Validity

University of Maine
Office of Assessment
Brian Doore
Director of Assessment
Agenda

• Definitions and background
• Content Validity
  – Overview
  – Examples
• Use of blueprint template on your own measure(s)
Validity

- Three Key Concepts in Judging the Quality of an Assessment or Measure
  - Validity
  - Reliability
  - Usability
Why should you be bothered with these concepts anyway?

- Appreciate why all measures contain error
- Know the various sources of error
- Understand that different kinds of measures are prone to different kinds of error
- Build measures with less error
- Know how to measure your error
- Know what is safe—and not safe—to conclude from your results
- Decide when certain measures should not be used
Validity: a definition

• Appropriateness of how scores are interpreted [and used]
  – That is, to what extent does your instrument measure what you say it does [and is as useful as you claim]?
  – Stated another way: To what extent are the interpretations and uses of a measure justified by evidence about its meaning and consequences
Validity: Key Points

– 1. a matter of degree ("how valid")
– 2. always specific to a particular purpose ("validity for...")
– 3. a unitary concept (four kinds of evidence to make one judgment—"how valid?")
– 4. must be inferred from evidence; cannot be directly measured
Validity: Four interrelated types of evidence

1. Content
2. Construct
3. Criterion
4. Consequential
Questions guiding Validation

1. **What are my objectives?**
   Did my measure really address those particular objectives?

2. **Do the students' test scores really mean what I intended?**
   What may have influenced their scores?
   - Growth
   - Instruction
   - Intelligence
   - Cheating etc.

3. **Did testing have the intended effects?**
   What were the consequences of the testing process and scores obtained?
Content-Related Evidence

• The extent to which an assessment’s tasks provide a relevant and representative sample of the *domain of outcomes* you are intending to measure.

  – The evidence:
    • most useful type of validity evidence for classroom tests
    • domain is defined by learning objectives
    • items chosen with table of specifications
# Table of Specifications

<table>
<thead>
<tr>
<th>Subject Content</th>
<th>Knowledge &amp; Comprehension</th>
<th>Application</th>
<th>Analysis, Synthesis &amp; Evaluation</th>
<th>Total # of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Art History</td>
<td>1,2,3,4,5,6,7,8</td>
<td>9</td>
<td>10</td>
<td>10 (50%)</td>
</tr>
<tr>
<td>Art Appreciation</td>
<td>13</td>
<td>12</td>
<td>11,14-20</td>
<td>10 (50%)</td>
</tr>
<tr>
<td>Total # of Items</td>
<td>9 (45%)</td>
<td>2 (10%)</td>
<td>9 (45%)</td>
<td>20 (100%)</td>
</tr>
</tbody>
</table>
Table of Specifications

- example
Content-related Evidence issues

• is an attempt to *build* validity into the test rather than assess it after the fact

• sample can be *faulty* in many ways
  – a. inappropriate vocabulary
  – b. unclear directions
  – c. omits higher order skills
  – d. fails to reflect content or weight of what actually taught
  – e. "face validity" (superficial appearance) or label does *not* provide evidence of validity
  – f. *assumes* that test administration and scoring were proper