

The image features a vertical strip on the left side containing a repeating pattern of the University of Massachusetts (UMASS) logo. Each logo consists of a stylized red 'U' with a white outline, positioned above the word 'UMASS' in a white, sans-serif font. The background of the slide is black.

A Theory of Action for Test Validation

Stephen G. Sireci

***Center for Educational Assessment
University of Massachusetts Amherst***

**Presentation for the 13th Annual Maryland
Assessment Conference**

“The Next Generation of Testing”

November 8, 2013

University of Maryland, College Park

Validity

- **A concept that has evolved and is still evolving**
- **The most important consideration in educational and psychological testing**
- **Simple, but complex**
 - **Can be misunderstood**
 - **Disagreements regarding what it is, and what is important**
 - **Disagreements over what needs to be validated.**

Purposes of this presentation

- Provide a brief description of the evolution of validity, focusing on the *Standards for Educational & Psychological Testing*
- Describe the *Standards'* validation framework
- Discuss “Theories of Action” (TOA) for testing programs
- Discuss how TOAs can be used to provide a framework for validating an assessment *program*.

Evolution of Validity: The AERA/APA/NCME *Standards*

Edition	Validity
1954	Construct, concurrent, predictive, content
1966	Construct, criterion-related, content
1974	Construct, criterion-related, content
1985	Unitary (but, content-related evidence, etc.)
1999	Unitary: 5 sources of evidence

The *Standards* (1999) succinctly defined validity

“Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests.” (p. 9)

Note the focus on test USE

What guidance does the *Standards* give us for validation research?

- **Five “sources of evidence that might be used in evaluating a proposed interpretation of test scores for particular purposes”**

Standards' Validation Framework

Validity evidence based on

- 1. Test content**
- 2. Response processes**
- 3. Internal structure**
- 4. Relations to other variables**
- 5. Testing consequences**

These sources will be retained in the (2014?) version.

With respect to *validation*, the 1999 *Standards* took a practical stance

- **based on the “argument-based approach” to validity.**
 - **Cronbach (1971, 1988)**
 - **Kane (1992, 2006, 2013)**

The *Standards* essentially endorsed the argument-based approach to validation:

“Validation can be viewed as developing a scientifically sound validity argument to support the intended interpretation of test scores and their relevance to the proposed use.”

(AERA et al., 1999, p. 9)

The argument-based approach is a compromise between sophisticated validity theory and the reality that at some point, we must make a judgment about the defensibility and suitability of use of a test for a particular purpose.

But what the *Standards* and validity literature also tell us

- **Validation must focus on testing purpose and use.**
- **And based on these two points (focus on purpose, use the *Standards'* framework), we have clear and simple guidance for validation**

Before May 2012, I argued the AERA et al. (1999) *Standards* could be used as a validation framework.

Why?

The *Standards* as a validation framework:

- **Provide a system for categorizing validity evidence so a coherent argument can be developed.**
- **Provide a way of standardizing the reporting of validity evidence.**
- **Focus on both test construction and test score validation.**
- **Emphasize the importance of evaluating consequences.**

Sireci (2012, 2013)

- **Validation can be viewed as a 5-step process.**

Validation Steps

- 1. Identify testing purposes**
 - Should be explicitly stated in technical manuals and official documents/web sites of testing agencies.
- 2. Identify potential test misuse**
- 3. Prioritize validity questions based on explicit purposes and potential misuse**

Validation Steps (continued)

- 4. Determine sources of evidence needed for each purpose**
- 5. Cross validity questions with sources of evidence**

Validation Steps (1)

1. Identify testing purposes

- Should not be hard to do—they are explicitly stated in technical manuals and official documents/web sites of testing agencies!**

2. Identify potential test misuse

3. Prioritize validity questions based on explicit purposes and potential misuse

Validation Steps (2)

- 4. Determine sources of evidence needed for each purpose**
- 5. Cross validity questions with sources of evidence**

MA Adult Proficiency Test: Validity FRAMEWORK

Purpose/Validity Question	Source of Validity Evidence				
	Content	Internal Structure	Relations w/ Ext. Variables	Response Processes	Testing Consequences
Measure correct skills?					
Congruent w/ frameworks?					
Accurate?					
Measure progress?					
Meet NRS requirements?					
Useful for program Evaluation?					
Inappropriate diagnostic use?					
Inappropriate placement?					
Effect on instruction?					

MA Adult Proficiency Test: Validity EVIDENCE

Purpose/Validity Question	Source of Validity Evidence				
	Content	Internal Structure	Relations w/ Ext. Variables	Response Processes	Testing Consequences
Measure correct skills?	√		√		
Congruent w/ frameworks?	√				
Accurate?		√	√	√	
Measure progress?	√	√	√		
Meet NRS requirements?	√	√			
Useful for program Evaluation?	√	√			
Inappropriate diagnostic use?					
Inappropriate placement?					
Effect on instruction?					

By reviewing validity evidence for the MAPT, we can see

- **No validity evidence based on testing consequences**
- **No validity evidence to support potential negative effects or positive effect on instruction**
- **Are we proud of this? No.**
- **But we know what our next steps are.**
 - **Currently surveying teachers**

I was happy with this simplified framework until I was asked to help Smarter Balanced develop a “comprehensive research agenda” in June 2012.

- **The problem was evident with Step #1, “Identify Testing Purposes.”**

Many testing programs, such as the 4 major consortia,

- **have Theories of Action (TOA) that go beyond using tests to make inferences at the *student* level.**
- **These TOAs are designed to affect change at multiple levels.**

Race to the Top RFP

“In determining the extent to which the consortium’s proposed governance structure will enable the successful design, development, and implementation of the proposed assessment system, we will consider—

Race to the Top RFP

(a) The consortium's vision, goals, role, and key deliverables (e.g., assessment components, scoring and moderation system, professional development activities), and the consistency of these with the consortium's THEORY OF ACTION;"

RttT Scoring Criteria

5 points for the Theory of Action

The extent to which the applicant's theory of action is logical, coherent, and credible, and will result in improved student academic outcomes...we will consider the description of, and rationale for—

- (a) Each component of the proposed assessment system and the relationship of the component to other components in the system;
- (b) How the assessment results produced by each component will be used;
- (c) How the assessments and assessment results will be incorporated into a coherent educational system (*i.e.*, a system that includes standards, assessments, curriculum, instruction, and professional development); and
- (d) How the educational system as a whole will improve student achievement and college- and career-readiness

If you're like me, you're thinking...

- **What is a Theory of Action? and,**
- **How do we validate that?**

Bennett (2010)

“Theory of action is a common notion in the program evaluation literature... [and] is closely associated with *logic model*”

So, to validate a TOA

- **We go beyond validating use of a test for a particular purpose**
 - which starts with articulating testing purposes
- **And validate the assessment *program***
 - which starts with articulating claims and goals of the *program*

How do we incorporate TOA into a validity argument

- **Not much work or guidance in the area of validating TOA.**
 - **Bennett (2010)**
 - **Bennett, Kane, & Bridgeman (2011)**
 - **Sireci (2013)**

Bennett (2010)

- **Theory of Action for an assessment system should include**
 - **Intended effects of system**
 - **Components and rationale for each**
 - **Interpretive claims from assessment results & action mechanisms**
 - **Potential unintended negative effects and what will be done to guard against them**

Bennett, Kane, Bridgeman (2011)

Suggest a two-part “interpretive argument”

- 1. Measurement argument**
 - 2. Theory of action argument**
- TOA argument “focuses on use of the assessments to enhance individual or *institutional* performance” (p. 3).**

(Selected) Bennett et al. (2011) recommendations

- 1. Focus validity argument on assessments and TOA**
- 2. Be explicit in TOA—state action mechanisms and intended effects**
- 3. Collect data from key stakeholders: how are assessments *used*?**
- 4. Collect data *now* (to assess change)**

Bennett (2010, p. 82)

- **Is TOA logical, coherent, and scientifically defensible?**
- **Was assessment system implemented as designed?**
- **Were interpretive claims empirically supported?**
- **Were intended effects on individuals and institutions achieved?**
- **What negative effects occurred?**

From the little, but excellent, research that is out there, we have good initial guidance for incorporating a TOA into a validation endeavor.

So, to validate a TOA

- **We go beyond validating use of a test for a particular purpose**
 - which starts with articulating testing purposes
- **And validate the assessment *program***
 - which starts with articulating claims and goals of the *program*

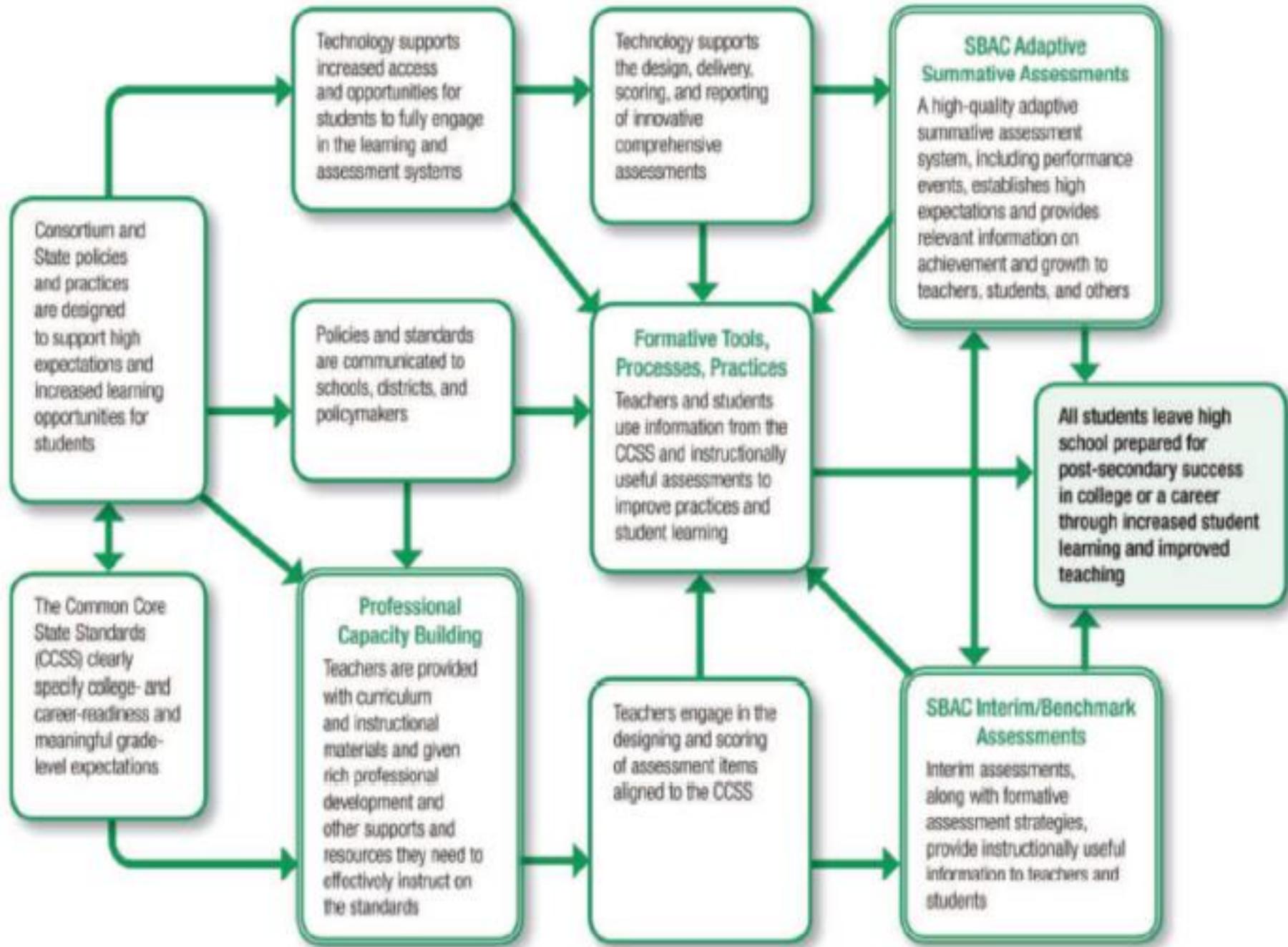
Proposed Smarter Balanced Validity Research Agenda

- **Available from Smarter Balanced**
 - **The next few slides describe how it was developed**

Smarter, Balanced Validation

- **Step #1: “Translate” claims in TOA to “purpose statements” to be the focus of validation.**
 - **For Smarter Balanced, this process took about 5 months**
 - **Smarter Balanced leadership, TAC, Test Administration & Design Work Group, Validation & Psychometrics Work Group**

Smarter Balanced Theory of Action



Examples of TOA claims

- **“Policies and standards are communicated to schools, districts, and policymakers.”**
- **“Teachers are provided with curriculum and instructional materials and given rich professional development...”**

Smarter Balanced Comprehensive Research Agenda: Steps

- 1. Isolate components of TOA that should be translated into purpose statements for validation.**
- 2. Determine which fit a traditional validation paradigm and which pertain to program evaluation.**
- 3. For those fitting traditional model, cross testing purposes and potential misuses with *Standards'* 5 sources of validity evidence.**

Smarter Balanced Comprehensive Research Agenda

- 4. Design studies to evaluate purposes that relate to efficiency, fairness, quality, and success of the system (program evaluation)**
 - These studies are those more commonly found in program evaluation (e.g., audits of procedures, observations, interviews, focus groups, usage studies, surveys).**

Smarter Balanced Comprehensive Research Agenda

5. Prioritize studies based on their importance, how much information they will contribute to the validity argument, and the relative importance of the purpose(s) they will support.

Examples of Smarter Balanced Purpose Statements for Validation

- **Split up by components:**
 - **Summative (7)**
 - **Interim (4)**
 - **Formative Assessment Resources (5)****(16 total)**

**Purposes of the Smarter Balanced
Summative assessments are to provide
valid, reliable & fair information about,**

- 1. students' ELA and Math achievement with respect to those CCSS measured by the ELA & Math summative assessments**
- 2. whether students < Grade 11 demonstrated sufficient academic proficiency in ELA & math to be on track for college readiness**
- 3. whether Grade 11 students have sufficient academic proficiency in ELA and Math to be ready to take credit-bearing college courses ...(next 4 not shown)**

Purposes of the Smarter Balanced *Interim* assessments are to provide valid, reliable & fair information about,

- 1. student progress toward mastery of the skills measured in ELA & Math by the summative test**
- 2. students' performance at the content cluster level so teachers & administrators can track student progress throughout the year and adjust instruction accordingly**
- 3. individual and group (e.g., school, district) performance at the claim level in ELA and math to determine whether teaching/learning are on target**
- 4. student progress toward the mastery of skills measured in ELA & Math *across all students & subgroups of students***

Purposes of the Smarter Balanced *Formative Assessment Resources* are to provide measurement tools and resources to,

- 1. Improve teaching and learning**
- 2. Monitor student progress throughout the year**
- 3. help teachers and other educators align instruction, curricula, and assessment**
- 4. help teachers & other educators use the Summative and Interim assessments to improve instruction at the individual student and classroom levels**
- 5. illustrate how teachers and other educators can use assessment data to engage students in monitoring their own learning**

Smarter Balanced Research Agenda

- **55 research studies recommended**
- **Many part of standard test development and validation practices**
- **But many were not**

The purposes of the Smarter Balanced Interim assessments are to provide valid, reliable and fair information about,	Source of Validity Evidence				
	Content	Internal Structure	Relations w/ Ext. Variables	Response Processes	Testing Consequences
1. student progress toward mastery of the skills measured in ELA and Mathematics by the summative assessment	✓	✓		✓	
2. students' performance at the content cluster level so teachers and administrators can track student progress throughout the year and adjust instruction accordingly	✓	✓			✓
3. individual and group (e.g., school, district) performance at the claim level in ELA and mathematics to determine whether teaching and learning are on target		✓	✓		✓
4. student progress toward the mastery of skills measured in ELA	✓	✓	✓	✓	✓

So, how do we incorporate TOA into a validity argument?

- **Not much work or guidance in the area of validating TOA.**
 - Bennett (2010)
 - Bennett, Kane, & Bridgeman (2011)
 - Sireci (2013)
- **Based on these references and recent experience, suggestions follow.**

Summary: How to incorporate TOA into validity argument

- 1. Evaluate TOA and isolate components that should be translated into purpose statements for validation**
 - Evaluate coherence of TOA
 - Specify claims of testing system at individual and institutional levels
- 2. Determine which purposes fit into test score use/interpretation, and which should be validated using a program evaluation approach**

Summary: How to incorporate TOA into validity argument

3. Cross testing purposes related to score use with *Standards'* five sources of validity evidence.

- Also consider potential test misuse

4. Design studies to evaluate efficiency, fairness, quality, and success of *system*.

- i.e., for purposes related to the system
- Gather information from stakeholders who represent institutional levels

Summary: How to incorporate TOA into validity argument

5. Prioritize studies based on (a) importance of purpose and (b) expected contribution to validity argument

Steps 1, 3, and 5 are not new, Steps 2 and 4 are.

21st Century Validation

- 1. Focuses on test use.**
- 2. Requires evidence test measures what it “purports” to measure**
- 3. Requires evidence of test “utility.”**
- 4. Requires evidence test is doing more good than harm.**
- 5. Includes focus at individual and institutional levels**

Validation Past & Present vs. Future

- **Validity evidence based on testing consequences is the bridge between validating the use of an assessment for a particular purpose, and validating an assessment system.**
 - **Must consider (validate) derivative measures (SGPs, VAEs, etc.)**

Haertel (2013)

- **Measurement has its roots in psychology, which focuses on individual differences**
- **Colleagues in sociology, anthropology, economics, law, etc., may be helpful for moving to assessment of consequences at institutional level**

Tips for incorporating TOA into validity argument

- 1. Start data collection as early as possible (before implementation?)**
- 2. Validation (evaluation) should be formative, as well as summative**
- 3. For validation of institutional claims, program evaluation techniques should supplement 5 sources of validity evidence**
 - See Bennett's (2010) research questions**

Tips for incorporating TOA into validity argument

- 4. Validity evidence based on testing consequences is particularly important, and is the bridge between validating the use of an assessment for a particular purpose, and validating an assessment system.**

Closing remarks

- **21st century educational assessment offers exciting new ways of assessing students' knowledge and skills**
- **Use of educational assessments in the 21st century presents validation challenges**
- **But we can do it!**
 - **Incorporating TOA into validity argument is new first step.**

- **Thanks to UMD for the invitation,
and to you for your attention!**

Sireci@acad.umass.edu