

Designing Valid Assessments

Kanna Hudson, Title III Researcher
Office of Planning & Effectiveness
Clark College

November 30, 2012

Today's participant outcomes...

1. Explain that there are multiple facets and types of validity
2. Define two or more types of validity
3. Apply information about validity to your own assessments

What is assessment?

- Explores fundamental questions about learning and the conditions for learning
- Organized and systematic
- Allows us to make inferences about learning
- Serves many purposes, internal and external



What is validity?

- Does it measure what it's supposed to measure?

Impression of Student Learning



Impression of Student Learning



Impression of Student Learning



Chris Jordan's *Cans Seurat 2007* depicts 106,000 aluminum cans arranged as *Sunday Afternoon on the Island of La Grande Jatte* (Georges Seurat)

Good outcomes make valid assessment much easier!

Not so good: Students will understand metric measurement of length.

Better: Given a metric ruler, students will be able to measure the length of a common linear object to the nearest millimeter.

Not so good: Students will know how the digestive system works.

Better: Students will be able to explain how food is processed through the digestive system.

Source: “Writing Effective Course Outcomes: No ‘Understanding,’ ‘Appreciating,’ or ‘Learning’ Allowed!”, Whatcom Community College

Reliability: a condition of validity

- Is it repeatable?
- Does it yield the same results even if it is administered in different settings or by different people?
- Is the scoring process applied consistently among different people? (inter-rater reliability)
- Reliability is a necessary condition of validity.

Interest to participate: a condition of validity

- A successful assessment project requires that everyone involved is motivated and interested to participate
- The assessment method should be...
 - Engaging
 - Respectful
 - Meaningful
 - Practical
 - Authentic
- Interest to participate is a necessary condition of reliability and validity.

Triangulation: supports validity

- Study something from more than one standpoint
- Multiple assessment strategies support the same conclusion
- Can happen within one assessment process (e.g. a test with multiple choice and essay questions)
- Triangulation can help support validity.

Types of validity

Construct validity

- Does it measure the construct it's supposed to measure?
 - For example, to what extent does the assessment measure “college readiness”?

Content validity

- Does it measure the course content and/or outcomes that it's supposed to measure?
 - For example, to what extent does the assessment measure knowledge of photosynthesis?

Evaluating content validity

Student Learning Outcomes : Students will be able to...

	...list the fifty states.	...recall each state's founding date	...hypothesize about the significance of each state's founding date in relation to historical events.
Question 6. List the fifty states and their founding dates.	4	4	1
Question 10. Summarize separately how Virginia and Hawaii were founded, and then compare the historical contexts of each.	2	3	4

1	No Match
2	Minimal
3	Moderate
4	High Match

Predictive validity

- How dependable is the relationship between the assessment results and a particular future outcome?
 - For example, does an assessment in Biology 1 measure how well the student will do in Biology 2?

Formative validity

- How well does the assessment provide information that is useful for improving what is being assessed?
 - For example, can you use the results of a pre-test to guide what you teach this term?

Sampling validity

- How well does the assessment reflect the full range of what is being assessed?
 - For example, does the test only include questions about the 1960s in the U.S., when it's supposed to cover the whole 20th century in the U.S.?

Important considerations

Relevance and representativeness

- The test should cover what the program covered and should place emphasis in proportion to the program's emphases

Internal test structure

- You may include multiple items in an assessment that assess the same content
- This is a form of triangulation

Over time and across groups and settings

- In studying validity evidence over time, some outcome measures should increase over time
- For example, a post test should have higher scores than a pre test

Fairness and bias

- If a test or rating scale is (often unintentionally) biased against certain groups of people, that test or scale is not valid, nor reliable.
 - For example, a math word problem may require cultural knowledge not covered in the class that is more/less familiar to certain populations (e.g. familiarity with the rules of football, etc.

Evaluating content validity

- Fill in the worksheet
- Use either the outcomes you brought, or the assessment you brought, or both
 - (Or use your imagination)

Assessment is...

*“...much more about **the conversations one can stimulate** around whatever data or evidence are collected than about the precise nature of the evidence gathered.”*

*(Bill Moore, “Introduction,” *Journal of Applied Research in the Community College*, vol. 9, # 2, Spring 2002.)*

Get in touch!

- **Ann Fillmore**, Assessment Liaison (CTE)
 - afillmore@clark.edu / 992-2365
- **Toby Peterson**, Assessment Liaison (Transfer)
 - tpeterson@clark.edu / 992-2084
- **Kanna Hudson**, Planning and Effectiveness
 - khudson@clark.edu / 992-2265