

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



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:

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

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

Test Session Observations Checklist

Check only one category for each item.

Level of conversational proficiency

- 1. Very advanced
- 2. Advanced
- 3. Typical for age/grade
- 4. Limited
- 5. Very limited

Self-confidence

- 1. Appeared confident and self-assured
- 2. Appeared at ease and comfortable (typical for age/grade)
- 3. Appeared tense or worried at times
- 4. Appeared overly anxious

Level of cooperation

- 1. Exceptionally cooperative throughout the examination
- 2. Cooperative (typical for age/grade)
- 3. Uncooperative at times
- 4. Uncooperative throughout the examination

Care in responding

- 1. Very slow and hesitant in responding
- 2. Slow and careful in responding
- 3. Prompt but careful in responding (typical for age/grade)
- 4. At times responded too quickly
- 5. Impulsive and careless in responding

Level of activity

- 1. Seemed lethargic
- 2. Typical for age/grade
- 3. Appeared fidgety or restless at times
- 4. Overly active for age/grade resulted in difficulty attending to tasks

Response to difficult tasks

- 1. Noticeably increased level of effort for difficult tasks
- 2. Generally persisted with difficult tasks (typical for age/grade)
- 3. Attempted but gave up easily
- 4. Would not try difficult tasks at all

Attention and concentration

- 1. Unusually absorbed by the tasks
- 2. Attentive to the tasks (typical for age/grade)
- 3. Distracted often
- 4. Consistently inattentive and distracted

LEVEL 1: QUALITATIVE INFORMATION

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. Appeared to have limited understanding of grade- or age-appropriate math application tasks
- 4. None of the above, not observed, or does not apply

Available for ACH Tests 1-11

Qualitative Observation

On Spelling, which of the following best describes the automaticity of the individual's performance? (Mark only one response.)

- 1. Spelled words easily and accurately
- 2. Spelled initial items easily and accurately; spelling of latter items reflected a need for further skill development (typical)
- 3. Spelled words in a laborious manner (nonautomatic)
- 4. None of the above, not observed, or does not apply

QUALITATIVE OBSERVATIONS

NEW!

- Data collected in standardization to determine percentage of **age** mates assigned each rating
- Use to determine how typical or atypical examinee's performance is on task
- Tables located in Chapter 5 of the WJ IV ACH Examiner's Manual
- Data available for 9 of the 10 tests with Qualitative Observation Checklists (*Word Attack does not have data*)
- Test 8: Oral Reading offers a Qualitative Observation Tally (quantifies the types of errors made)

DATA FOR QUALITATIVE OBSERVATIONS

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

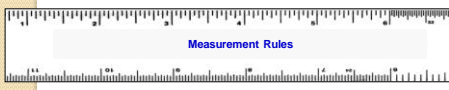
- A criterion-referenced scores are most appropriate when an educator wants to assess the specific skills or concepts a student has learned.
- There are “cut scores” that determine level of success.
- Tells us how they perform against a learning objective versus another student.

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

LEVELS OF DEVELOPMENT W-scale




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



We can see where on the ruler each examinee's performance is located and how far it is from the average performance for their age or grade.

**Age 10-0
Grade 5.0**

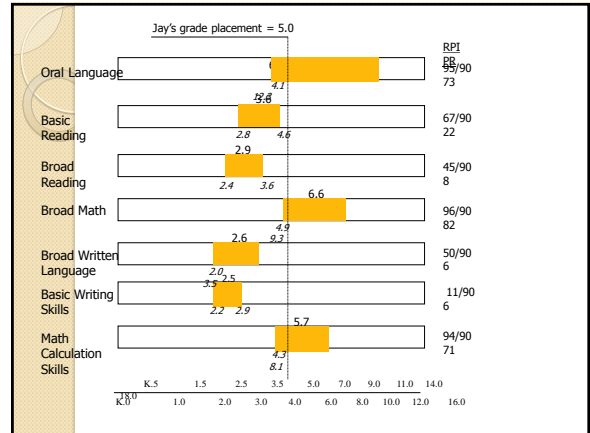


Examinee, 5.0

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.

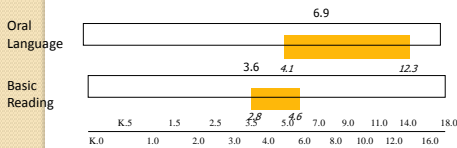
Criterion-Referenced Interpretation of RPI Scores

RPI	Instructional Level
96/90 to 100/90	Independent
76/90 to 95/90	Instructional
75/90 and below	Frustration

Interpretation of RPI Scores

W Difference Values	Reported RPIs	Proficiency	Implications
+31 & 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	average to advanced	easy
-6 to +6	82/90 to 95/90	average	manageable
-13 to -7	67/90 to 82/90	limited to average	difficult
-30 to -14	24/90 to 67/90	limited	very difficult
-50 to -31	3/90 to 24/90	very limited	extremely difficult
-51 & below	0/90 to 3/90	extremely limited	nearly impossible

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

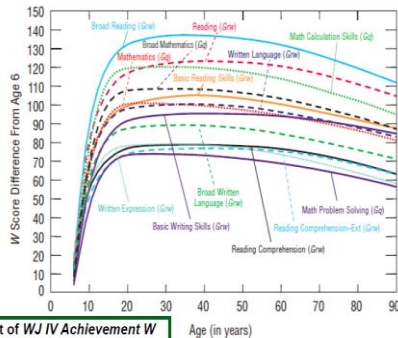
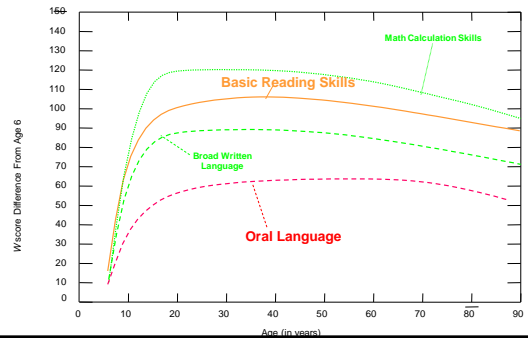


Different developmental curves exist for each measured skill or ability.

- wide zone – less developmental change (COG and Language)
- narrow zone – more developmental change

DEVELOPMENTAL CURVES FOR SELECTED ACHIEVEMENT CLUSTERS AND ORAL LANGUAGE

Compare the curves for BRS and Oral Language



Plot of WJ IV Achievement W score difference curves for clusters by age.

Houghton Mifflin Harcourt

HOW DO WE COMMUNICATE RPI IN A REPORT?

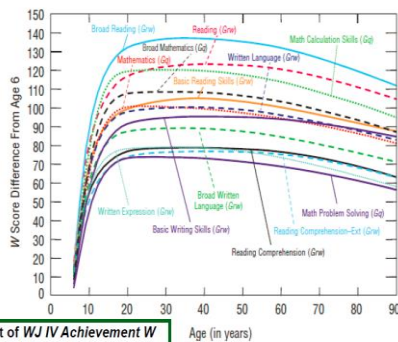
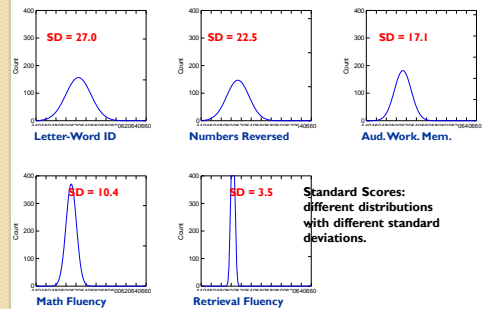
Sam's RPI of 21/90 on the Phoneme/Grapheme cluster indicates that on similar tasks, in which the average fourth-grade student would demonstrate 90% proficiency, Sam would demonstrate 21% proficiency. Sam's knowledge of phoneme-grapheme correspondence and spelling patterns is **very limited**.

- 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)

DENSITY PLOTS OF SELECT WJ TESTS FOR NORM SUBJECTS AGES 11-14: STANDARD DEVIATIONS



HOW DO WE COMMUNICATE RPI IN A REPORT?

- 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.

Level 4

- 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)

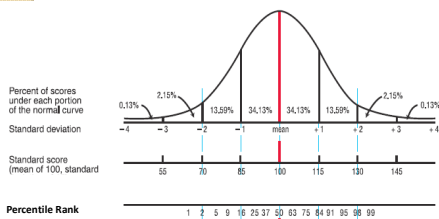
Standard score

- 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/superior, 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



Standard Scores	Percentile Rank	WJ IV Descriptive Labels	NOTES:
131 and above	98 to >99.9	Very Superior	Different tests use different ranges and labels.
121 to 130	92 to 97	Superior	
111 to 120	76 to 91	High Average	
90 to 110	25 to 75	Average	85-115 is "average" on some tests.
80 to 89	9 to 24	Low Average	
70 to 79	3 to 8	Low	
69 and below	< 0.1 to 2	Very Low	

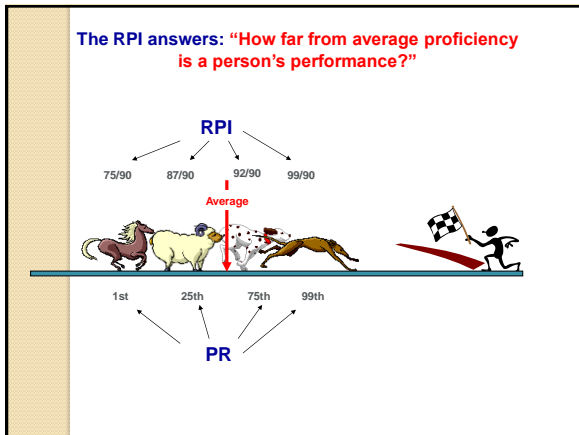
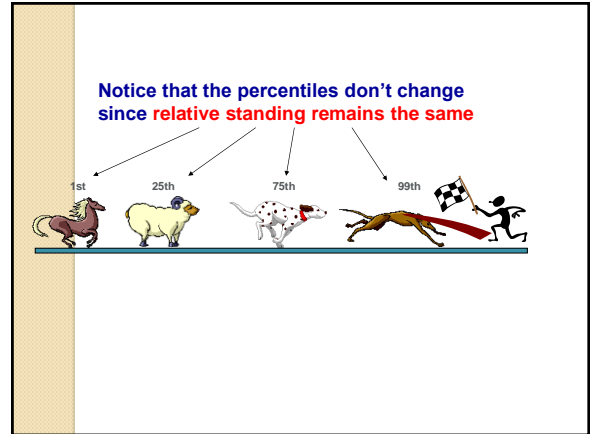
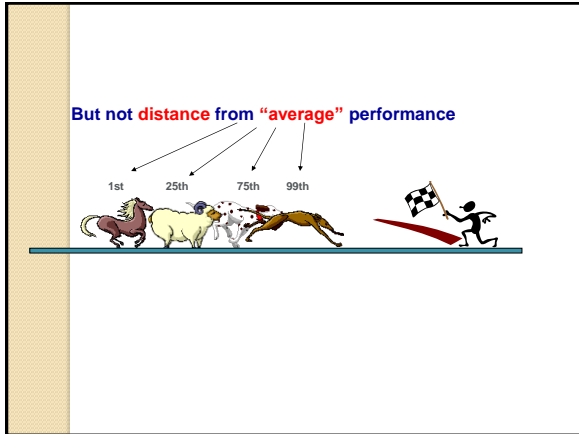
STANDARD SCORES VS RELATIVE PROFICIENCY INDEX (RPI)

Understanding the RPI Scores and Their Relation to Standard Scores

RPI's versus Percentile Ranks/Standard Scores

Percentiles reflect relative standing

Standard scores also reflect relative standing in a group



Danny SS versus RPI

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

CLUSTER/TEST	W	SS	PR	SS (80-90)	PR
READING	464	3.1	69/90	80 (80-90)	26
Letter-Word Identification	478	3.0	53/90	89 (86-92)	23
Passage Comprehension	490	3.5	81/90	94 (90-99)	35
BROAD READING	477	3.0	52/90	89 (86-92)	23
Letter-Word Identification	478	3.0	53/90	89 (86-92)	23
Passage Comprehension	490	3.5	81/90	94 (90-99)	35
Sentence Reading Fluency	464	2.9	22/90	88 (83-92)	21
BASIC READING SKILLS	478	2.7	56/90	87 (85-90)	19
Letter-Word Identification	478	3.0	53/90	89 (86-92)	23
Word Attack	479	2.3	60/90	85 (80-90)	16
READING COMPREHENSION	495	4.1	89/90	98 (95-101)	45
Passage Comprehension	490	3.5	81/90	94 (90-99)	35
Reading Recall	499	5.5	93/90	105 (101-108)	63
READING COMP (Ed)	482	3.7	85/90	95 (92-98)	37
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	70/90	90 (85-95)	25
READING FLUENCY	470	2.7	34/90	85 (82-89)	16
Oral Reading	477	2.0	49/90	83 (80-87)	13
Sentence Reading Fluency	464	2.9	22/90	88 (83-92)	21
READING RATE	455	2.8	18/90	87 (83-90)	19
Sentence Reading Fluency	464	2.9	22/90	88 (83-92)	21
Word Reading Fluency	446	2.7	12/90	85 (80-90)	16

- ### 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

W Score versus SS

Passage Comprehension

Age	W Ability	Increase	SS	%tile
8-3	462	-	84	14
11-2	488	+26	84	14

JEREMY, AGE 7-9

- 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

CLUSTER/Test	SS	RPI	Proficiency	Instructional Implications
BASIC READING SKILLS	81	3/90	Negligible	Impossible
Passage Comprehension	83	12/90	Very limited	Extremely difficult
Spelling	80	26/90	Limited	Very difficult

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 (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., hasl/house, eher/here)

Don't Forget...

- **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**

QUESTIONS?...

 Houghton
Mifflin
Harcourt

Contact information:

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

Lauren.defrancesco@hnhco.com

317.364.8432