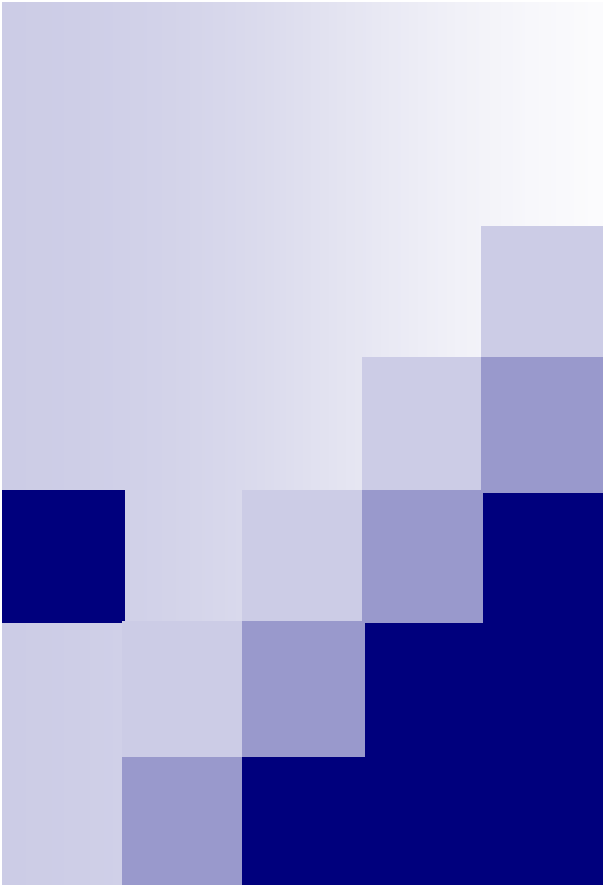


Scores Module Pre-Test



*Please complete the Pre-Test
before continuing.*



Scores and What They Mean



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?**



"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 S. 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



Level 1: Qualitative Information

- **Observe and analyze behaviors**
- **Validate interpretation of individual's test performance**
- **Analyze task demands and response processes**
- **Infer processing strengths and weaknesses**
- **Analyze errors**



Exercise to Increase Your Use of Qualitative Information

- **Analysis of task and response demands**
- **Error analysis**
- **Observations/comments**



Determine the task and response demands:

Examiner directions: Point to each word and say, “What word is this?”

Item scoring: Credit is given if the examinee says the word as a connected whole. No penalty for earlier attempts to sound out the word as long as final attempt is a smooth pronunciation.

Examinee Page:

on

was

it

web

coming

brother

point

eleven



Task Demands

Task requires reading real words. Does not require knowing meaning.

Response Demands

- **Response requires reading the word orally.**
- **No penalty for time or self-corrections.**
- **Word must be said altogether as a whole word, not in parts.**



Analyze these errors from an instructional perspective:

no for on

if for it

co-ming for coming

pont for point

saw for was

wed for web

bother for brother

even for eleven



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.



Instructional Implications

- Visual aspects – orthography (vision?)
- Vowels, rules for long/short
- Check phoneme/grapheme knowledge
- Oral vocabulary ok (check)
- Appears to need specific instruction
- Needs extended time to complete assigned readings
- May need shorter assignments
- Needs scaffolding



Level 2: Age or 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)**



Level 2: Age or Grade Equivalents

Sample Descriptions

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.



Level 2: Age or Grade Equivalents

Write descriptions for the following scores:

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

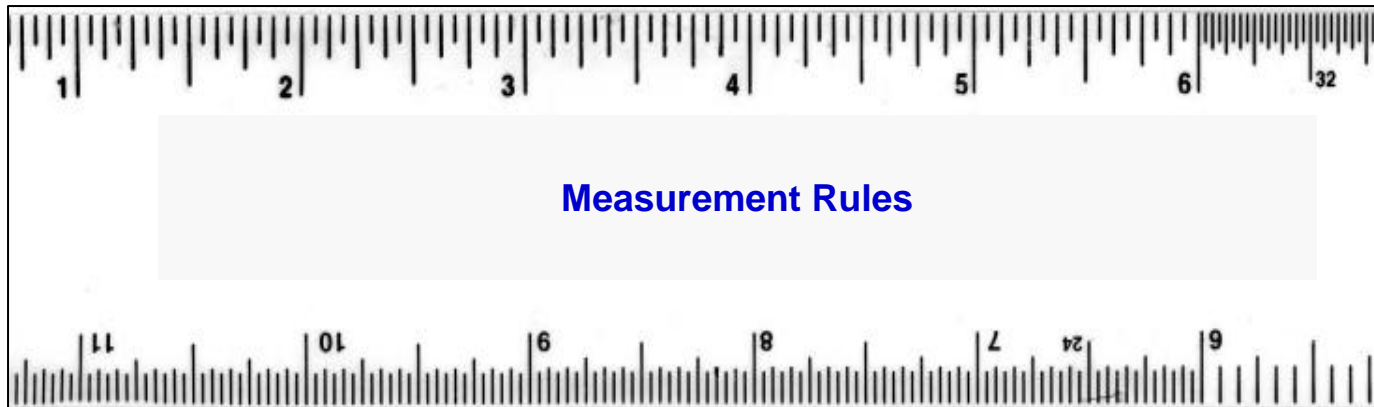
April, 5 years old, AE of 8-1 on fine motor task



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**

Envision that an equal interval ruler underlies the test.

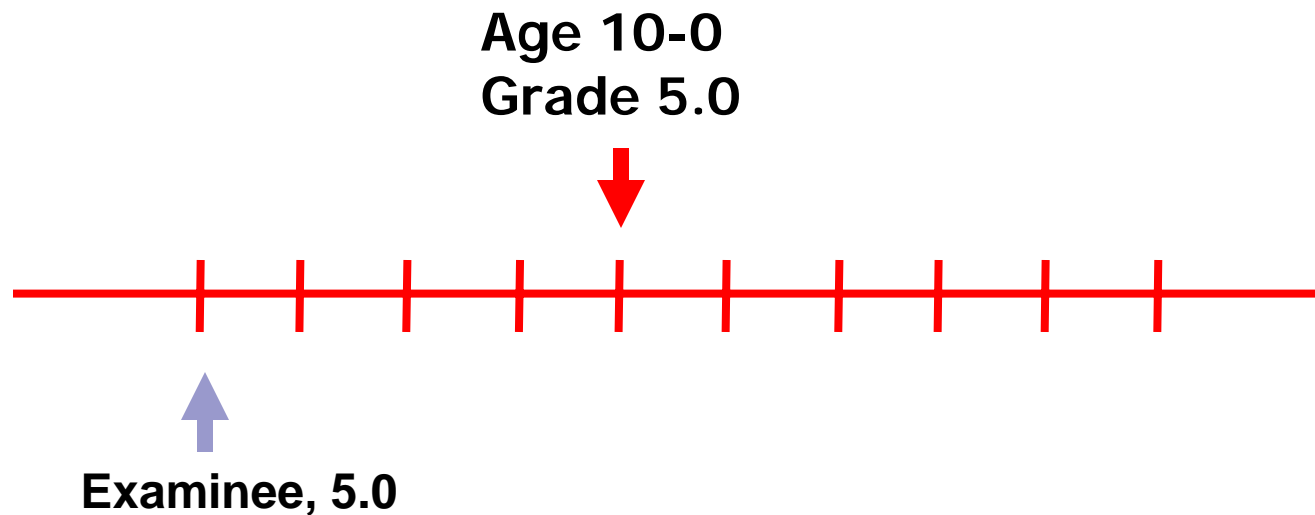


Every item in the test is located at a point on the ruler.

Every person's performance is located at a point on this same ruler.

The mean or median for each norm group is located at a point on this ruler.

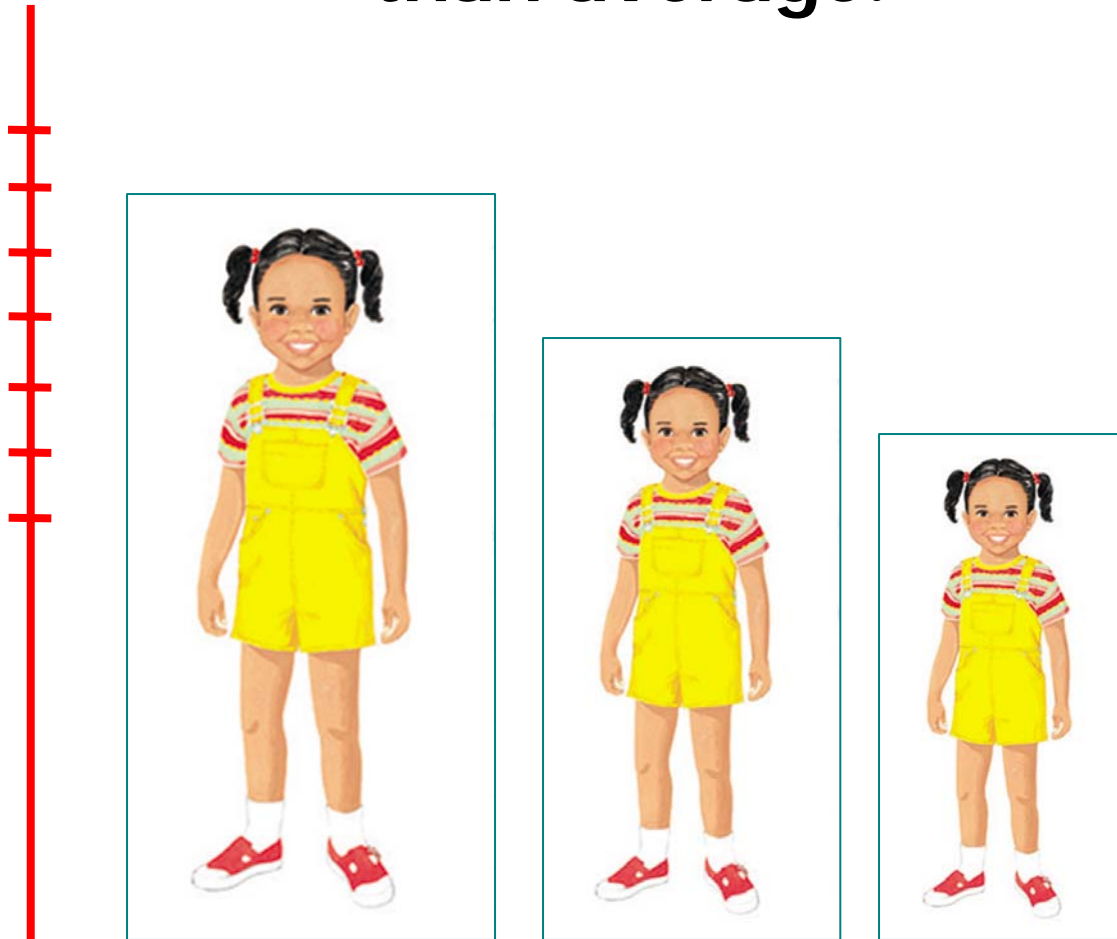
We can see where on the ruler each examinee's performance is located and how far it is from the average performance for their age or grade.



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

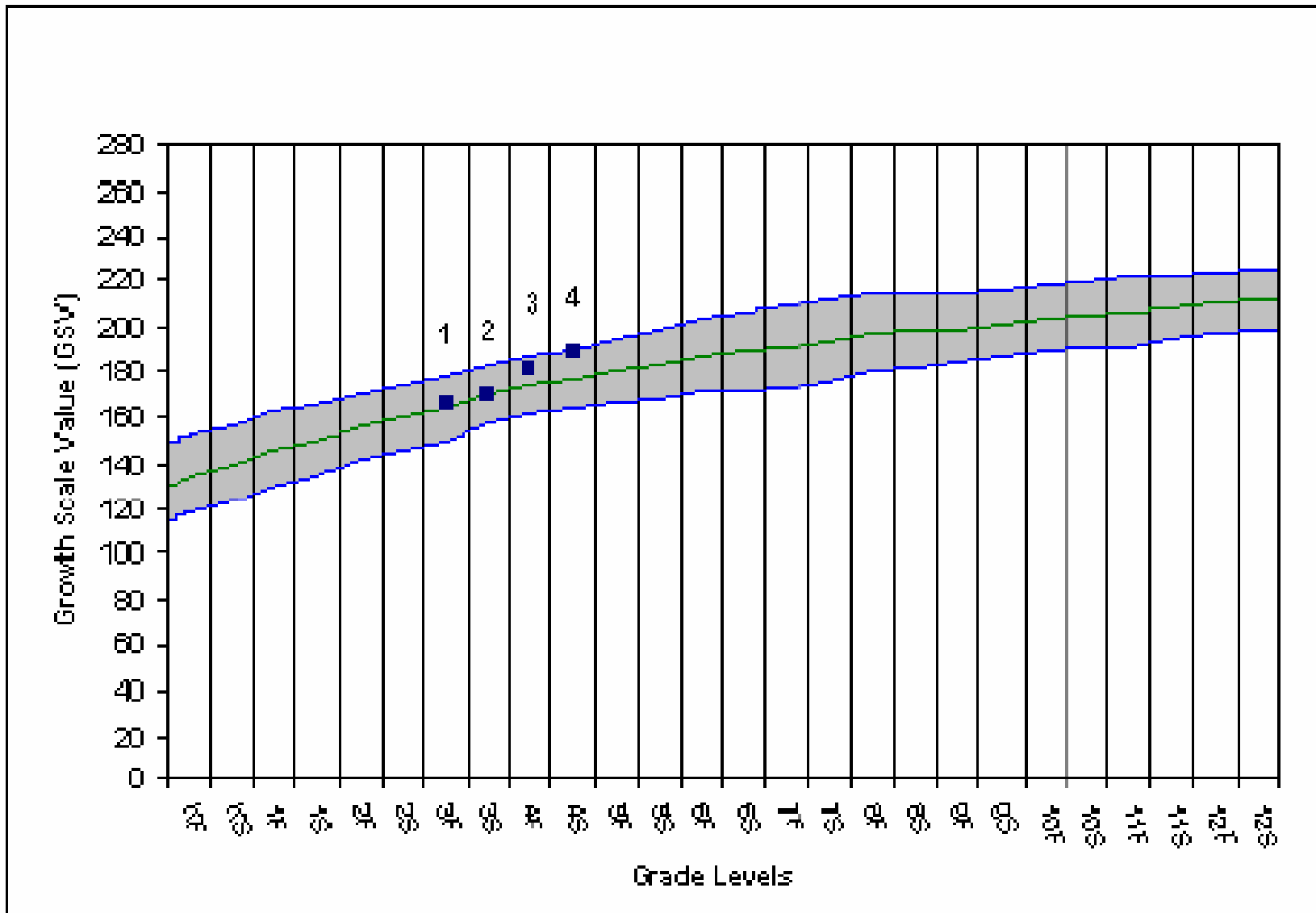
Growth can be documented even if the child is “shorter than average.”

Scale



GSV Chart

Peabody Picture Vocabulary Test (PPVT-4)



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.

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.

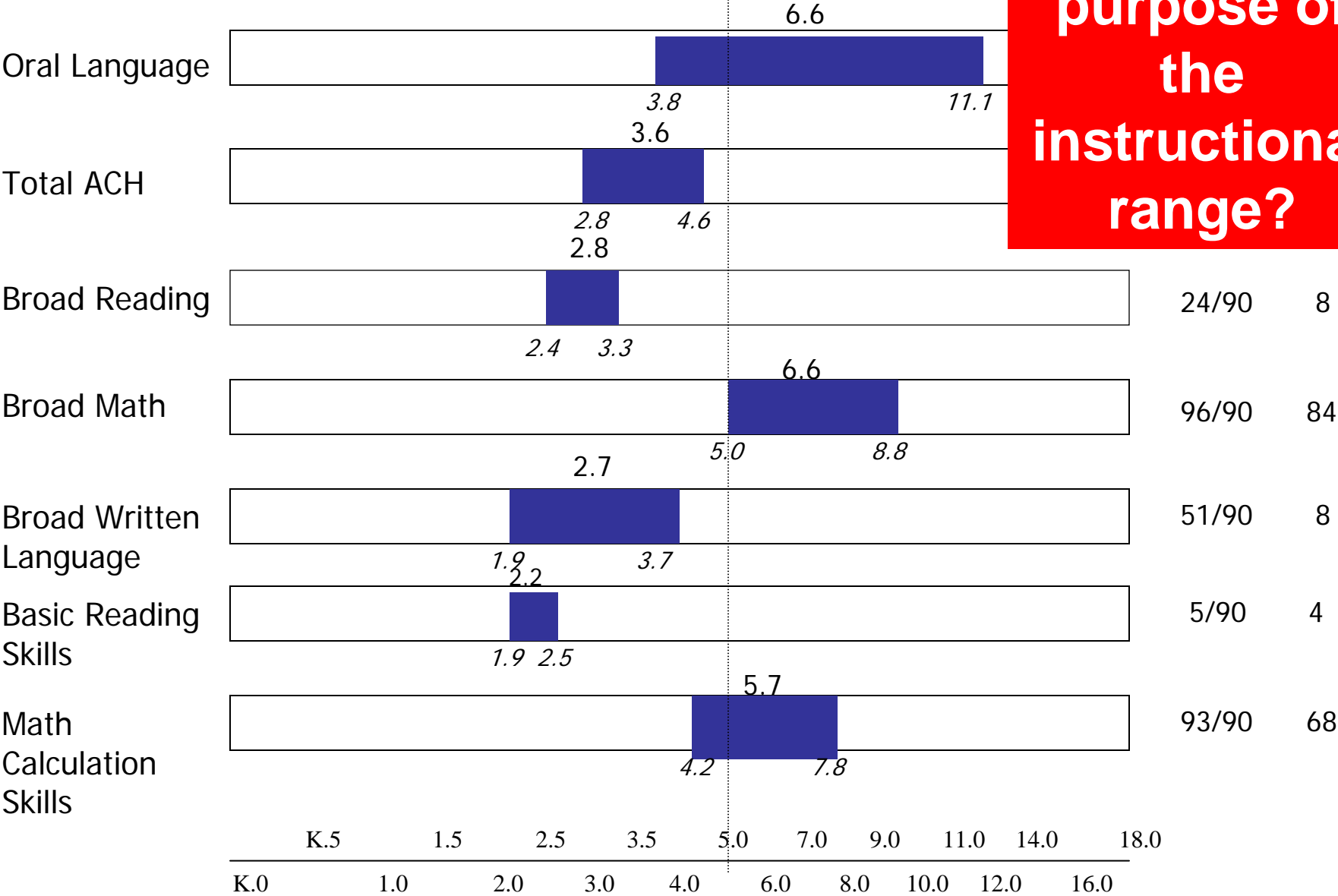
RPI	Instructional Level
96/90 to 100/90	Independent
76/90 to 95/90	Instructional
75/90 and below	Frustration

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.

RPI	Proficiency	Functional Level	Development/ Delay
100/90	Very Advanced	Very Advanced	Very Advanced
98/90 to 100/90	Advanced	Advanced	Advanced
95/90 to 98/90	Average to Advanced	Within Normal Limits to Advanced	Age-appropriate to Advanced
82/90 to 95/90	Average	Within Normal Limits	Age-appropriate
67/90 to 82/90	Limited to Average	Mildly Impaired to Within Normal Limits	Mildly Delayed to Age-appropriate
24/90 to 67/90	Limited	Mildly Impaired	Mildly Delayed
3/90 to 24/90	Very Limited	Moderately Impaired	Moderately Delayed
0/90 to 3/90	Negligible	Severely Impaired	Severely Delayed

What is the purpose of the instructional range?

Jody's grade placement = 5.0





Level 3: Proficiency, Growth, & Instructional Ranges

Sample Descriptions

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.



Level 3: Proficiency & Instructional Range

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 Comparisons-Relative Standing

- **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.**

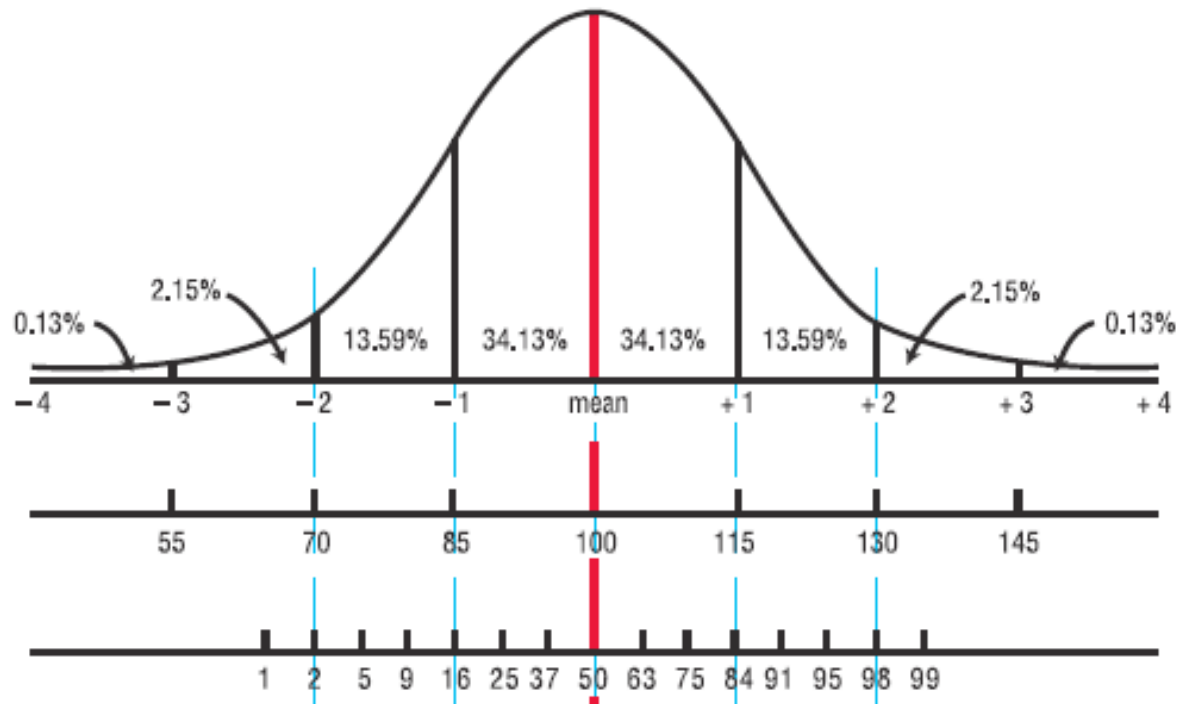
Reviewing the Normal Curve

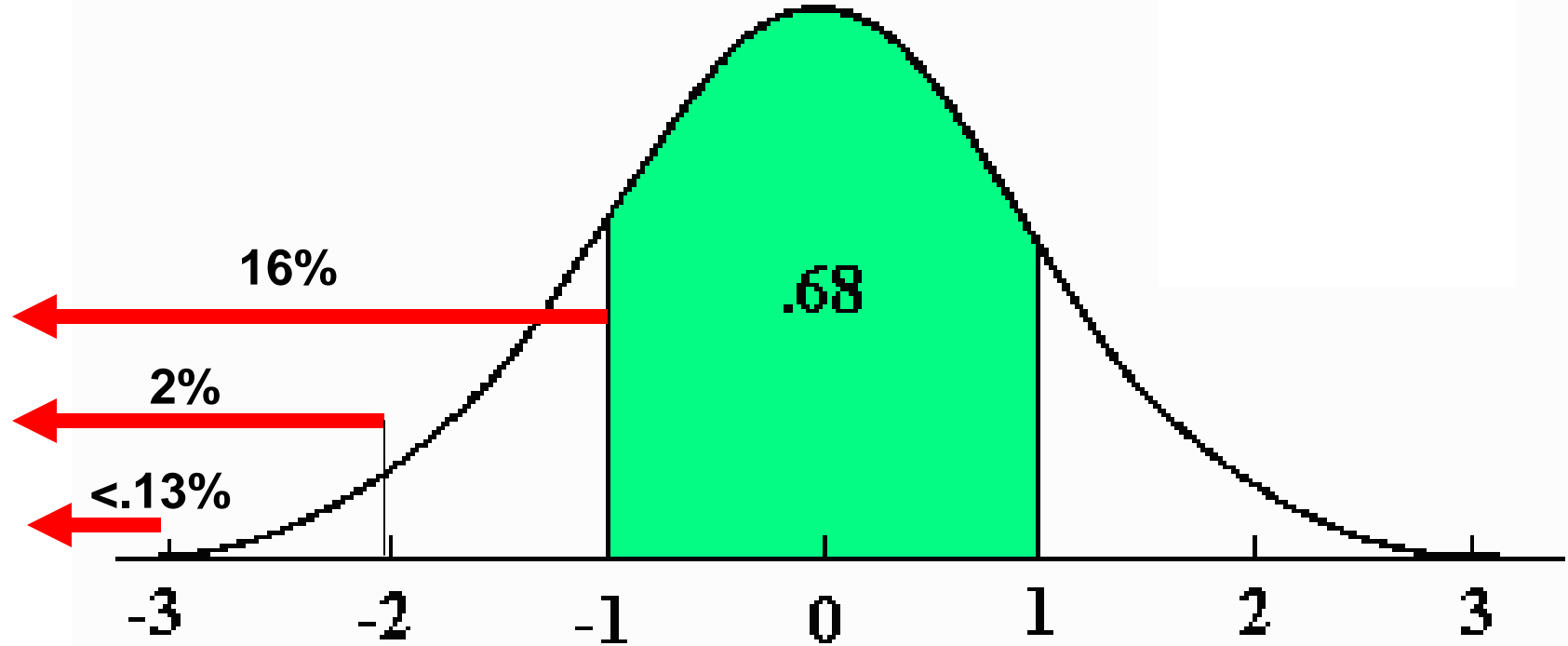
Percent of scores under each portion of the normal curve

Standard deviation

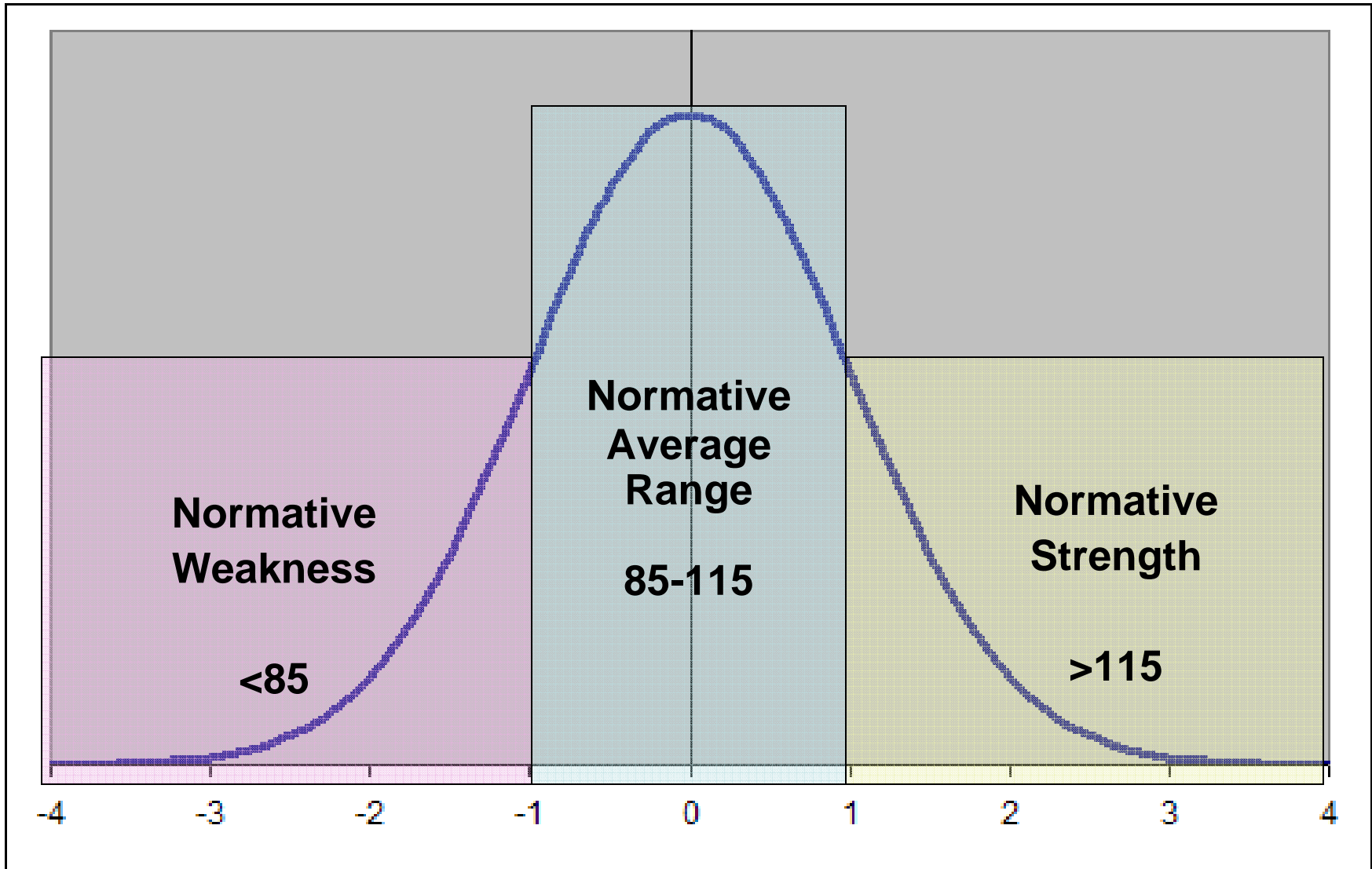
Standard score
(mean of 100, standard deviation of 15)

Percentile rank





- 68% of the population falls between -1 and +1**
- 32% of the population falls below -1 or above +1**
- 4% of the population falls below -2 or above +2**
- <.3% of the population falls below -3 or above +3**



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



Statistically Significant Differences

There really are three tests:

1. Are the scores significantly different (not chance variations of essentially identical scores)?
2. Is the difference also unusual?
Base rates, Discrepancy PR
3. Unusual or not, does the difference have educational implications for the individual?



Confidence Bands/Intervals

Increases “confidence” that the examinee’s true score falls within an identified range.

Uses the standard error of measure (SEM) around the obtained standard score to create the range.

Commonly available confidence levels are: 68%, 80%, 90%, and 95%. The higher the confidence level, the wider the band.



Confidence Band Examples

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

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

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

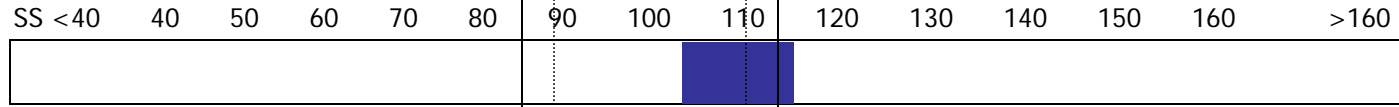
71-77 at the 68% level (+/-1 SEM) +/-3

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

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

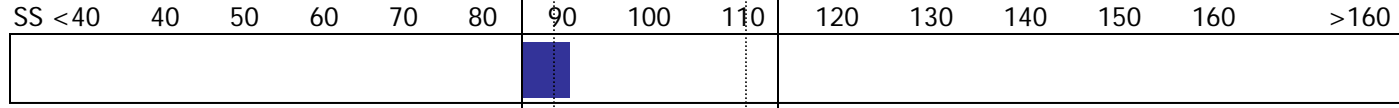
Average

Oral Lang.



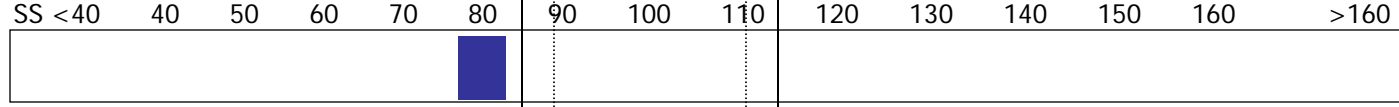
PR <0.1 .1 .5 1 2 5 7 10 15 20 30 40 50 60 70 80 85 90 93 95 98 99 99.5 99.9 >99.9

Total ACH



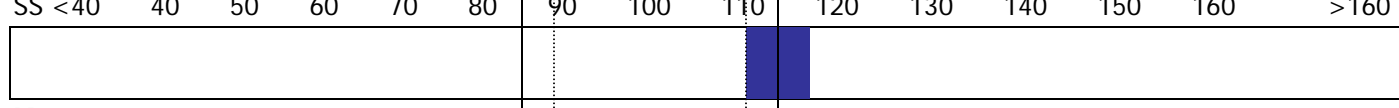
PR <0.1 .1 .5 1 2 5 7 10 15 20 30 40 50 60 70 80 85 90 93 95 98 99 99.5 99.9 >99.9

Broad Reading



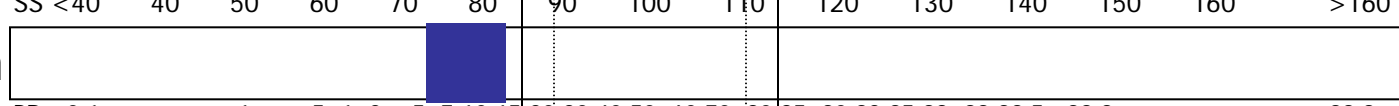
PR <0.1 .1 .5 1 2 5 7 10 15 20 30 40 50 60 70 80 85 90 93 95 98 99 99.5 99.9 >99.9

Broad Math



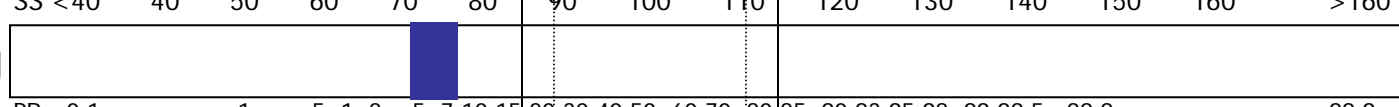
PR <0.1 .1 .5 1 2 5 7 10 15 20 30 40 50 60 70 80 85 90 93 95 98 99 99.5 99.9 >99.9

Broad Written Language



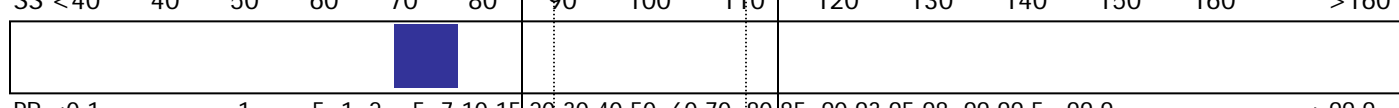
PR <0.1 .1 .5 1 2 5 7 10 15 20 30 40 50 60 70 80 85 90 93 95 98 99 99.5 99.9 >99.9

Basic Reading Skills



PR <0.1 .1 .5 1 2 5 7 10 15 20 30 40 50 60 70 80 85 90 93 95 98 99 99.5 99.9 >99.9

Basic Writing Skills



PR <0.1 .1 .5 1 2 5 7 10 15 20 30 40 50 60 70 80 85 90 93 95 98 99 99.5 99.9 >99.9

Average



Mental Retardation

....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: 71**
- **Consider one SEM (+/- 3) = 68-74**

Conclusion: **Student may be found eligible for special education services based on a diagnosis of mental retardation.**



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**
- **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**



Level 4: Peer Comparison-Relative Standing

Sample Descriptions

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).



Level 4: Peer Comparison-Relative Standing

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



Which Score To Use?

2nd grader (2.9)

College Senior (16.9)

SS

75

75

PR

5

5

GE

1.1

6.3

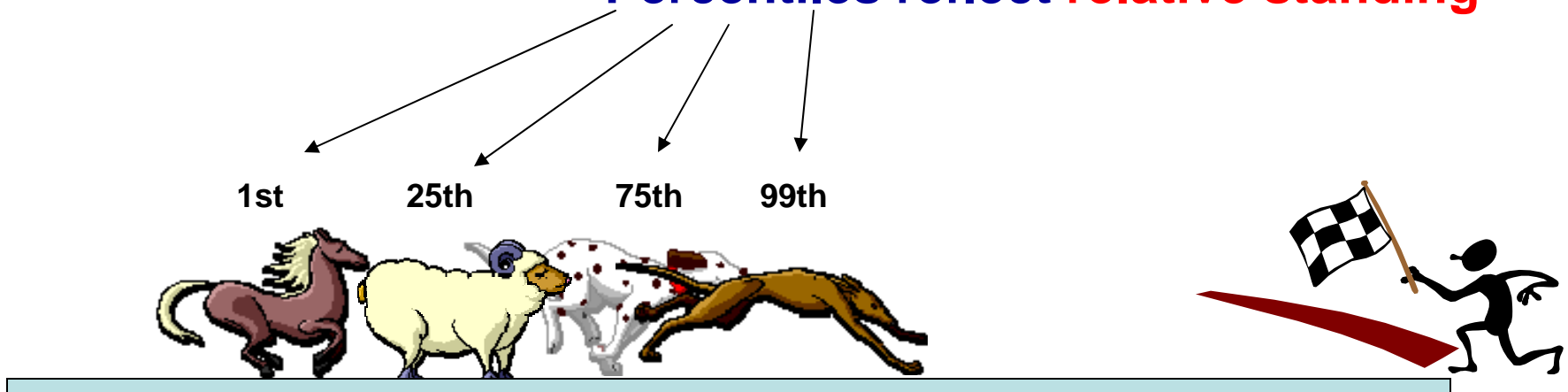
RPI

10/90

68/90

Results from Word Attack.

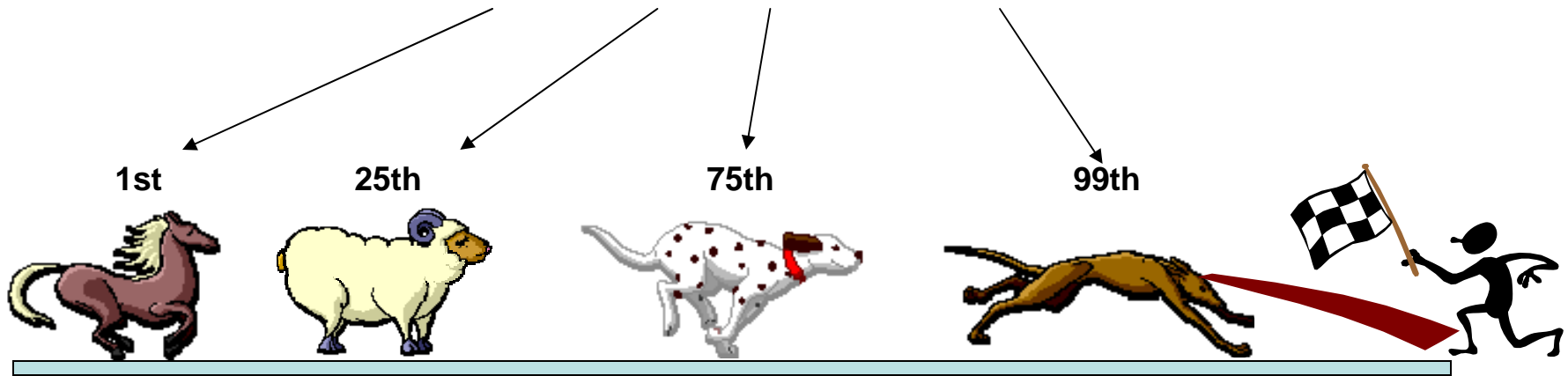
**Standard scores and
Percentiles reflect relative standing**



Sprint Analogy: All racers finish close together.

Norms: A narrow distribution (individuals did not vary too much on the task)

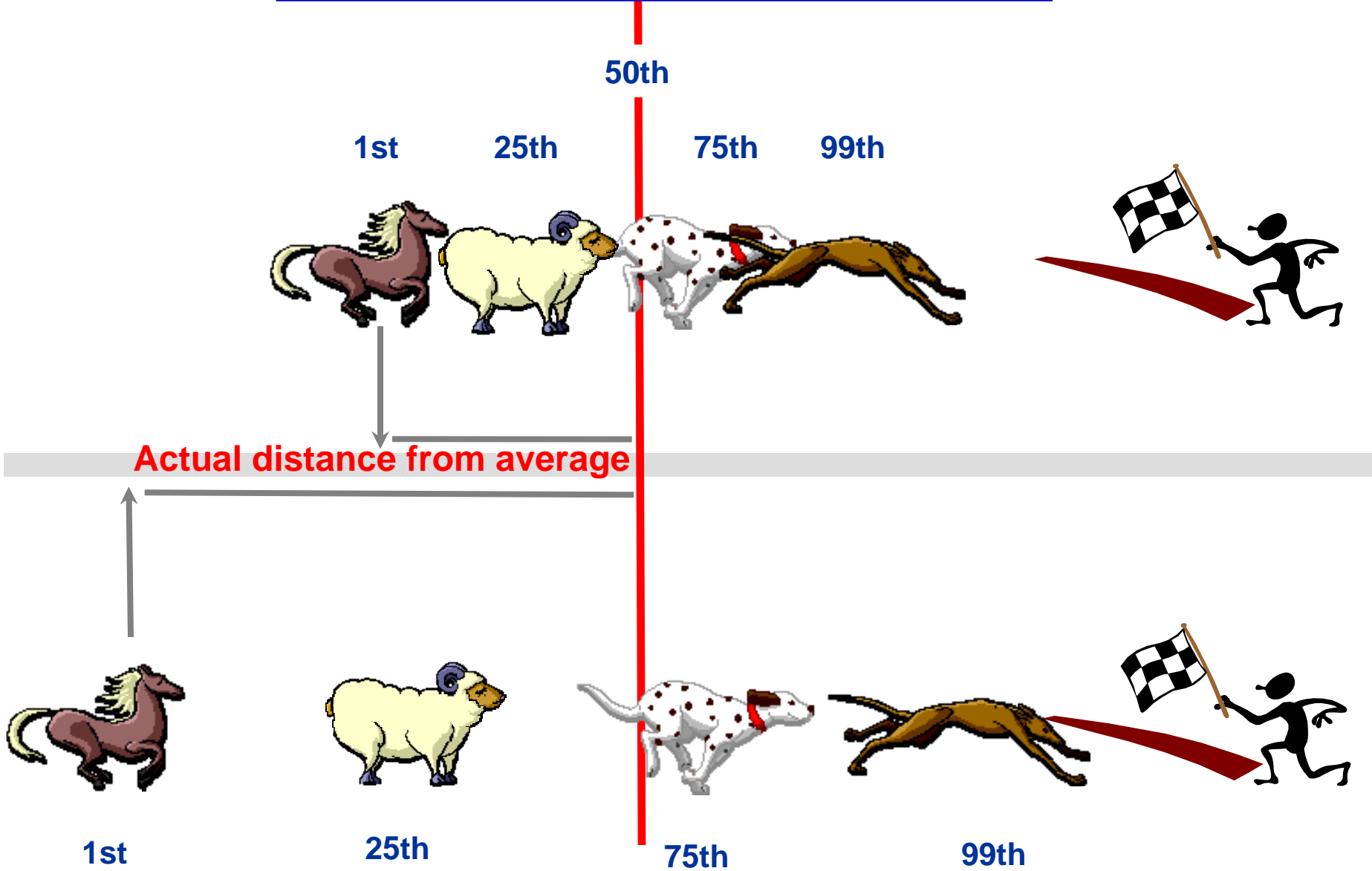
Notice that the percentiles don't change since **relative standing remains the same**



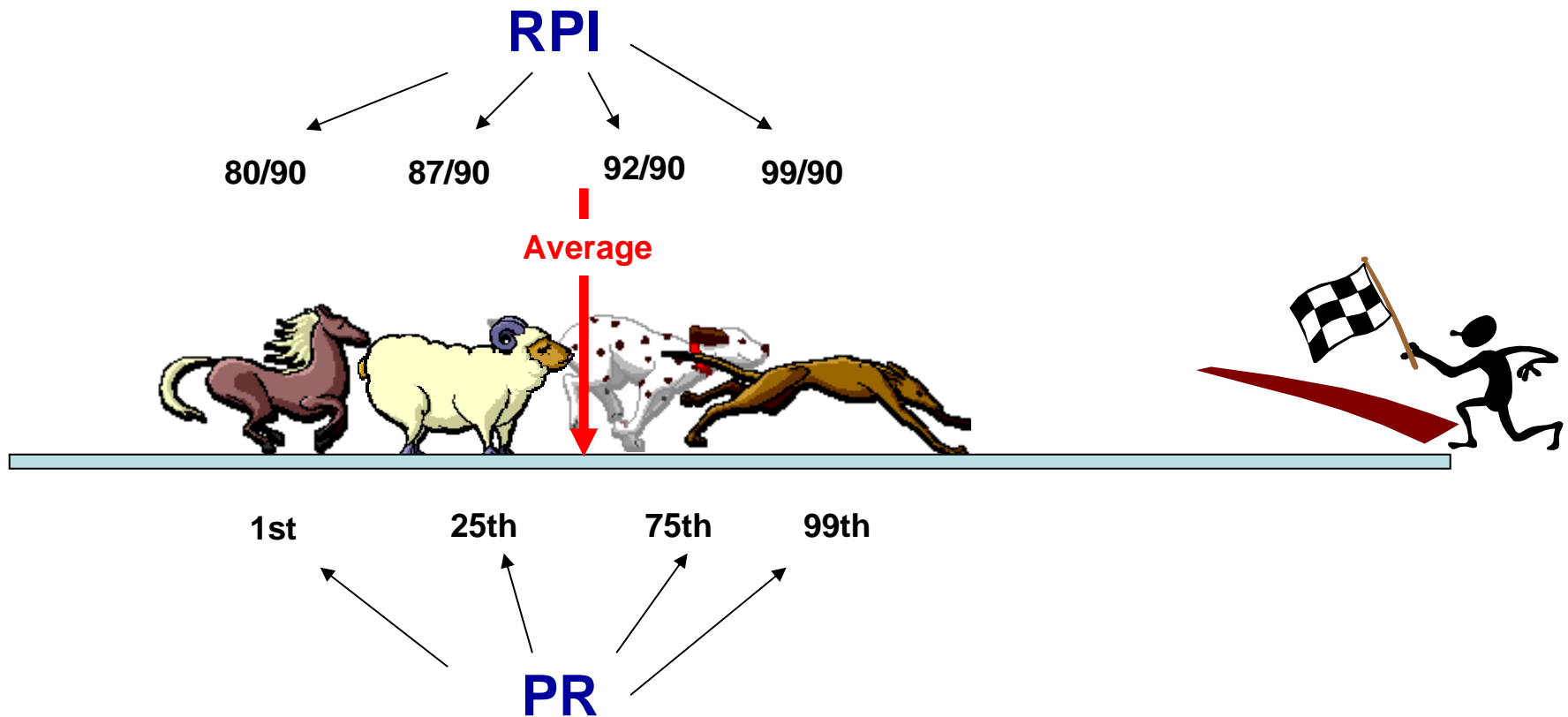
Marathon Analogy: Racers are spread out at the finish.

Norms: A wider distribution (individuals vary widely on the task)

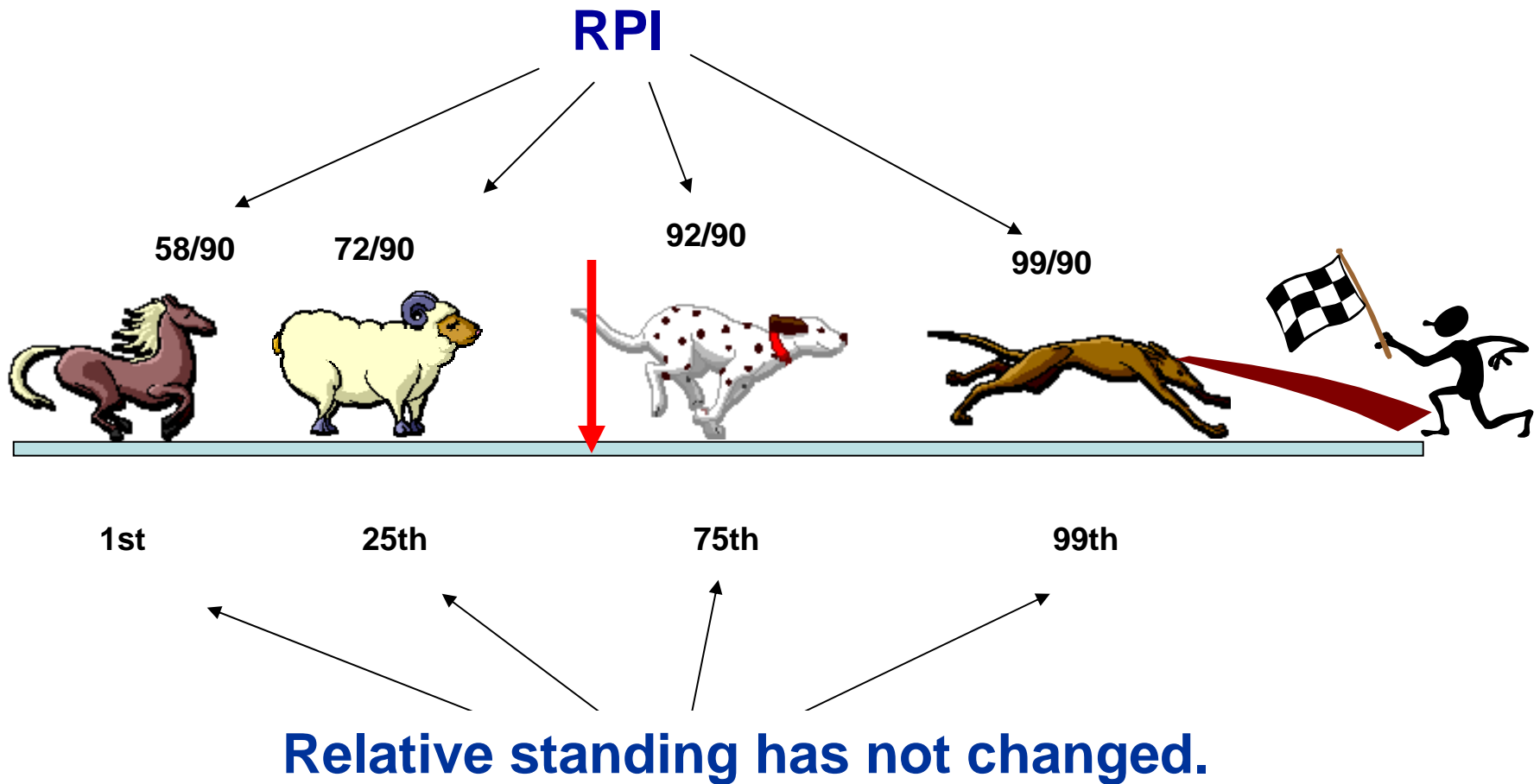
Average level of performance or proficiency



“How far from average is a person’s performance?”



Absolute distance from average has changed.





Which 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 (5th Grader)

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?



Educational & Family History

Performance in other achievement areas

Phonemic Awareness, Language, Abilities

Qualitative Information:

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



Tricky Score Issues



Tricky Issue #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?**



Composite Scores

Total or composite scores will be more extreme (farther from the mean) than the average of the component scores (unless all of the component scores are perfectly correlated).


Many composites are comprised of subtests that measure distinctly different skills/abilities, so they don't have perfect correlations.



WJ III Example

GIA (EXT)	82
Comprehension-Knowledge	97
Long-Term Retrieval	95
Visual-Spatial Thinking	97
Auditory Processing	102
Fluid Reasoning	79
Processing Speed	60
Short-Term Memory	91

(Average is 88.7 or 89)



Comprehensive Assessment of Spoken Language (CASL) Example

Core composite: 72

Antonyms 74

Morphemes 80

Sentence Comp 87

Nonliteral language 76

Pragmatic Judgment 73



WISC-IV Example

Verbal Comprehension Index	91
Perceptual Reasoning Index	98
Working Memory Index	71
<u>Processing Speed Index</u>	<u>75</u>
Full Scale (average is 84)	81



Tricky Issue #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?**



Differences Within a Composite

Cognitive Fluency = 75

Decision Speed = 98

Retrieval Fluency = 70

Rapid Picture Naming = 71

If significant differences exist among the tests within a cluster, report performance on the narrow abilities and attempt to explain the reasons for the difference.



Tricky Issue #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?**



Declining Standard Score Issue




In Grade 2.0, a student obtained a SS of 106 on a test.

In Grade 4.0, the same student obtained a SS of 89 on that same test.

How would you explain this decline in Standard Scores?

**Has the student made any progress?
How would you determine this?**



“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. (2001). Psychological testing and psychological assessment. American Psychologist, February

Scores Module Post-Test



*Please complete the Post-Test.
Compare your results from the
Pre- and Post-Tests.*