GETTING THE MOST OUT OF YOUR TESTS

Mika Hoffman
Center for Educational Measurement
Excelsior College
Why do you test?

To obtain information that helps someone make a decision
What makes a good test?

The *information* obtained from it is

• **Accurate**

  and

• **Relevant**

  to the *decision* being made
Validity

Two ways to think about validity:

- The degree to which a test actually measures what it claims to measure
- The extent to which inferences, conclusions, and decisions made on the basis of test scores are appropriate and meaningful
The validity argument

- A series of statements of the **conclusions** to be drawn from the test scores and **evidence** showing that the conclusions are warranted

  - Amount and strength of the evidence varies with the stakes of the test
  - Evidence comes from all stages of development
What information does the test provide?

Information from test items
  • Task difficulty
  • Task discrimination

Information from test scores
  • Individual level
    • Overall ability
    • Ability by content area or task type
  • Group level
    • Comparing sections of a course
    • Comparing students from different backgrounds
Who uses the information?

Decision-makers can be

- Instructors for the course
- Instructors for other sections
- Curriculum designers
- Department heads / deans / other administrators
- Institutional effectiveness offices
- Accreditors
What kinds of decisions are made?

Decisions can be of different types:

- Decisions at the group level
- Decisions at the individual level
- Formative
- Summative

What are the consequences of these decisions?

- For examinees
- For instructors
- For administrators
- For programs
How can we maximize test validity?

Test Plan document
• What the test is supposed to measure
• Specifications for its items (content, cognitive level, types)
• Administration specifications
• Scoring specifications

Good item-writing practices
• Avoid construct-irrelevant variance

Scoring rubrics
• Defining levels of performance clearly
• Different judges should rate the same performance the same
The test plan

• Write down the exam’s defining characteristics
  • Information the exam scores will provide
  • Decisions to be made on the basis of the scores
  • Consequences of the decisions
  • Who will make the decisions
  • Proctored or not?
  • Open book or not?
  • Number of points per item
  • How grades or passing will be determined

• What content is to be covered?
• What learning outcomes are to be covered?
• How many items?
• What type of items?
  • Essay
  • Short Answer
  • Matching
  • Multiple Choice
• What cognitive level of items?
A few words on scores

Two ways of looking at test scores
- Norm-referenced
- Criterion-referenced

How do you determine what a passing score is? Or cut points for different grades?
- Scoring rubrics for essays or performance
- Getting enough items right overall?
- Getting enough of the “hard” items right?
Item writing

• Test only what you want to test, no more no less
• Don’t add extra verbiage
• Don’t write multiple-choice options that are silly or tricky
• Phrase all questions clearly
• Make it clear to students what is expected of them
• Make questions that are worth answering (testing relevant and important aspects of the subject)
Scoring rubrics

- Promote reliability in rating student performance
- Help reduce “halo effect”
- Keep raters focused on what is actually being tested
Ways of promoting validity

1. Link assessment to standards/learning outcomes
   - Review standards/outcomes for each class or unit
   - Assign weight to standards/outcomes based on their importance and consequences of non-mastery [construct a test plan]
   - Assess standards/outcomes in direct proportion to assigned weights
Ways of promoting validity

2. Match item format to the type of cognitive behavior targeted by the standard
   ➢ Recall vs. application
   ➢ Process vs. product
   ➢ Objective vs. subjective
Ways of promoting validity

3. Reduce construct-irrelevant variance
   - Reading level is appropriate for audience
   - Biased language and stereotypes have been avoided
   - Irrelevant “window-dressing” is avoided
   - Scoring rules including penalties for guessing are made known to test-takers
   - All test-takers have equal access to content beforehand
   - Adequate time is allowed to complete the test
Ways of promoting validity

4. Compare assessment outcomes with scores from other assessments
Ways of promoting validity

5. Faculty colleagues review tests prior to class administration noting errors and potentially confusing content
Ways of promoting validity

6. For extended response and other “subjective” assessments, develop scoring guidelines with examples of characteristics/behaviors for each score point.

As an example, develop descriptions of “exceptional”, “highly competent”, “competent”, “marginal”, and “below competent” performance.
Ways of promoting validity

7. Use multiple assessment formats during course

A large body of student work results in more accurate evaluation of student achievement
Ways of promoting validity

8. Ask students for feedback

- Did they understand the instructions?
- Did the questions make sense?
Ways of promoting validity

9. Analyze assessments to determine their statistical and psychometric properties (difficulty, discrimination, non-response rate, reliability)

Results can be used to verify answer key and improve quality
Ways of promoting validity

10. Compile “banks” of previously administered items and use banks to compose new tests

CAUTION
Make sure that all tests yield equivalent scores!
A word about test security

The confidentiality and privacy of assessment content and results

- Lack of security reduces validity because student scores are based on a construct other than target cognitive behavior
Summary

- Write down a test plan
- Think about the validity argument at all times
- Check everything at least twice
- Avoid construct-irrelevant variance
- Analyze results
Questions?

Mika Hoffman
Executive Director, Test Development Services
Center for Educational Measurement
Excelsior College
mhoffman@excelsior.edu
(518) 464-8773
Web resources

http://blog.questionmark.com/category/general/quizzes
  Quizzes to test your knowledge of testing

http://www.adprima.com/measurement.htm
  Brief on testing in education

http://ncme.org/linkservid/983E4748-1320-5CAE-6EFB9C4290DCE72E/showMeta/0/
  Bibliography of testing resources

http://www.charlesdennishale.com/
  E-book on assessment

http://cpd.excelsior.edu/test-development-training/
  Online professional development course
Other resources

• ALTE item writing guidelines
• Saskatchewan publication on rubrics
• Wisconsin papers on various testing topics
• AERA/APA/NCME Standards for Educational and Psychological Testing