to retain 90% (see Table 6). The instructional implications are that Jeremy will find tasks requiring this type of

memory ability very difficult. Additionally, Jeremy is somewhat limited in perceptual speed, lending support to error patterns indicating an additional problem in orthographic processing. In this case, the RPIs, not the standard scores, inform the examiner as to the specific cognitive weaknesses that contribute to Jeremy's significant difficulties in learning basic reading and spelling skills.

An additional point relates to the concept of "significance." Jeremy's school district uses 1.5 standard deviations as the criterion for a significant discrepancy between a student's predicted standard score and his or her actual score, after correcting for regression. The *Woodcock-Johnson III Normative Update Compuscore and Profiles Program* (Schrank & Woodcock, 2007) reports discrepancies (i.e., Intellectual Ability/Achievement, Oral Language/Achievement, and Predicted Achievement/Achievement) that show Jeremy meets this criterion on Basic Reading Skills but not on Written Expression. Yet his RPI of 69/90 indicates that his written expression skills are limited to average and that he will find grade-appropriate writing tasks difficult. Although the discrepancy based on peer-comparison scores is not significant by his school district's standards, the RPI describes a writing weakness that is educationally significant and is confirmed by his classroom writing samples. Additionally, the reported Intra-Cognitive Variations and GIA Standard/Cognitive Discrepancies show no significant cognitive weaknesses, whereas the RPIs describe weaknesses in memory abilities and, to a lesser extent, in perceptual speed that clearly have educational effects. The point is that evaluators cannot depend solely on discrepancies among peer-comparison scores for analysis of a person's abilities and the factors contributing to his or her learning difficulties. They must also consider the possible educational significance of an individual's quality of performance on a test, and this is best represented by the RPI.

Considerations for Using the RPI in View of the Individuals with Disabilities Education Act and in Developing Interventions

In describing evaluation procedures, the Individuals With Disabilities Education Act (IDEA 2004) regulations state that assessment tools must be technically sound (IDEA 2004, §300.304[b][3]) and provide relevant information that will be of direct assistance in determining the educational needs of the child (IDEA 2004, §300.304[c][7]). Additionally, in determining that a child has a specific learning disability (SLD), one factor that may be considered is the pattern of strengths and weaknesses relevant to SLD (IDEA 2004, §300.309[a][2][ii]) (U. S. Department of Education, 2006).

One of the most relevant aspects of information regarding a student's educational needs is his current proficiency in cognitive abilities, language abilities, and academic skills or knowledge. Can the student manage a task, or a set of tasks, that are used to assess a skill or ability at the same level of difficulty as his or her peers? Whether or not the student can do so and the *level* at which he or she can do so is critical information, irrespective of the standard scores. Using the RPI and peer-comparison scores, along with knowledge of the research regarding which cognitive abilities facilitate specific academic skills and those implicated in specific academic deficiencies, an evaluator can better understand or theorize the reasons for a student's failure to acquire certain skills or knowledge. Most importantly, this information can guide the educational team's selection or design of instructional methods that are most likely to be effective for the student.

Consider Jon and Andrea, both of whom have RPI scores in basic reading and spelling that indicate limited proficiency. Jon has weak perceptual speed but adequate working

memory and phonological awareness. Andrea has adequate processing speed but weak phonological awareness and lexical retrieval. Both will need instructional programs that are systematic, that are cumulative in reviewing and incorporating previously learned material at each step, and that ensure the student has mastered a skill or concept before moving on. Jon, however, is likely to need supplementary procedures to help him develop instantaneous recognition of letter patterns such as odd spellings (e.g., enough, height), syllable patterns, and morphemes. Andrea will need a strong emphasis on developing automaticity in phonemic awareness skills in conjunction with phonics and overlearning of new vocabulary for automatic retrieval. The standard scores (or percentile ranks) *may* indicate the cognitive and academic profiles that will lead the evaluator to this insight, but it is more likely that the RPIs, along with qualitative analysis, will do so. IDEA 2004 mandates the use of technically sound assessment instruments in conducting an evaluation, but it is, as always, incumbent on the evaluator and other qualified professionals to decide what information is the most relevant to assist in determining the educational needs of the child.

Interpreting Achievement and Growth

When selecting test scores to report to parents, teachers, and the individual who was evaluated, some scores are more easily explained than others. Peer-comparison scores are often explained in terms of the percentile rank (which can be visualized easily as the person's position in a line of people) rather than by standard scores. Percentile ranks or other peer-comparison scores, however, do not convey the information in which people are most interested—the person's proficiency in the trait being measured compared to what is expected for his or her age or grade. Consequently, the recipients of the report often ask, "What grade level is that?" Well-informed psychologists and other evaluators are aware of the problems that accompany the use of grade or age equivalents in interpreting an individual's test performance. Although many of the problems associated with the interpretation of these scores do not occur in the WJ III assessments, two issues should be considered: reporting a person's level of achievement and quantifying his or her growth in the trait assessed.

Age and grade equivalents on appropriately designed tests do represent a level of development in a skill, ability, or area of knowledge. If ninth-grade Johnny's *W* ability on Applied Problems is the same as that of the average sixth grader's *W* ability on Applied Problems, he is, in fact, performing most similarly to an average sixth grader—on the test. The statement describing his achievement in these terms, however, is likely to be too broad. "Johnny's ability to solve practical math problems is at the sixth-grade level" implies that all aspects of his problem-solving abilities are at the same level. In reality, Johnny may handle some areas of math application as well as the average sixth-grade student and other areas of math less well. His parents and teachers are likely to understand Johnny's situation more accurately if told, "Johnny's RPI in Applied Problems is 45/90. This means that he is likely to be about 45% successful when attempting to solve practical math problems that an average ninth grader can solve with 90% success."

The use of age and grade equivalents are more problematic when interpreting growth because they are not on equal-interval scales. The significance of any increase in a child's AE or GE score is dependent on the trait being measured and the amount of development that typically occurs during the interval of time since the last assessment. For example, 1 year's growth in basic reading skills in the primary grades indicates considerably more learning than 1 year's growth in middle school because most learning of these skills

occurs in the primary grades. It is both more meaningful and more accurate to say, "Since we last tested Johnny, his RPI has increased from 35/90 to 75/90. Whereas a year ago, he was likely to handle grade-level reading material with about 35% success, his current scores indicate that he'd be about 75% successful." Whether the student is in 2nd grade or 12th grade, the statement conveys the same meaning.

The *Woodcock-Johnson III Normative Update Compuscore and Profiles Program* (Schrank & Woodcock, 2007) provides a Parent Report, which includes brief descriptions of the reading, writing, and math tests administered and a graph depicting the child's level of proficiency in each area. This report uses the criterion-referenced interpretations of RPI scores (see Table 4). Alternatively, the evaluator may use the interactive Instructional Zone Profile Worksheet and Developmental Zone Worksheet (Mather & Jaffe, 2004) appended to this bulletin, based on RPI instructional implications.

Interpretation of Proficiency vs. Relative Standing in Clinical Research

The *Woodcock-Johnson III Normative Update Technical Manual* (McGrew, Schrank, & Woodcock, 2007a, pp. 78–85) presents and discusses descriptive statistics for groups of individuals representing a variety of clinical diagnoses. These include anxiety spectrum disorder, attention-deficit/hyperactivity disorder (ADHD), autistic spectrum disorder, depressive spectrum disorders, mental retardation, language disorders, mathematics disorders, reading disorders, and written expression disorders. The clinical samples represent two age levels: below age 19 and age 19 and above. Comparison statistics are presented for nonclinical groups at each age level. For each clinical group, scores are presented for various combinations of the following clusters: Comprehension-Knowledge, Long-Term Retrieval, Visual-Spatial Thinking, Auditory Processing, Fluid Reasoning, Short-Term Memory, Brief Reading, Brief Math, Brief Writing, Academic Knowledge, Listening Comprehension, and Oral Expression. The Brief clusters come from the *Woodcock-Johnson III Tests of Achievement Form C/Brief Battery* (Woodcock, Schrank, Mather, & McGrew, 2007b) and are reported in the *WJ III NU Compuscore and Profiles Program* (Schrank & Woodcock, 2007).

Presenting abilities in terms of standard scores and RPIs provides some interesting advantages. One benefit is the description of the cognitive, linguistic, and academic strengths and weaknesses of each clinical group. Another is the opportunity to compare which abilities are least proficient within a clinical group based on each metric. These issues are addressed in the following discussions of some of the clinical samples.

ADHD

For both the child and adolescent (below age 19) and the adult (age 19 and above) samples, all of the cognitive, oral language, and academic cluster standard scores are within the lower end of the average range (SS 90–100) with no significant discrepancies among them. The lowest scores for both groups are in Long-Term Retrieval and Processing Speed (SS 91–92). The RPIs, however, indicate that each age group has areas of relative weakness not represented by the standard scores. The younger group evidences limited-to-average proficiency in Short-Term Memory and in all three Brief academic clusters. The adult group also demonstrates limited-to-average proficiency in Short-Term Memory as well as Comprehension-Knowledge, Brief Reading, Brief Math, Academic Knowledge, and both language clusters. The finding of weak short-term memory in children, adolescents, and adults with ADHD, indicated by the RPIs but

not by the standard scores, is consistent with the results of other research (Marusiak & Janzen, 2005; Quinlan & Brown, 2003). The interpretation of research findings regarding cognitive, academic, and language weaknesses associated with ADHD might be different depending on whether the investigator considers only the standard scores or both the standard scores and the RPIs. Similarly, researchers' decisions regarding which abilities to investigate further within this clinical group might differ based on whether standard scores or RPIs were reported in previous research.

Autism Spectrum Disorders

In recent years, autism has received increased research attention. Within the younger clinical sample, the median standard score for Processing Speed is in the low average range. Other research has found more pronounced weaknesses on processing speed tests (Calhoun & Mayes, 2005; Mayes & Calhoun, 2007). The RPI results are in accordance with this research. For this sample, RPI results for academic achievement are particularly important. Standard scores for Brief Reading, Brief Math, and Brief Writing are in the low average range, which some would interpret as indicating mild difficulty in those skills. In contrast, the RPIs indicate limited proficiency (54/90 to 57/90), suggesting that tasks requiring these skills will be very difficult for individuals typical of this group.

Language Disorders

The RPIs in the child and adolescent language disorders group describe statistical correlations and, possibly, ramifications of this disorder. Subjects demonstrated limited to average proficiency in Short-Term Memory (74/90), Comprehension-Knowledge (74/90), and Brief Math (71/90) and limited proficiency in Fluid Reasoning (RPI 58/90), Listening Comprehension (65/90), and Brief Reading (58/90). They did not show the weaknesses that might be expected in Brief Writing (80/90) or Oral Expression (80/90).

Several of these results might be surprising. Individuals with diagnosed language disorders might be expected to demonstrate lower proficiency in tests of short-term memory and higher proficiency in tests using nonverbal stimuli, such as the Fluid Reasoning test. Unexpected findings should prompt the examiner to consider the following three possible factors.

1. *Do the standard score confidence bands of the narrow abilities within a cluster overlap?* Each test within a cluster assesses a different aspect of the broader ability represented by the cluster. For example, Short-Term Memory comprises a test of memory span (Memory for Words) and a test of working memory (Numbers Reversed). A person with a language disorder might be able to repeat a string of unrelated words but have more difficulty holding and rearranging a sequence of digits in memory. If the confidence bands of the two tests do not overlap, the individual's ability in memory span may be significantly different from that of working memory. In this case, the cluster score represents the broader aspect of memory in this person only on average and over a variety of tasks. A statement regarding only the cluster score would give incomplete and possibly misleading information regarding the person's memory abilities. The discrepancy between the narrower abilities presents important information for investigation regarding the aspect of memory that is deficient. In such a case, more testing needs to be done to further investigate the narrower abilities.
2. *What are the demands of the task?* The name of a cluster or test does not always reflect the task demands. Analysis of the task demands will help ascertain the reasons for unexpected results. For example, it may be surprising that

the language disorders group did not obtain a lower RPI in Brief Writing. In analyzing the task demands of the Writing Samples test, however, it is apparent that many of the skills a person with a language disorder would be expected to have difficulty with are not required. In most of the items, spelling, capitalization, and minor usage and grammatical errors are not penalized. Additionally, the person is asked only to write one sentence so that concept organization at a more complex level is not required.

3. *Are the weak abilities causes or consequences of the clinical condition?* One interpretation of the low scores in Fluid Reasoning is that a language disorder causes the individual to have difficulty forming new concepts or, as some psychologists have suggested, understanding the instructions. Another potentially important interpretation is that a weakness in concept formation causes the disordered development of language abilities. The direction of influence may go either or both ways (R. W. Woodcock, personal communication, June 4, 2007).

Investigation of the RPIs of any clinical sample may help professionals better understand the nature of these disabilities where standard scores may not show discrepancies. McGrew et al. (2007) note that within the statistics presented for the clinical samples, "a standard score of 94 . . . can be associated with RPIs as low as 62/90 (interpreted as limited proficiency) or as high as 88/90 (interpreted as average proficiency). Therefore, both the standard score and the RPI should be considered when evaluating the performance of an individual or a group" (p. 85).

Summary

The *W* score and relative proficiency index are two unique metrics available in the WJ III assessments. Because *W* scores are linked to normative data, and *W* abilities (describing the ability level of the person) and *W* difficulties (describing the difficulty level of the items) are on the same equal-interval scale, these values provide a mathematical basis for predicting a person's proficiency at any level of task difficulty. The *W* score and the RPI are also useful in interpreting and reporting the amount of change (gain or loss of proficiency) a person makes in any assessed trait over time. The RPI is particularly useful in indicating educationally significant difficulties that may not be obvious solely from peer-comparison scores. Consequently, psychologists and diagnosticians who are unaware of the value of proficiency scores may overlook important information regarding an individual's skills, abilities, and areas of knowledge. Hopefully, the information in this bulletin will lessen that possibility.

Appendix

Table A–1.
Instructional Zone Profile Worksheet—WJ III Tests of Achievement

Achievement Clusters/ Achievement Tests	Very Difficult (67/90 & below)	Difficult (67/90 to 82/90)	Appropriate (82/90 to 95/90)	Easy (95/90 to 98/90)	Very Easy (98/90 & above)
Oral Language (Std)					
Story Recall					
Understanding Direction					
Oral Language (Ext)					
Story Recall					
Understanding Directions					
Picture Vocabulary					
Oral Comprehension					
Oral Expression					
Story Recall					
Picture Vocabulary					
(Academic Knowledge)					
(General Information—COG)					
Listening Comprehension					
Understanding Directions					
Oral Comprehension					
Broad Reading					
Letter-Word Identification					
Reading Fluency					
Passage Comprehension					
Basic Reading					
Letter-Word Identification					
Word Attack					
Reading Comprehension					
Passage Comprehension					
Reading Vocabulary					

Table A–1, continued

Instructional Zone Profile Worksheet—WJ III Tests of Achievement

Achievement Clusters/ Achievement Tests	Very Difficult (67/90 & below)	Difficult (67/90 to 82/90)	Appropriate (82/90 to 95/90)	Easy (95/90 to 98/90)	Very Easy (98/90 & above)
Broad Written Language					
Spelling					
Writing Fluency					
Writing Samples					
Written Expression					
Writing Fluency					
Writing Samples					
Basic Writing Skills					
Spelling					
Editing					
(Punctuation & Capitalization)					
(Spelling of Sounds)					
Phoneme/Grapheme					
Word Attack					
Spelling of Sounds					
(Sound Awareness)					
Broad Math					
Calculation					
Math Fluency					
Applied Problems					
Math Calculation Skills					
Calculation					
Math Fluency					
Math Reasoning					
Applied Problems					
Quantitative Concepts					
Cross Academic Clusters					
Academic Fluency					
Reading Fluency					
Writing Fluency					
Math Fluency					

Table A–1, continued

Instructional Zone Profile Worksheet—WJ III Tests of Achievement

Academic Skills					
Letter-Word Identification					
Spelling					
Calculation					
Academic Applications					
Passage Comprehension					
Applied Problems					
Writing Samples					

Note. From *Woodcock-Johnson III: Reports, Recommendations, and Strategies* [CD] by N. Mather and L. E. Jaffe, 2004, New York: John Wiley & Sons. Copyright 2004 by John Wiley & Sons. Reprinted with permission.

Tests in parentheses are not included in the cluster but provide important information regarding the broad academic skill assessed by the cluster.

Table A–2.

Developmental Zone Worksheet—WJ III Tests of Cognitive Abilities and Diagnostic Supplement

Cognitive Factor/Clusters Cognitive Tests	**Very Difficult (67/90 & below)**	**Difficult (67/90 to 82/90)**	**Appropriate (82/90 to 95/90)**	**Easy (95/90 to 98/90)**	**Very Easy (98/90 & above)**
Comprehension-Knowledge (*Gc*)					
Verbal Comprehension					
General Information					
(Academic Knowledge–ACH)					
Long-Term Retrieval (*Glr*)					
Visual-Auditory Learning					
Retrieval Fluency					
Visual-Spatial Thinking (*Gv*)					
Spatial Relations					
Picture Recognition					
Visual-Spatial Thinking–3 (*Gv*) (DS)					
Spatial Relations					
Picture Recognition					
Visual Closure					
Visualization (*Gv*) (DS)					
Spatial Relations					
Block Rotation					

Table A–2, continued

Developmental Zone Worksheet—WJ III Tests of Cognitive Abilities and Diagnostic Supplement

Cognitive Factor/Clusters Cognitive Tests	**Very Difficult (67/90 & below)**	**Difficult (67/90 to 82/90)**	**Appropriate (82/90 to 95/90)**	**Easy (95/90 to 98/90)**	**Very Easy (98/90 & above)**
Auditory Processing (*Ga*)					
Sound Blending					
Auditory Attention					
Sound Discrimination (*Ga*) (DS)					
Sound Patterns–Voice					
Sound Patterns–Music					
Fluid Reasoning (*Gf*)					
Concept Formation					
Analysis-Synthesis					
Fluid Reasoning–3 (*Gf*) (DS)					
Concept Formation					
Analysis-Synthesis					
Number Matrices					
Numerical Reasoning (*Gq*) (DS)					
Number Series					
Number Matrices					
Processing Speed (*Gs*)					
Visual Matching (1 or 2)					
Decision Speed					
Perceptual Speed (*Gs*) (DS)					
Visual Matching (1 or 2)					
Cross Out					
Short-Term Memory (*Gsm*)					
Numbers Reversed					
Memory for Words					
Auditory Memory Span (*Gsm*) (DS)					
Memory for Words					
Memory for Sentences					

Table A–2, continued

Developmental Zone Worksheet—WJ III Tests of Cognitive Abilities and Diagnostic Supplement

Cognitive Factor/Clusters Cognitive Tests	**Very Difficult (67/90 & below)**	**Difficult (67/90 to 82/90)**	**Appropriate (82/90 to 95/90)**	**Easy (95/90 to 98/90)**	**Very Easy (98/90 & above)**
Phonemic Awareness (*Ga*)					
Sound Blending					
Incomplete Words					
(Sound Awareness–ACH)					
Working Memory (*Gsm*)					
Numbers Reversed					
Auditory Working Memory					
Associative Memory (*Glr*) (DS)					
Visual-Auditory Learning					
Memory for Names					
Associative Memory–Delayed (*Glr*) (DS)					
Visual-Auditory Learning–Delayed					
Memory for Names–Delayed					
Broad Attention					
Numbers Reversed (*Gsm*)					
Auditory Attention (*Ga*)					
Pair Cancellation (*Gs*)					
Auditory Working Memory (*Gsm*)					
Cognitive Fluency					
Retrieval Fluency (*Glr*)					
Decision Speed (*Gs*)					
Rapid Picture Naming (*Gs*)					
Executive Processes					
Concept Formation (*Gf*)					
Planning (*Gf*)					
Pair Cancellation (*Gs*)					

Note. From *Woodcock-Johnson III: Reports, Recommendations, and Strategies* [CD] by N. Mather and L. E. Jaffe, 2004, New York: John Wiley & Sons.

Tests in parentheses are not included in the cluster but provide important information regarding the broad cognitive ability assessed by the cluster.

*(DS) = Clusters that require administration of tests from the Diagnostic Supplement

References

Calhoun, S. L., & Mayes, S. D. (2005). Processing speed in children with clinical disorders. *Psychology in the Schools*, *42*, 333–343.

Marusiak, C. W., & Janzen, H. L. (2005). Assessing the working memory abilities of ADHD children using the Stanford-Binet Intelligence Scales, Fifth Edition. *Canadian Journal of School Psychology*, *20*, 84–97.

Mather, N., & Jaffe, L. E. (2002). *Woodcock-Johnson III: Reports, recommendations, and strategies*. New York: Wiley.

Mather, N., & Jaffe, L. E. (2004). *Woodcock-Johnson III: Reports, recommendations, and strategies* [CD]. New York: Wiley.

Mather, N., & Woodcock, R. W. (2001). Examiner's Manual. *Woodcock-Johnson III Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside Publishing.

Mayes, S. D., & Calhoun, S. L. (2007). Learning, attention, writing, and processing speed in typical children and children with ADHD, autism, anxiety, depression, and oppositional-defiant disorder, *Child Neuropsychology*, *13*, 469–493.

McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). Technical manual. *Woodcock-Johnson III Normative Update*. Rolling Meadows, IL: Riverside Publishing.

McGrew, K. S., & Woodcock, R. W. (1989). Technical manual. *Woodcock-Johnson Psycho-Educational Battery–Revised*. Rolling Meadows, IL: Riverside Publishing.

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

Quinlan, D. M., & Brown, T. E. (2003). Assessment of short-term verbal memory impairments in adolescents and adults with ADHD. *Journal of Attention Disorders*, *6*, 143–152.

Schrank, F. A., & Woodcock, R. W. (2002). Manual and checklists. *Report Writer for the WJ III*. Rolling Meadows, IL: Riverside Publishing.

Schrank, F. A., & Woodcock, R. W. (2007). *WJ III Normative Update Compuscore and Profiles Program* (Version 3.0) [Computer software]. Rolling Meadows, IL: Riverside Publishing.

U. S. Department of Education. 2006. (August 14). Assistance to states for the education of children with disabilities and preschool grants for children with disabilities; Final rule. Part II: 34 CFR Parts 300 and 301. *Federal Register* 71(156). Retrieved July 14, 2007, from http://edocket.access.gpo.gov/2006/pdf/06-6656.pdf

Woodcock, R. W. (1978). Development and standardization of the *Woodcock-Johnson Psycho-Educational Battery*. Rolling Meadows, IL: Riverside Publishing.

Woodcock, R. W. (1999). What can Rasch-based scores convey about a person's test performance? In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 105–127). Mahwah, NJ: 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). Circle Pines, MN: American Guidance Service.

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

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

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001c). *Woodcock-Johnson III Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside Publishing.

Woodcock, R. W., McGrew, K. S., Mather, N., & Schrank, F. A. (2003). *Woodcock-Johnson III Diagnostic Supplement to the Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside Publishing.

Woodcock, R. W., Schrank, F. A., Mather, N., & McGrew, K. S. (2007a). Technical manual. *Woodcock-Johnson III Normative Update.* Rolling Meadows, IL: Riverside Publishing.

Woodcock, R. W., Schrank, F. A., Mather, N., & McGrew, K. S. (2007b). *Woodcock-Johnson III Tests of Achievement Form C/Brief Battery.* Rolling Meadows, IL: Riverside Publishing.

Woodcock, R. W., Schrank, F. A., McGrew, K. S., & Mather, N. (2007). *Woodcock-Johnson III Normative Update.* Rolling Meadows, IL: Riverside Publishing.

RIVERSIDE


HOUGHTON MIFFLIN HARCOURT

3800 Golf Road, Suite 100
Rolling Meadows, IL 60008-4015

800.323.9540
www.woodcock-johnson.com

1423722