

## Scores and What They Mean

Participant's Handout

**NOTE: Please complete the Pre-Test for the Scores Module before continuing.**

### ✓ Warm-Up: Thinking About Scores

1. What score do you use most frequently when reporting results?
2. Name some of the other types of scores available on tests you use.
3. Why do tests offer more than one type of score?
4. How comfortable are you when explaining various scores to others?

### Scores and Our Role

- ✓ Read this quote and paraphrase. What is the main point?

"Group-statistic based interpretations provide the "best jumping off points for interpretations of tests." But, individuals being tested can change the nature of interpretation (approach tasks differently, inflate specificity, reduce influence of ability being measured). This is part of the whole "intelligent" testing philosophy and my belief that "we (you) are the instrument." It is the job of a good clinician to know when the interpretation of a test may need to shift slightly away from the group-based most likely hypotheses. It is what we are trained to do..."

(Kevin McGrew, 2004)

### Levels of Interpretive Information

Level 1: Qualitative, informal, error analysis. observations	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	Standard Scores Percentile Ranks

# Scores and What They Mean

Participant's Handout

## Level 1: Importance of Qualitative Information

- **Observe and analyze behaviors:**  
Look for individual differences in information processing and learning style.
- **Validate interpretation of individual's test performance:**  
Use examples of observable behaviors
- **Analyze response processes:**  
Aspect of test validity (*Standards for Educational and Psychological Testing, 1999*).
- **Infer processing strengths and weaknesses:**  
Compare an individual's performance on one ability to performance on another.  
(*Lohman, 2001*)

### Exercise to Increase Your Use of Qualitative Information

- **Determine the task and response demands: (see slide)**
  - ✓ Task Demands: \_\_\_\_\_
  - ✓ Response Demands: \_\_\_\_\_
- **Analyze errors from an instructional perspective (see slide)**
  - ✓ Types of errors: \_\_\_\_\_
  - ✓ Instructional needs: \_\_\_\_\_
- **Observations made during testing:**
  - hesitant, long delay between words
  - did not say words quickly and automatically
  - tried to sound words out
  - errors were typically real words
  - rubbed eyes
  - stated "reading is hard."
- **What are the instructional implications you can derive from all of this information?** (task & response demands, error analysis, and observations)

✓ **Take a moment to list them now.**

---

---

---

---

## Scores and What They Mean

Participant's Handout

### Level 2: Age and Grade Equivalents

- Based on raw score
- Not effected by choice of age or grade norms
- Reflects age or grade level in norming at which average score is the same as the examinee's raw score
- Abbreviated AE or GE
- Written with hyphen or period (10-4, 6.8)

#### Sample Descriptions of Level 2 Scores:

- On the phonemic awareness 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.

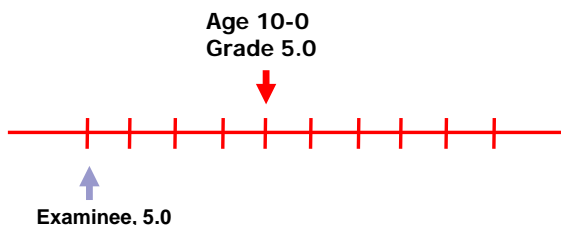
#### ✓ Write descriptions for the following scores:

- Jon, 5th grader, GE of 2.5 on word recognition task

- April, 5 years old, AE of 8-1 on fine motor task

### Level 3: Proficiency, Growth, Instructional Ranges

- Criterion-referenced information
- Indicates the quality of performance
- Helps monitor progress
- Indicates the range of development or instruction (independent to frustration)
- Types of Level 3 Scores: w scores, RPI, instructional or developmental ranges, change sensitive scores, growth scores, growth scale values



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

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

## Scores and What They Mean

Participant's Handout

### Relative Proficiency Index (RPI) *(only found on Woodcock tests)*

- 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

RPIs are expressed as a fraction with the denominator fixed at 90. The numerator indicates the examinee's proficiency on that task and can range from 0-100.

90/90: Examinee has average proficiency on task.

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

For a detailed chart of RPIs and descriptive labels, see the Examiner Manual for the Woodcock test you may be using.

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.

### Instructional Ranges

- Often seen in informal reading inventories
- Usually indicate Independent, instructional, frustration levels on a task
- Provided on some achievement tests (e.g., WJ III ACH)

✓ What is the purpose of an instructional range? \_\_\_\_\_

---

### Sample Descriptions of Level 3 Scores:

- Julie's RPI of 5/90 on spelling indicates she has very limited proficiency compared to average grade mates.
- Nick is making grade-appropriate progress in vocabulary as evidenced by his Growth Scale Value (GSV) score of 171, average for 5th grade.
- Karen will find decoding tasks easy at a beginning 3rd grade level, but difficult at a mid-4th grade level.

# Scores and What They Mean

## Participant's Handout

### ✓ Write descriptions for the following scores:

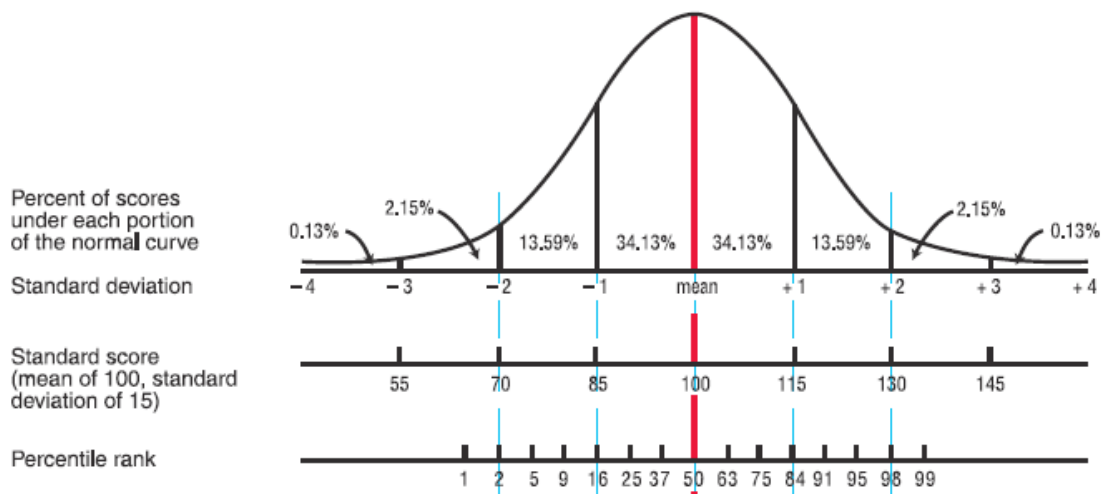
- Juan, 8th grade, RPI=45/90 on written expression

- Lena, 5th grade, instructional range on reading comprehension is 2.5 to 3.8.

### Level 4: Peer Comparison Scores

- Compares examinee to age or grade peers
- Standard Scores (equal interval)
  - Describes performance relative to the average performance of the comparison group.
  - Examples:  $M=100, SD=15$  or  $M=10, SD=3$
- Percentile Ranks (not equal interval)
  - Describes performance as relative standing in the comparison group on a scale of 1 to 99.
  - Indicates the percentage of comparison group who had scores the same as or lower.

### Reviewing the Normal Curve



### Statistically Significant Differences

3 tests:

- Are the scores significantly different? (not chance variation)
- Is the difference also unusual? (base rates, discrepancy PR)
- Educational implications?

## Scores and What They Mean

Participant's Handout

### Confidence Bands and Confidence Levels

- Increases “confidence” that the true score falls within an identified range.
- Uses the standard error of measure (SEM) around the obtained standard score to create the band or interval.
- Commonly available confidence levels are: 68%, 80%, 90%, and 95%. The higher the confidence level, the wider the band.

Obtained score +/- 1 SEM = 68% level of confidence

Obtained score +/-2 SEMs = 95% level of confidence

If the obtained score is 74 and the SEM is +/-3, then the range will be:

71-77 at the 68% level (+/-1 SEM) +/-3

68-80 at the 95% level (+/-2 SEM) +/-6

Also used to look for statistically significant differences between test scores.

### Mental Retardation: Eligibility Criteria

....has been determined to have significantly sub-average intellectual functioning as measured by a standardized, individually administered test of cognitive ability in which the overall test score is at least two standard deviations below the mean, when taking into consideration the standard error of measurement of the tests;

AND

Concurrently exhibits deficits in at least two of the following areas of adaptive behavior: communication, self-care, home living, social/interpersonal skills, use of community resources, self-direction, functional academic skills, work, leisure, health, and safety.

### When You Might Need to Consider the SEM

- Student has required deficits in adaptive behavior
- Obtained full-scale IQ >70 (possibly in the 71-73 range)
- Consider one SEM for test – For example, IQ=71 (+/- 3) = range of 68-74

**Conclusion:** \_\_\_\_\_

### MR Eligibility

- The required deficits in adaptive behavior must be present
- If the global intelligence score is already at 70 or lower, then the SEM doesn't matter
- Use the SEMs specified for the test you use (There is no “universal” SEM)
- Use only one SEM (68% level of confidence)
- If the test's computer scoring program does not provide a 68% level of confidence, you must look up the size of one standard error of measure in the manuals for the test you use.

# Scores and What They Mean

Participant's Handout

## Sample Descriptions of Level 4 Scores:

- Only 2% of Betsy's age mates scored higher than she did on rapid word reading (PR=98).
- Less than 1% of grade mates, scored as low or lower than Bret on spelling (PR=.5).
- Compared to other 6th graders, Jesse's performance in reading was in the low average to average range (SS=88-96).

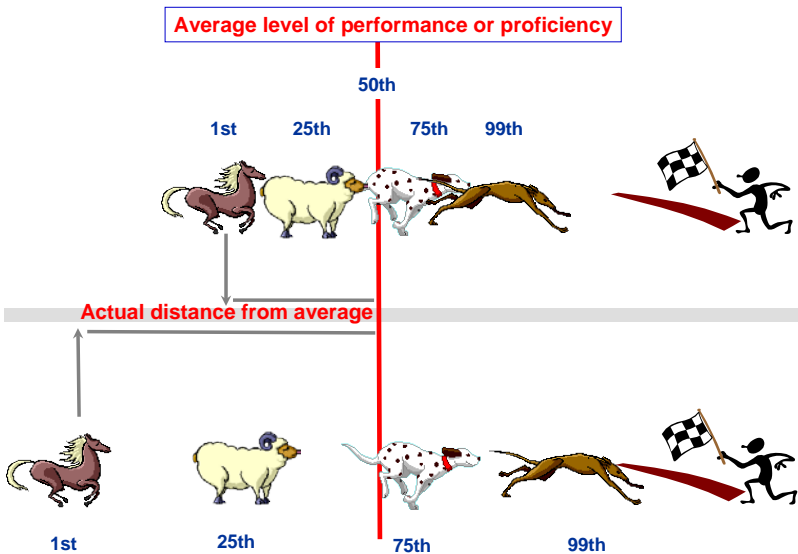
### ✓ Write descriptions for the following scores:

- Manuel, 4th grade, SS=142 in math reasoning

- Lacy, 2nd grade, SS=66-74 (68% confidence) in word reading

- Josh, 9th grade, PR=25 in calculation

## Making Use of Other Scores



Sprint analogy=Narrow distribution. Even though horse is at PR 1, it really isn't that far from average.

Marathon analogy = Wide distribution. The horse is still at PR 1 but is now further from average.

Standard Scores/Percentile Ranks describe relative standing within the norm group of the test you are using. The underlying distribution (wide or narrow) reflects how variable the age or grade group was on the task.

## Scores and What They Mean

### Participant's Handout

Which statement is most helpful for instructional planning?

1. The student has a standard score of 80 in reading comprehension.
2. The student finds reading comprehension tasks easy at the beginning third grade level and difficult at the end-fourth grade level.
3. On grade level tasks, this student has limited proficiency in reading comprehension. He will have 3 percent success when average grade mates have 90% success (RPI=3/90).
4. Four percent of grade mates scored this low or lower in reading comprehension.
5. In reading comprehension, this sixth grade student had the same number correct as the average student in grade 3.5.

### Example Using All Scores: 5<sup>th</sup> Grade Student

Norm-Referenced Information:

Reading Comprehension, SS=90

Word Reading, SS= 91

Criterion-Referenced Information:

Reading Comprehension, RPI = 74/90

Word Reading, RPI=61/90

Oral reading fluency, 50 wcpm (138 is benchmark)

Developmental/Instructional Information:

Reading Comprehension, Instructional Zone: 2.5 to 4.9

Word Reading, Instructional Zone: 2.9 to 4.3

✓ What else do we need to know?

---

---

Qualitative Information:

- Makes guesses based on visual characteristics
- Lacks word knowledge
- Knows beginning sounds
- Does not apply phonological analysis
- No apparent strategies for comprehension

## Scores and What They Mean

Participant's Handout

### Tricky Score Issues

#1 There are times when a composite does not seem to “hang” with the subtest scores. The composite seems too high or too low.

- Aren't composites an average of the component subtests?
- Why does this happen with composites?

✓ Explanation: \_\_\_\_\_

\_\_\_\_\_

#2 What should I do when the subtests within a cluster or composite are very different from one another?

- Can I still use the cluster/composite score?
- What should I do?

✓ Explanation: \_\_\_\_\_

\_\_\_\_\_

#3 When re-evaluating a student, her standard scores went down. I know she has made progress. What's going on?

- Why didn't the standard score go up?
- Can I use standard scores to monitor progress?
- What can I do to document progress?

✓ Explanation: \_\_\_\_\_

\_\_\_\_\_

“Tests do not think for themselves, nor do they directly communicate with patients. Like a stethoscope, a blood pressure gauge, or an MRI scan, a psychological test is a dumb tool, and the worth of the tool cannot be separated from the sophistication of the clinician who draws inferences from it and then communicates with patients and professionals.”

*Meyer et al. (FEB 2001). Psychological testing and psychological assessment. American Psychologist.*

**NOTE: Please complete the Post-Test for the Scores Module. Compare the results from your Pre- and Post-Tests.**