

Putting the pieces together: Validity evaluations of states' alternate assessment systems

Scott Marion, Center for Assessment

Jacqui Kearns, NAAC/Univ of KY

Melissa Fincher, GA Dept of Education

Janet Stuck, CT Dept of Education

Derek Briggs, Univ of Colorado

CCSSO National Conference on Student Assessment

June, 2008

Validation is “a lengthy, even
endless process”

(Cronbach, 1989, p.151)

The NAAC-GSEG

- Builds off of previous projects (NHEAI and first phase of NAAC) that focused on technical documentation
- Collecting and evaluating validity evidence was a significant challenge
- This project focuses on having states (5) create and evaluate validity arguments
- Our focus is NOT on single studies or even sets of studies, but on how the studies support or refute a validity argument

Five-State Consortia

- Begin building validity arguments for their alternate assessments based on alternate achievement standards
 - By convening a panel of measurement and special education experts, with a purposeful overrepresentation of well-known validity theorists,
 - Guiding states in developing a validity evaluation plan and constructing a validity argument.
 - Considering the maturity of their system,
 - the assessment format, and
 - Other unique factors related to assessing this diverse population

Intended Outcomes for the Consortia Project

- Demonstrate through our partnerships with states high quality validity evaluation models,
- Provide models of validity-based technical documentation for AA-AAS and eventually for AA-MAS and general assessments,
- Add to the growing research base on high quality, technically sound AA-AAS to provide technical assistance to states as they endeavor to conduct their validity evaluation studies, and
- Provide a range of research-to-practice products that explicate the process and results.

- But what is a validity argument and how do we evaluate the validity of our inferences?
- In other documents (NHEAI/NAAC, Marion & Pellegrino, 2006) we outlined the history of validation studies from the content to the construct model through Kane's argument-based approach...

Kane's argument-based framework

- "...assumes that the proposed interpretations and uses will be explicitly stated as an argument, or network of inferences and supporting assumptions, leading from observations to the conclusions and decisions. Validation involves an appraisal of the coherence of this argument and of the plausibility of its inferences and assumptions (Kane, 2006, p. 17)."

Two Types of Arguments

- An interpretative argument specifies the proposed interpretations and uses of test results by laying out the network of inferences and assumptions leading from the observed performances to the conclusions and decisions based on the performances
- The validity argument provides an evaluation of the interpretative argument (Kane, 2006)

The Interpretative Argument

- Essentially a mini-theory—the interpretative argument provides a framework for interpretation and use of test scores
- Like theory, the interpretative argument guides the data collection and methods and most importantly, theories are falsifiable as we critically evaluate the evidence and arguments

- One of the most effective challenges to interpretative arguments (or scientific theories) is to propose and substantiate an alternative argument that is more plausible
 - With AA-AAS we have to seriously consider and challenge ourselves with competing alternative explanations for test scores, for example...
 - “higher scores on our state’s AA-AAS reflects greater learning of the content frameworks” OR
 - “higher scores on our state’s AA-AAS reflects higher levels of student functioning” OR
 - “higher scores on our state’s AA-AAS reflect greater understanding by the teachers on how to gather evidence or administer the test”

More Simply...

- Following Shepard (1993, p. 429)...
 - What does the testing practice claim to do;
 - What are the arguments for and against the intended aims of the test; and
 - What does the test do in the system other than what it claims, for good or bad?
- Test validation is basically the process of offering assertions about a test or a testing program and then collecting data and posing logical arguments to refute those assertions
 - If the assertions cannot be refuted, we can say that they are tentatively supported (and that's the best we can do!)

Criteria for Evaluating Interpretative Arguments

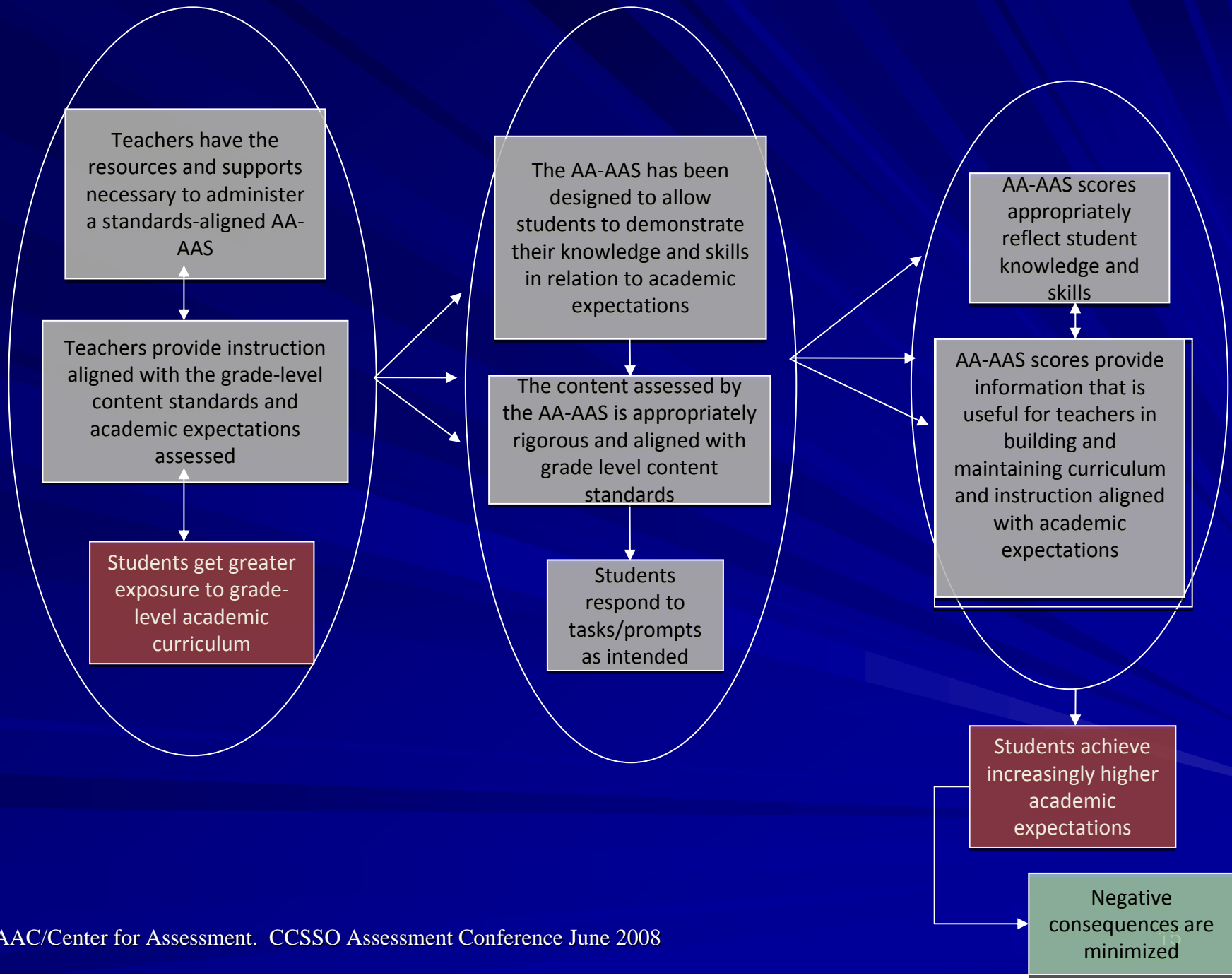
- Clarity—should be clearly stated as a framework for validation. Inferences and warrants specified in enough detail to make proposed claims explicit.
- Coherence—assuming the individual inferences are plausible, the network of inferences leading from the observations to conclusions and decisions make sense
- Plausibility—particularly of assumptions, are judged in terms of all the evidence for and against them

Values and Consequences

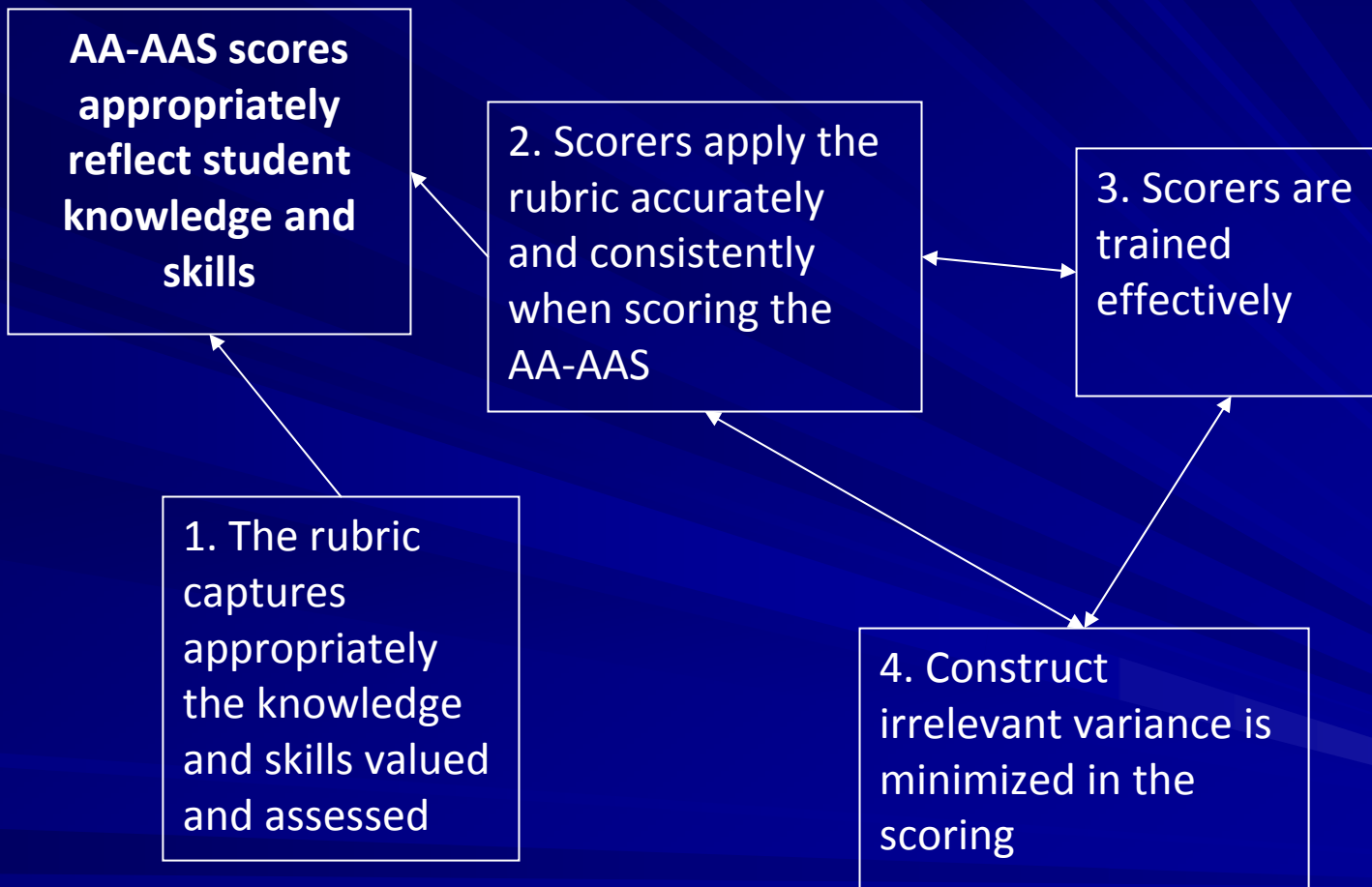
- Evaluating a decision procedure requires an evaluation of values and consequences
- “To evaluate a testing program as an instrument of policy [e.g., AA-AAS under NCLB], it is necessary to evaluate its consequences” (Kane, 2006, p.53)
- Therefore, values inherent in the testing program must be made explicit and the consequences of the decisions as a result of test scores must be evaluated.

Getting Started

- Katherine Ryan (2002) and others have suggested that laying out a more general “theory of action” is a useful starting point for developing a more complete validity argument
- My colleague, Marianne Perie, and I created the following EXAMPLE theory of action for an alternate assessment system...



Propositions underlying a single claim



Validity Evaluation

- As part of this project, we have asked the states to create a validity argument that is bigger than the specific expectations of this project
- Validity evaluation plan reflecting prioritization
- Undertake at least three studies
- Begin synthesizing study results to evaluate the validity argument

Evaluating the Validity Argument

- Haertel (1999) reminded us that the individual pieces of evidence do not make the assessment system valid or not, it is only by synthesizing this evidence in order to evaluate the interpretative argument can we judge the validity of the assessment program.

- We will now hear from Melissa Fincher and Janet Stuck from Georgia and Connecticut, respectively, who are leading their state's efforts to develop and evaluate their validity arguments
- Derek Briggs, from the University of Colorado, Boulder and a member of our expert panel, will offer some commentary and sage advice