

Unique Challenges and Opportunities in Assessment of 21st Century Skills

Steven L. Wise

Northwest Evaluation Association

Practical Measurement Issues

- Reliability & validity
- How are the scores to be used?
- Test stakes are important!
- Assessment test versus assessment program
- What is the practical limit on the amount of time we really have to assess students?

Reliability

- The consistency of scores from an assessment
- Size of standard error of measurement is a key issue.
- More important, perhaps, is classification decision consistency.

Validity

- The degree to which evidence and theory support the inferences made from the scores.
- Types of evidence: content domain coverage, correlations, mean differences
- If inferences are to be made at the student level, individual score validity is also important.

Individual Score Validity

- The extent to which a score is free from construct-irrelevant factors.
- Potential threats:
 - low stakes: test-taking motivation
 - high stakes: item/task exposure, familiarity with items/tasks, language proficiency, cheating, coaching, socially desirable responding, test anxiety, test-taking motivation

Item/Task Complexity

- Low complexity (multiple-choice; other highly-structured, easily scored items)
- High complexity (e.g., constructed response, oral presentations, performances, simulations)
- Practical Issues:
 - cost to develop
 - cost to score
 - domain coverage
 - scalability

Some Key Measurement-related Questions to Consider

- How scalable is the assessment to an assessment program?
- How vulnerable is the assessment to validity threats? Which validity threats are of greatest concern?
- How simple/complex are the standard setting needs?
- What are the costs of misclassifications of students?