Theory of Action and Validity Argument in the Context of Through-Course Summative Assessment

Brent Bridgeman
Randy Bennett
Michael Kane

Educational Testing Service
bbridgeman@ets.org

Overview

- Two-part interpretive argument for through-course assessments
  - Measurement claims
  - Theory-of-action claims
- Determining weights for through-course components
- Psychometric quality issues
- Threats to validity for both measurement and theory-of-action arguments
Evaluating Claims for Through-Course Assessments

- Kane proposes two-part argument-based approach to validation
  - *Interpretive argument* lays out chain of inferences from observed performances to conclusions and decisions
  - *Validity argument* evaluates the evidence for the interpretive argument

- Measurement argument
  - Chain from observations (test items) to claims about mastery of the Common Core State Standards (CCSS)

- Theory-of-Action argument
  - Chain from test scores to enhanced performance by individuals (e.g., students and teachers) and institutions (e.g., schools and districts)

Measurement Argument

- Item scoring rules are appropriate and applied consistently
- Items are representative of the full CCSS domain
- Enough items to control sampling error
- Items combined across multiple occasions in a meaningful manner
Theory-of-Action Argument

• Example action mechanisms
  – Teachers
    • Review the CCSS and communicate key elements to students
    • Create assignments and formative tasks consistent with the CCSS (without just teaching to the test)
    • Review periodic feedback reports to adjust instruction for all students (struggling, meeting expectations, or excelling)
  – Schools and districts
    • Use results to identify teacher development needs
    • Hold teachers and administrators accountable
    • Use data to recruit and assign teachers

• Ultimate goal is more students prepared for college or career

Determining Weights for Through-Course Assessments

• Simple sum of 4 assessments is simplest, but may not be best
• For a clear learning progression, heavier weight on later tasks makes sense
• For separate skills with no clear learning progression (e.g., sentence punctuation and apostrophe rules), no reason to give more weight to skill learned later
Determining Weights for Through-Course Assessments (continued)

- Equal weights do not make sense if some assessments focus on skills that are clearly more important for college and career readiness
- Reliability of the composite can be improved by giving more weight to the more reliable components, BUT maximizing reliability may not maximize validity

Psychometric Quality Issues

- Parts of the theory of action set a low psychometric bar
  - focusing teacher and student attention on the standards
- Parts of the theory of action set a high psychometric bar
  - “Routine use of [through-course assessment] results to adjust instruction, including for students who are struggling, meeting expectations or excelling” requires highly reliable and appropriately focused instruments
  - Comparability across tasks
  - Comparability across raters
Threats to Validity of the Theory of Action

• Construct underrepresentation
  – “The overarching goal of the SBAC is to ensure that all students leave high school prepared for postsecondary success in college or a career”
  – “State leaders…share one fundamental goal: building their collective capacity to dramatically increase the rates at which students graduate from high school ready for school success” (PARCC)

• BUT only ELA and mathematics are assessed
  – Are these the only skills necessary for college and career success in the 21st century?

Threats to Validity of the Theory of Action (continued)

• Reliable and valid assessments do not imply that appropriate actions will be taken based on those assessments; both the assessments and the actions must be considered in the theory-of-action validity argument
What is the appropriate action given this score report?
Threats to Validity of the Theory of Action (continued)

- Test security
  - NCLB tests were high stakes for teachers but low stakes for students, hence little incentive for students to cheat (though there were some cases of teachers cheating)
  - Security may be more of an issue for SBAC and PARCC assessments for two reasons
    - Student level stakes are higher, such as for college placement, and possibly eventually for high school graduation or college entrance
    - Performance events that take place, at least in part, outside of the classroom may provide an opportunity for students to present the work of others as their own

Seven Recommendations

1. Focus validity argument not only on the assessments, but on the full theory of action
2. Be explicit in the statement of the theory of action
3. Take advantage of within-course assessment task types that cannot typically be included in an end-of-course assessment
4. Have a strong rationale for weights assigned to through-course components
Recommendations (continued)

5. Collect data from all key stakeholders
6. Begin data collection now so that changes in teaching and learning can be documented
7. Evaluate and control threats to validity of measurements (e.g., lack of comparability across tasks and raters) and threats to validity of proposed actions

Thank You