


# Validity and Reliability and Other Terms

Assessment Literacy part 2-Admin meeting 3-5-19

1



## Validity

Inferences educators draw from a valid test should be a strong predictor of student's level of proficiency

Items measure what they intend to measure

There needs to be items that differentiate student skill levels (DOK should be varied)


2



### Complexity vs. Difficulty, Cont. Webb's DOK Levels

|  |                  |               |         |                       |
|--|------------------|---------------|---------|-----------------------|
| C<br>O<br>M<br>P<br>L<br>E<br>X<br>I<br>T<br>Y | Level of Thought | EVALUATION    | Level 1 | RECALL OF INFORMATION |
|  |                  | SYNTHESIS     | Level 2 | BASIC REASONING       |
|  |                  | ANALYSIS      | Level 3 | COMPLEX REASONING     |
|  |                  | APPLICATION   | Level 4 | EXTENDED REASONING    |
|  |                  | COMPREHENSION |         |                       |
|  | KNOWLEDGE        |               |         |                       |
|  |                  |               |         | Amount of Effort      |
|  |                  |               |         | DIFFICULTY            |

3



### Hess Cognitive Rigor Matrix

| Webb's DOK | DOK LEVEL 1 Recall & Reproduction   | DOK LEVEL 2 Basic Skills & Concepts                           | DOK LEVEL 3 Strategic Thinking & Reasoning   | DOK Level 4 Extended Thinking   |
|------------|---|---|--|---|
| Remember   | -Recall, locate basic facts, definitions, details, events                   |   |  |   |
| Understand | -Select appropriate words for use when intended meaning is clearly evident. | -Specify, explain relationships                               | -Explain, generalize, or connect ideas using supporting evidence                               | -Explain how concepts or ideas specifically relate to other content domains or content        |
| Apply      | -Use language structure or word relationships to determine meaning          | -Use context to identify word meanings                        | -Use concepts to solve non-routine problems  | -Devise an approach among many alternatives to approach a novel problem                       |
| Analyze    | -Identify the kind of information contained in graphic, table, visual, etc. | -Compare theory elements, facts, terms, events                | -Analyze or interpret author's craft to critique a text  | -Analyze multiple sources or texts  |
| Evaluate   |   | -Analyze organization & text structures                       | -Evaluate evidence and develop a logical argument for conjectures based on one text or problem | -Evaluate relevance, accuracy & completeness of information across texts/sources              |
| Create     | -Generate ideas, concepts, etc. related to a topic or concept               | -Generate hypotheses based on observations or prior knowledge | -Develop a complex model for a situation   | -Synthesize information across multiple sources or texts, articulate new voice or perspective |

4

## Purchased Assessments

Do they address the standards as rigorously as your state assessment?  
Do they provide items that encourage deep critical thinking (multiple standards assessed in one question)?

Do they reflect item stems from the state assessment?

Do they provide rubrics?

Do wrong answers illuminate understanding? (Can you review the questions and answers?)

5

## Reliability



How consistently will a student score on this test over time? The test should give you the same scores time after time.

Fewer items for a standard or skill=less reliability

6

## Standardized Assessments



Data is valuable

Data must be triangulated with other assessments

You should not use one assessment to explain a student's proficiency.

7

## Norm-Referenced Test



Are designed to describe performance by comparing other test taker's performance

Growth is usually seen in the middle

8

### Criterion-Referenced Test

Are designed to describe performance by comparing to level of proficiency (cut scores)



9

### Standards-Referenced Test

Similar to a criterion-referenced test, but usually a scale of 1-4 is attached to each standard and a scaled score is provided as well (scaled score is not a good reporting tool) WHY?



10

### Triangulate the Data

Over and over it is emphasized to triangulate data

Next time, we will look at our assessment inventory and the data we have available



11