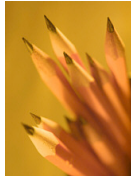


Assessment

Fall Quarter 2007



High-Stakes Testing



- All Tests Have Shortcomings
- Tests Tend to Concentrate on Discrete Thinking
 - Knowledge/skill in bits and pieces
 - Emphasis on recognition rather than production of knowledge

When Testing Has A Lesser Role

- Discrete Test Items Sample From A Larger Valued Set of Proficiencies
- But These Test Items Do Not Embody All that We Mean by A High-Quality Education.
- Still, Inferences to the Domain are Valid



When Testing has a Larger Role

If the Stakes are Raised, Then the Content of Testing Begins to Define Educational Goals

- Instruction is Shaped Accordingly
- Test Scores Are No Longer Predictive of a Broader and More Significant Educational Achievement
 - Test scores lose their meanings
 - They no longer "sample" from the larger domain"

How to Respond?

- Possibly, Reduce the Extent of Testing
 - But there is growing accountability
- Possibly, Make Tests More Worthy of Their New Role
 - So that teaching to the test does good rather than harm

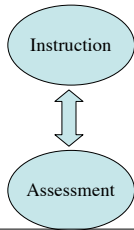


Features of New Assessments

- Alternative, Authentic, and Performance Assessment All Involve Tasks That Are Complex, Extended in Time, and Intrinsically Worthwhile
- The Emphasize Student Production
- It's Implicit That Students Are Thinking at a High Level
 - Critical thinking and problem solving rather than recognition and recall

Assessment in the Service of Instruction

- Blending of Assessment and Instruction
 - Classroom or formal assessment
- Teaching (Like Testing) Should Involve Diagnosis and Response
- Assessment (Like Teaching) Should Always Have Value for Learning





Where Do Assessments Come From?

- Goals and Objectives for Education Are the Prerogative of Multiple Stakeholders
- Standards Frameworks
 - National and state frameworks
 - Illustrative tasks
 - General Rubric
 - E.g., characteristics of a good essay

Creating the Assessment

- Creating the Assessment Framework (Specifications)
- Knowledge objectives by process dimensions
 - Knowledge: Planets, molecules, vertebrates
 - Process: Reasoning, representational forms, types of items (e.g., constructed response, multiple-choice).
- Fairness Entails Opportunity to Learn



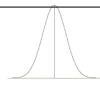


Ensuring Reliable Scoring

- Clear, Well-Defined, and Defensible Criteria
- Well-Trained Scorers
- Evidence of Inter-Rater Agreement
 - Document this
- Procedure of Resolving Score Divergence
 - A third scorer?
 - Beware of Rater Fatigue; Rater Differences
- Is Training of Scorers Professional Development?



Validity



- Validity is Concerned with the Meaning of What is Measured
- Validity is an Argument That a Test Really Measures What it Purports to Measure
- Construct Validity is the Most General Form
 - A *construct* is the psychological dimension measured
- Other Forms of Validity
 - Content: Established partly by test specifications
 - Predictive: Does the test relate to future success?
- Reliability Places an Upper Bound on Validity
 - If you can't measure it accurately, it can't be meaningful