Constructing a Validity Argument for Alternate Assessments based on Modified Achievement Standards (AA-MAS)

Scott F. Marion, Center for Assessment

The contents of this publication were developed under cooperative agreement S283B050019 with the U. S. Department of Education. However, the contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.
The AA-MAS requires a more careful validity evaluation than one might undertake for either the AA-AAS or the general assessment.

- uncertain conceptual framework supporting this assessment initiative
- novelty of the enterprise
Framework

- Unified conception of validity, but using Kane’s (2006) argument-based approach to validity evaluation
  - Offers several pragmatic advantages over evaluations based on the construct model
    - primarily in terms of prioritizing studies
    - synthesizing the results of the various studies
  - At its simplest, Kane’s approach asks the evaluator to search for and evaluate all the threats to the validity of the assessment inferences
Why an Argument?

- Validity argument serves to:
  - organize studies,
  - provides a framework for analysis and synthesis, and
  - forces critical evaluation of claims using a falsification orientation.

- For example, “the AA-MAS is measuring grade-level knowledge and skills.” The content-related evidence then should include information that would allow one to challenge this grade-level claim if, in fact, the test was measuring below grade-level content.
Kane’s Argument-Based Framework

- **Interpretative argument** -- specifies the proposed interpretations and uses of test results by laying out the network of inferences and assumptions leading to the observed performances to the conclusions and decisions based on the performances,
  - essentially a mini-theory...the interpretative argument guides the data collection and methods for conducting the validity analyses. Most importantly, theories are falsifiable

- **Validity argument** provides an evaluation of the interpretative argument.
Theory of Action—As a Starting Point

- A more general theory of action can be a useful starting point for developing a more complete interpretative argument.
  - Really a simplified interpretative argument that requires the explication of the intended components of an assessment and decision system, as well as the mechanisms by which a test user could reasonably expect to get from one step to the next.

- A well-developed theory of action for AA-MAS is perhaps even more critical than it might be for other validation initiatives.
  - Policy makers, developers, stakeholders, and technicians should have to very explicitly lay out why they think that implementing an AA-MAS will lead to improved educational opportunities for eligible students.
An example theory of action

Same academic content standards as the general assessment

Essentially same blueprint as general assessment with certain modifications

The assessment is designed to measure grade-level content and high achievement expectations accurately.

Achievement standards include recognition of specific supports, but focus on high expectations.

The AA-MAS reinforces appropriate instructional and formative assessment strategies for use in classrooms/schools.

Teachers provide instruction aligned with academic expectations.

Student scores on the AA-MAS provide an accurate estimate of what students know.

Test performance is used by educators to provide more appropriate supports and programs.

Improved student/school performance on the AA-MAS leads to higher accountability scores.
Prioritizing Validity Evaluation Questions

- The prioritization should be influenced by the philosophy and orientation guiding the program.

- Following Kane (2006), the state should not select questions and design a validity evaluation to confirm their guiding philosophy. Rather, the validity evaluator should purposefully design studies to contradict the states’ beliefs and claims.
Classes and Sources of Evidence

- **Joint Standards** (AERA, et al, 1999) are a good starting point. Evidence based on:
  - Content
  - Internal structure
  - Response processes
  - Relations with other variables
  - Consequences

- But, this is not enough, especially for AA-AAS
  - Who are the students?
  - How do they learn?

- Maps onto of **Assessment Triangle** (Pellegrino, et al, 2001)
Additional focus on consequences

- Accountability function of AA-MAS
  ...consequences of such policy actions must be incorporated into the validity evaluation

- Many intended positive consequences, but some serious potentially unintended negative consequences (e.g. lower expectations) with AA-MAS
  - Example proposition “The increase in the percentage of special education students scoring proficient as a result participating in the AA-MAS has not led to an increase in schools falsely meeting AYP targets (Type II errors).”
Synthesis & Evaluation

- Individual pieces of evidence (i.e. studies) do not make an assessment system valid or not...
- The evidence and logic must be synthesized to evaluate the interpretative argument.
- The evaluative argument provides the structure for evaluating the merits of the interpretative argument (Kane, 2006).
- Major challenge: new evidence being collected along the way, while our understanding of alternate assessments and the students they serve evolves much more rapidly than in many other programs.
  - will require evaluators to (re)examine evidence in light of these newer understandings.
Dynamic Evaluation

- Evaluators rarely have all the evidence in front of them to make conclusive judgments at a single point in time.

- In current context, evaluators do not have the luxury of concluding, “The system is not working; let’s start over.”
  - State leaders and test developers must act as if the dynamic results were part of a formative evaluation…search for ways to improve the system.
  - Evidence might be so overwhelmingly stacked against the intended claims… only option might be starting over.

- The state should use the guiding principles and purposes of the AA-MAS to determine how to weigh various sources of evidence to arrive at an evaluative judgment.