Technical Adequacy

Validity, Reliability, and Alignment
Alternate Assessment ReVision Group
November 16, 2005
What Is Technical Quality?

Handbook for Professional Development in Assessment Literacy

By:
Jan Sheinker, Ed.D.
Doris Redfield, Ph.D.
VALIDITY

• Are we measuring what we say we are measuring?

• Can we make valid interpretations/inferences?
INCREASING VALIDITY

- Use more than one measure
- Make sure it measures *what it says it does*
- Base it on *agreed-upon standards*
- *Align* with expected performance
EVIDENCE

Do Tests Measure What They are Supposed To?

• Items measure knowledge & skills from:
  - Content Standards
  - Curriculum Frameworks

• Scores on the test would be
  - Similar to scores on similar tests

• Item formats are appropriate for
  - Measuring what is supposed to be measured

• Tests represent agreement about
  - What students should know & be able to do as measured by the test
RELIABILITY
is consistency of test results

From one test to another
From one part to another
From one time to another
From one scorer to another

August 2002
Validity and Reliability

A reliable measure may not be valid.

A valid test must be reliable to be valid.
Evidence of RELIABILITY

- **Test to Test**: 2 tests same material → Same results
- **Time to Time**: 1 test 2 times → Same results
- **Items to Items**: 2 halves 1 test → Same results
- **Scorer to Scorer**: 2 scorers 1 test → Same results

August 2002
Documentation of Alternate Assessment Technical Adequacy

ASES DAATA Project
January 19 & 20, 2005
Orlando MEGA SCASS
Content Validity and Alternate Assessment—

* A Brief Primer of Alignment

One aspect of alternate assessment technical adequacy
What is test validity?

It is an evaluative statement about an assessment. It addresses the acceptable uses of assessment results.
What validity is and is not:

**IS**
- A statement of confidence in alternate assessment results.
- A continuous improvement *target, in the beginning a baseline.*
- Overseen by the state.

**IS NOT**
- A single score that provides a pass/fail rating about assessment quality.
- A mysterious judgment made by nameless professionals behind closed doors.
Achievement Tests Typically Focus On Content Validity

- English Language Arts
- Mathematics
- Science
- Social Science
Content Alignment

Content Standards

Standards direct curriculum, instruction, and assessment provides opportunity to learn.

Alternate Assessment

Standards direct assessment content

Curriculum and Instruction

Curriculum and instruction
Content Validity Arguments have **RELATED** (Linn, 2002)

<table>
<thead>
<tr>
<th>Assessment Content</th>
<th><strong>TO</strong></th>
<th>Domain Content</th>
</tr>
</thead>
</table>

Represented by

- Tables of Test Specifications
- Content of Test Items
- Definitions for Reporting Strands

**Mapped to:**

Represented by

- **Content Standards** with specific content (e.g., $+ - \times /$) and
- **Process** (e.g., factual knowledge, conceptual understanding, problem solving)
This is called “Alignment”

Webb’s process examines

• Test items & strands or reporting clusters
And
• Content Standards
Tindal adapted Webb’s Alignment Process for Special Education

Using same four components
- Categorical Concurrence
- Range of Knowledge
- Depth of Knowledge
- Breadth of Knowledge
Alignment Questions:

• **Categorical Concurrence:** *Does the alternate assessment address subject matter contained in the state Content Standards?*

• **Range of Knowledge:** *Does the alternate assessment address every Content Standard? To what extent is each standard represented? A lot, a little?*

• **Depth of Knowledge:** *Does the alternate assessment address the standard at the level of difficulty (cognitive complexity) reflected in the content standard?*

• **Balance of Representation:** *For each content standard (or indicator) represented in the alternate assessment, how often does it appear?*
Content Alignment

Adapted from Grisham-Brown and Kearns (2001)

- **Content Standards**: Standards direct curriculum and instruction
- **Alternate Assessment**: Standards direct assessment content
- **Curriculum and Instruction**: Alignment between curriculum, instruction, and assessment provides opportunity to learn.
Addressing Alignment in Parallel Ways

General Assessment

• Content Standards and Grade Level Expectations
• Test specifications
• Score Reporting Category or Strand Descriptions

Alternate Assessment

Less complex performances of content standards and/or grade level expectations
Access to the General Curriculum: A Continuum of Learning
(Mathematics)  Dan Wiener & Pam Green 2002

Grade 7-8
Learning Standard #2 for Algebra:
Solve simple algebraic expressions for given values
Example: $3a^2 - b$, for $a=3$ & $b=7$

‘Entry Points’

Less Complex
Match pictures & objects to create and compare sets

More Complex
Understand symbols and meaning of:
* addition +
* subtraction -
* equal to =

‘Entry Points’

More Complex
Solve simple one- and two-digit number sentences
Example:
1 + 1 + 1 = x
2 + x = 5
3x + 8 = 29

Standard ‘as written’
Access to the General Curriculum: A Continuum of Learning
(ELA – Reading and Literature)
Dan Wiener & Pam Green 2002

Grade 7-8
Learning Standard #16.10 for Reading and Literature:
Identify and analyze mythologies from different cultures
Example: Student creates a hero tale, using epic tale conventions (e.g., quest, special weapons)

‘Entry Points’

Less Complex  More Complex

Respond to epic tales read aloud by selecting/drawing pictures related to the story
Recognize that an epic tale is fictional
Example: Student reads (or listens to) adapted stories, and categorizes each as ‘make-believe’ or ‘real’
Identify elements of fiction in an epic tale
Example: Student reads an epic tale, identifying details related to characters, setting and plot

Standard ‘as written’
Recognize that genes are passed from parent to child

Example: Student conducts genetic crosses after researching a particular human trait (e.g., blood type)

Give examples demonstrating that some genes are weaker (recessive) and some are stronger (dominant)

Example: Student researches human traits and classifies different expressions of the trait as ‘weak’ or ‘strong’

Perform genetic crosses to determine the probability of inheriting certain traits/characteristics

Example: Student surveys peers about eye color, noting similarities/differences between student and their parents

Recognize and match similarities/differences in same species (e.g., hair color in humans)

Entry Points

Less Complex

More Complex

Standard ‘as written’