WJ IV: Putting all the pieces together
Lauren DeFrancesco, M.S.
National Account Executive
Houghton Mifflin Harcourt

Agenda:
- Overview of the different scores available with the WJ IV
  - What are they and how do we use them?
  - Practical Application
- Variation and Comparison Methods
  - How do we interpret these?
  - Practical Application
- Putting the Pieces Together
- WJ IV KEY Essentials

Review of Levels of Information Available

Levels of Data:

<table>
<thead>
<tr>
<th>Level</th>
<th>Qualitative, informal, error analysis</th>
<th>Useful for instructional planning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Test Session Observations Checklist</td>
<td>Useful for behavioral observations</td>
</tr>
<tr>
<td>Level 2</td>
<td>Level of Development</td>
<td>Age Equivalent</td>
</tr>
<tr>
<td></td>
<td>Level of Instruction</td>
<td>Grade Equivalent</td>
</tr>
<tr>
<td>Level 3</td>
<td>Level of Proficiency</td>
<td>Relative Proficiency Index, OALP</td>
</tr>
<tr>
<td></td>
<td>Easy to Difficult Range</td>
<td>Developmental/Instructional Zone</td>
</tr>
<tr>
<td>Level 4</td>
<td>Relative Standing in Group</td>
<td>Standard Scores</td>
</tr>
<tr>
<td></td>
<td>Rank Order</td>
<td>Percentile Ranks</td>
</tr>
<tr>
<td></td>
<td>Significantly high or low standing</td>
<td>Discrepancy PP, SD</td>
</tr>
</tbody>
</table>

Age or Grade-Based Norms

- Choose most appropriate reference group
- Grade norms: K-12, 2-year college, and 4-year college including first year of graduate school
- Age norms: 2-95+
- Use same reference group when comparing results from different tests (i.e., age to age, grade to grade)
- Generally grade norms are preferable in school-based settings; age norms in ungraded settings
- When examinee’s age and grade are not consistent, score results both ways.
- AE/GE are not impacted by choice of norm group

Levels OF DATA:
- The four levels are cumulative
- Each level builds on information from the previous level
- Level 1 = Qualitative and often used to support a hypothesis
- Levels 2-4 = Provide data options from which to select
Level 1

- Qualitative information (Criterion-Referenced):
  - How did the examinee behave?
  - How did they approach a task?
  - Predicting how an examinee may do in the classroom given a similar task
  - Interventions can be based from these

Level 1: Qualitative Information

Available for ACH Tests 1-11

Qualitative Observation

On Applied Problems, which of the following best characterizes the individual’s performance? (Mark only one response.)

- 1. Solved problems with no observed difficulties (good comprehension and analytical abilities)
- 2. Solved initial problems with no observed difficulty but demonstrated increasing difficulties solving the latter items (typical)
- 3. Appears to have limited understanding of grade or age-appropriate math application tasks
- 4. None of the above, not observed, or does not apply

Example at age 9 for Letter-Word Identification:

- 4% identified words rapidly and accurately (rating 1)
- 75% were rated as typical (rating 2)
- 13% identified initial items rapidly and accurately but had difficulty applying phoneme-grapheme relationships on more difficult items (rating 3)
- 7% had non-automatic word reading skills (rating 4)
- 1% did not apply phoneme-grapheme skills (rating 5)

Level 2

- Based on raw score
- Not affected by choice of age or grade norms
- Reflects age or grade level in norm sample at which median score is the same as the examinee’s raw score
- Abbreviated AE or GE
- Written with hyphen (AE) or period (GE)
  (AE: 10-4, GE: 6.8)
Level 2 Scores

**Grade Equivalent** (GE) scores reflect the level of task difficulty at which a student can perform and may be useful for instructional planning.

**Age Equivalent** (AE) scores reflect developmental level and may be useful in understanding the abilities of young children and may help with placement planning.

**W Scores**

Raw scores are converted into W scores

A transformation of Rasch ability scale

An intermediate step in the interpretation process

Provides equal-interval measurement

Centered on a value of 500 which is set to approximate the average performance of a 10-year-old (grade 5.0)

**W Scores to AE/GE**

If the median W score for students in the second month of fourth grade is 450, then a student with a W score of 450 would receive a grade equivalent score of 4.2.

If the median W score for individuals at age 11 years, 5 months was 510, then a person with a W score of 510 would receive an age equivalent score of 11-5.

Sample Descriptions of Level 2 Scores

On the reading comprehension task, 12 year old Lisa scored similarly to an average 6 year old.

The number of items Tom, a 7th grader, answered correctly on the math calculation task is comparable to the average student in early grade 4.

Level 3

- **Proficiency (Criterion-Referenced)**
  - Rasch difference scores
  - CALP Levels
  - RPI
  - Instructional or developmental zone
  - Proficiency on tasks of average difficulty for peers
  - Range of development or instruction
  - Indicates quality of performance
  - Helps monitor progress

**Criterion-referenced scores**

- A criterion-referenced scores are most appropriate when an educator wants to assess the specific skills or concepts a student has learned.
- There are “cut scores” that determine level of success.
- Tells us how they perform against a learning objective versus another student.

(The Glossary of Education Reform 4/30/14; Bright Hub 2012)
**Levels of Development**

- Equal Interval Scale
- Difficulty levels of items
- Ability level that represents success on a test
- Linking – Reference W

**Measurement Rules**

Allows us to monitor growth and determine the individual’s functional or developmental range.

**Developmental or Instructional Zones**

- Instructional zones (on ACH tests)
  - Indicates independent, instructional, and frustration levels on a task
  - Uses W scale to create range around examinee’s obtained score (+/- 10 W points)
- Profile Report in the platform

**Relative Proficiency Index (RPI)**

- Provides a criterion-referenced index of a person’s proficiency or functionality.
- Compares person’s proficiency to average age or grade mates.
- Predicts level of success on similar tasks.
- Shows actual distance from average.
- Based on W Diff

**Understanding RPIs**

- RPIs are expressed as a fraction with the denominator fixed at 90. The numerator indicates the examinee’s proficiency on that task.
- Ranges from 0/90 to 100/90.
- 90/90: Examinee has average proficiency on task.
Criterion-Referenced Interpretation of RPI Scores

<table>
<thead>
<tr>
<th>RPI</th>
<th>Instructional Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>96/90 to 100/90</td>
<td>Independent</td>
</tr>
<tr>
<td>76/90 to 95/90</td>
<td>Instructional</td>
</tr>
<tr>
<td>75/90 and below</td>
<td>Frustration</td>
</tr>
</tbody>
</table>

Interpretation of RPI Scores

<table>
<thead>
<tr>
<th>W Difference Values</th>
<th>Reported RPIs</th>
<th>Proficiency</th>
<th>Implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>+31 &amp; above</td>
<td>100/90</td>
<td>very advanced</td>
<td>extremely easy</td>
</tr>
<tr>
<td>+14 to +30</td>
<td>98/90 to 100/90</td>
<td>advanced</td>
<td>very easy</td>
</tr>
<tr>
<td>+7 to +13</td>
<td>95/90 to 98/90</td>
<td>average</td>
<td>easy</td>
</tr>
<tr>
<td>6 to +6</td>
<td>92/90 to 95/90</td>
<td>average</td>
<td>manageable</td>
</tr>
<tr>
<td>13 to -7</td>
<td>67/90 to 82/90</td>
<td>limited</td>
<td>difficult</td>
</tr>
<tr>
<td>30 to -14</td>
<td>24/90 to 67/90</td>
<td>limited</td>
<td>very difficult</td>
</tr>
<tr>
<td>50 to -31</td>
<td>3/90 to 24/90</td>
<td>very limited</td>
<td>extremely difficult</td>
</tr>
<tr>
<td>51 &amp; below</td>
<td>0/90 to 3/90</td>
<td>extremely limited</td>
<td>nearly impossible</td>
</tr>
</tbody>
</table>

If all zones are 20 W points wide (+/-10 W points around score), why are some wider than others?

Different developmental curves exist for each measured skill or ability.
- wide zone – less developmental change (COG and Language)
- narrow zone – more developmental change

Compare the curves for BRS and Oral Language

Sam’s RPI of 21/90 on the Phoneme/Grapheme cluster indicates that on similar tasks, in which the average fourth-grade student would demonstrate 90% proficiency, Sam would demonstrate 21% proficiency. Sam’s knowledge of phoneme-grapheme correspondence and spelling patterns is very limited.
• Scores are derived differently.
  • SS uses Standard Deviation
  • RPI does not use SD

• Abilities develop differently.
  • People are more variable on some tasks.

*RPI uses average W for age or grade on task as beginning reference.

**RPI is determined by how many W points above or below that reference W the individual’s score falls.

(Reminder: The GE and AE are also calculated this way)

- Scores are derived differently.
  - SS uses Standard Deviation
  - RPI does not use SD

- Abilities develop differently.
  - People are more variable on some tasks.

While Nicholas’s standard score on the Mathematics Reasoning cluster was within the average range for seventh-grade students overall, his RPI (45/90) indicates that he will have considerably more difficulty than most of his same grade-peers in math problem solving.

Levels

• Relative Standing in a Group (Norm-referenced)
  - Standard Score
  - Percentile Rank

• Communication of an examinee’s position among peers
• Relative position
• Most commonly used to make decisions

Norm referenced scores

• The intention is to rank an examinee against another examinee
• Bell curve is forced, creating percentiles and standard deviations
• Used to make comparisons to other students
• Sensitive to slight raw score changes (one more question right or one more question wrong)
• Easy to use
• Provide a quick snapshot

(The Glossary of Education Reform 4/30/14)
**Standard score**

- Mean of 100, SD of 15
- Range from 0-200+
- More difficult for parents and other nonprofessionals to understand on their own.
- Verbal labels are provided to help provide clarification as to describing test results (i.e., exceptionally superior, very high/supior, average, etc.)

**Percentile rank**

- Scale from 1 to 99
- Describes performance relative to a specific age- or grade-level segment in the norm sample.

  - Example: Martha’s percentile rank of 99.5 on the Basic Math Skills cluster indicates that only five out of 1,000 students would have a score as high or higher.
  - Example: Martha’s percentile of 1 on the Basic Writing Skills cluster indicates that only one out of 100 third-grade students would obtain a score as low or lower than Martha.

**Reviewing the Normal Curve**

<table>
<thead>
<tr>
<th>Percentile Rank</th>
<th>Standard Scores</th>
<th>WJ IV Descriptive Labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>131 and above</td>
<td>98 to &gt;99.9</td>
<td>Very Superior</td>
</tr>
<tr>
<td>121 to 130</td>
<td>92 to 97</td>
<td>Superior</td>
</tr>
<tr>
<td>111 to 120</td>
<td>76 to 91</td>
<td>High Average</td>
</tr>
<tr>
<td>90 to 110</td>
<td>25 to 75</td>
<td>Average</td>
</tr>
<tr>
<td>80 to 89</td>
<td>9 to 24</td>
<td>Low Average</td>
</tr>
<tr>
<td>70 to 79</td>
<td>3 to 8</td>
<td>Low</td>
</tr>
<tr>
<td>69 and below</td>
<td>&lt; 0.1 to 2</td>
<td>Very Low</td>
</tr>
</tbody>
</table>

**Standard Scores vs Relative Proficiency Index (RPI)**

Understanding the RPI Scores and Their Relation to Standard Scores

Percentiles reflect relative standing

Standard scores also reflect relative standing in a group
But not distance from “average” performance

Notice that the percentiles don’t change since relative standing remains the same

The RPI answers: “How far from average proficiency is a person’s performance?”

Notice that the percentiles don’t change since relative standing remains the same

Monitoring Progress

- Norm-referenced tests like the WJ IV are not designed to be used as frequent progress-monitoring tools
- Level 3 information in the WJ IV helps document progress over time
- W scores are best metric for documenting change
- Relative standing may not change or decrease (SS/PR) even if raw score increases

Documenting Progress

- W scores are sensitive to even minimal changes in performance.
- W scores are the best metric for monitoring progress.
- SS/PR may decrease even when raw scores increase
  - the individual’s relative standing declines when gains are less than average

Danny
SS versus RPI

Danny

Woodcock-Johnson IV Tests of Achievement Form A (Norms based on age 9-11)

CLUSTER/Test W GE RPI SS (68% Band) PR

READING COMPREHENSION
Passage Comprehension 490 3.5 81/90 94 (90-99) 35
Reading Recall 499 5.5 93/90 105 (101-108) 63

BASIC READING SKILLS
Letter-Word Identification 478 3.0 53/90 89 (86-92) 23
Word Attack 479 2.3 60/90 85 (80-90) 16

READING FLUENCY
Oral Reading 477 2.0 49/90 83 (80-87) 13
Sentence Reading Fluency 464 2.9 22/90 88 (83-92) 21

READABILITY
Passage Comprehension 490 3.5 81/90 94 (90-99) 35
Reading Recall 499 5.5 93/90 105 (101-108) 63
Reading Vocabulary 487 3.1 76/90 90 (85-95) 25

READABILITY (Ext)
Passage Comprehension 490 3.5 81/90 94 (90-99) 35
Reading Recall 499 5.5 93/90 105 (101-108) 63

Monitoring Progress

- Norm-referenced tests like the WJ IV are not designed to be used as frequent progress-monitoring tools
- Level 3 information in the WJ IV helps document progress over time
- W scores are best metric for documenting change
- Relative standing may not change or decrease (SS/PR) even if raw score increases
W Score versus SS

Passage Comprehension

<table>
<thead>
<tr>
<th>Age</th>
<th>W Ability</th>
<th>Increase</th>
<th>SS</th>
<th>%tile</th>
</tr>
</thead>
<tbody>
<tr>
<td>8-3</td>
<td>462</td>
<td>-</td>
<td>84</td>
<td>14</td>
</tr>
<tr>
<td>11-2</td>
<td>488</td>
<td>+26</td>
<td>84</td>
<td>14</td>
</tr>
</tbody>
</table>

Limited progress in reading and spelling

Intense tutoring during 1st grade

Entered 2nd grade unable to sound out 3-letter words

Reading with dad every night. “We 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.”

Claiming sick before school and pleading to stay home

Parents referred for evaluation

<table>
<thead>
<tr>
<th>CLUSTER/Test</th>
<th>SS</th>
<th>RPI</th>
<th>Proficiency</th>
<th>Instructional Implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>BASIC READING SKILLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passage Comprehension</td>
<td>83</td>
<td>12/90</td>
<td>Very limited</td>
<td>Extremely difficult</td>
</tr>
<tr>
<td>Spelling</td>
<td>80</td>
<td>26/90</td>
<td>Limited</td>
<td>Very difficult</td>
</tr>
</tbody>
</table>

When sounding out 2-3 letter words, he usually produced the correct sounds individually, but was unable to retain their sounds and sequence when blending them (e.g., nap became “pen”).

Didn’t recognize common sight words (e.g., here, they)

Spelling mainly represented first and last sounds (e.g., kad/crawled)

Produced inaccurate/unlikely letter combinations (e.g., hast/house, ether/there)

• Know the purpose of each score
• Each score tells you something different
• Remember, the scores from different levels are not interchangeable
• To get the most out of your evaluations, include all levels of information

Contact information:

Lauren DeFrancesco, M.S.
National Account Executive- Clinical Products
Northeast Area

Lauren.defrancesco@hmhco.com

317.364.8432