Inclusive Next Generation Assessment: An Examination of Approaches for Evaluating Technical Quality

Edynn Sato, Ph.D.

Reidy Interactive Lecture Series
Cambridge, Massachusetts
October 21-22, 2010
Overview

• Research—Some Critical Points
  – Special student populations (SWDs, ELs)
  – Access
  – Validity Evidence

• Toward a Theory of Action for Inclusive Next Generation Assessment

• Technical Quality—Some Considerations

• Some Additional Considerations
What research supports: Some critical points

Special population students:

• Students with disabilities (SWDs) and English learner (EL) student populations are heterogeneous
  - e.g., EL students differ according to English language proficiency, native language proficiency, cultural proximity, time and consistency in U.S. schools, and U.S. learning and assessment experiences.
  - e.g., SWDs differ according to sensory, physical, cognitive, and/or linguistic capacities

• SWDs and EL students can learn complex, rigorous academic content

[See references at end of presentation.]
What research supports: Some critical points

Access:

• Access is arguably the most relevant threat to validity for SWDs and EL students.

• There are viable, systematic ways to address the range of students’ access needs such that students can fully demonstrate what they know and can do.

• Accommodations, Universal Design (UD), and Universal Design for Learning (UDL) are necessary but not sufficient for addressing the range of students’ access needs.

• Access should be considered and addressed throughout an assessment’s design, development, and implementation process.
What research supports: Some critical points

Validity Evidence:

• Evidence should help to examine interactions among student characteristics, access strategies, content, and assessment formats and their effect on the assessed construct, validity of interpretations of student outcomes, and consequences should be considered.

• Criteria for judging the technical quality of assessments of the general student population overlap with criteria relevant to special student populations; however, the criteria do not overlap completely.
Toward a Theory of Action for Inclusive Next Generation Assessment

• Extends knowledge and research of validity approaches (e.g., Messick’s Unified View of Validity, Kane’s Argument-based Approach, Mislevy’s Evidence-centered Assessment Design)

• Integrates emerging research on access and the assessment of special student populations

• While focused on special student populations, generalizes to all students—necessarily moves beyond UD, UDL, and targeted accommodations
Toward a Theory of Action for Inclusive Next Generation Assessment

- Supports an integrated, systemic approach to assessment design and development
- Applies to formative and summative assessments
- Necessary for computer-delivered assessments and innovative accessible tasks
- Serves as a “road map” for guiding the design, development, and implementation of inclusive next generation assessment
- Allows for multiple points of entry
Toward a Theory of Action for Inclusive Next Generation Assessment

Assessment Purpose(s) and Use(s)

Standards-based Assessment

- Design
  - Content quality
  - Technical quality

- Development
  - Training
  - Monitoring
  - Evaluation

- Implementation

Outcomes

Consequences

Access needs

Developmental research

Existing and needed

Situational and Socio-cultural context

- Sensory
  - Physical
  - Cognitive
  - Linguistic

- Knowledge, Skills, Abilities
  - Content
  - Language
  - Performance Expectations

Assessment Purpose(s) and Use(s)

Standards-based Assessment

- Design
  - Content quality
  - Technical quality

- Development
  - Training
  - Monitoring
  - Evaluation

- Implementation

Outcomes

Consequences

Access needs

Developmental research

Existing and needed

Situational and Socio-cultural context

- Sensory
  - Physical
  - Cognitive
  - Linguistic

- Knowledge, Skills, Abilities
  - Content
  - Language
  - Performance Expectations

Assessment Purpose(s) and Use(s)

Standards-based Assessment

- Design
  - Content quality
  - Technical quality

- Development
  - Training
  - Monitoring
  - Evaluation

- Implementation

Outcomes

Consequences

Access needs

Developmental research

Existing and needed

Situational and Socio-cultural context

- Sensory
  - Physical
  - Cognitive
  - Linguistic

- Knowledge, Skills, Abilities
  - Content
  - Language
  - Performance Expectations

Assessment Purpose(s) and Use(s)

Standards-based Assessment

- Design
  - Content quality
  - Technical quality

- Development
  - Training
  - Monitoring
  - Evaluation

- Implementation

Outcomes

Consequences

Access needs

Developmental research

Existing and needed

Situational and Socio-cultural context

- Sensory
  - Physical
  - Cognitive
  - Linguistic

- Knowledge, Skills, Abilities
  - Content
  - Language
  - Performance Expectations

Assessment Purpose(s) and Use(s)

Standards-based Assessment

- Design
  - Content quality
  - Technical quality

- Development
  - Training
  - Monitoring
  - Evaluation

- Implementation

Outcomes

Consequences

Access needs

Developmental research

Existing and needed

Situational and Socio-cultural context

- Sensory
  - Physical
  - Cognitive
  - Linguistic

- Knowledge, Skills, Abilities
  - Content
  - Language
  - Performance Expectations

Assessment Purpose(s) and Use(s)

Standards-based Assessment

- Design
  - Content quality
  - Technical quality

- Development
  - Training
  - Monitoring
  - Evaluation

- Implementation

Outcomes

Consequences
Technical Quality: Some Considerations Related to Inclusive Assessment

Given two assessment tasks developed to measure a particular construct, to what degree do they both measure the construct? (construct comparability)

• Alignment
  – Array of representations of the construct
    • Different semiotic representations
    • “Default, alternate, and supplemental content”
    • Content versus language
  – Effect of access strategy/strategies on the construct
    • Presentation, engagement, response, format/medium
    • Sensory, physical, cognitive, and/or linguistic
  – Effect of context on the construct
    • Situational
    • Socio-cultural
Technical Quality: Some Considerations Related to Inclusive Assessment

Do students from different subgroups attach the same meaning to the construct as a whole? (construct equivalence)

• Dimensionality
• Construct invariance
  – Non-monotonic increase
Do equivalent test scores have the same meaning for all students, regardless of their group membership? (measurement invariance)

- Selection bias
  - Different sample sizes of focal and reference groups
  - Different proficiency distributions between the focal and reference groups
- Dimensionality
- Variation at item level versus total score
- Flexibility versus standardization of task/administration
Some Additional Considerations

Inferences and consequences…

• Flexibility and standardization
• Targeted and non-targeted constructs
• Underestimation
• Under-representation
• Differential boost
  – How much and for whom?
• Incremental validity
  – How should this relate to possible marginal changes to the assessed construct and/or limited test equivalence?
References

References

For more information contact:
Edynn Sato
esato@wested.org