Science Assessment in Minnesota

Session III: Political and Practical Considerations in Implementing Innovative Assessments

Dirk Mattson, Ph.D.
Director, Research and Assessment

Workshop Series on Best Practices in State Assessment
Workshop 1

December 10-11, 2009

Overview of the Science Tests

- The Minnesota Comprehensive Assessments-Series II (MCA-II) fulfill the testing requirements under No Child Left Behind.
- In reading and math, our first generation (Series I) were used under IASA.
- The science tests were built specifically in response to NCLB: thus, they began with Series II with design efforts starting in 2004.

How They Work

- Designed as scenario-based, computer-delivered.
- Present students with realistic representations of classroom experiments and real-world phenomena.
- Sets items within a scenario by common context to give students the opportunity to consider science content at a higher cognitive level than would be possible with stand-alone items.
- Include graphic, audio and/or video media.
- Include various item types:
  - multiple choice (MC),
  - figural response (FR) in which the student manipulates graphic elements of an item, to select a point on a graphic or to complete a graph or diagram.
  - short constructed response (SCR) or
  - extended constructed response (ECR)

Minnesota Comprehensive Assessments-Series II in Science

A Demonstration

The Genesis

- Predecessor met with lead state science teachers on Day 2 of his job.
- These teachers did not want a multiple-choice test exclusively - wanted something that reached deeper to better represent science inquiry.
- State had just experienced a contentious end to statewide performance assessments. Thus, this option was off the table.
- Two ways to move toward stakeholder desires:
  1. Science kits – with the mess and challenge that this created, plus it was not novel and moving forward, or
  2. A computer-based test that used the technology to assess science knowledge in new ways
- Greatest concern came from administrators regarding school capacity.

A Brief History

2004: New standards and repeal of teacher-level performance assessments measuring old standards

2005: Conceptual design of science test begins

2006: Proof of concept in small number of school districts

2007: Statewide field testing

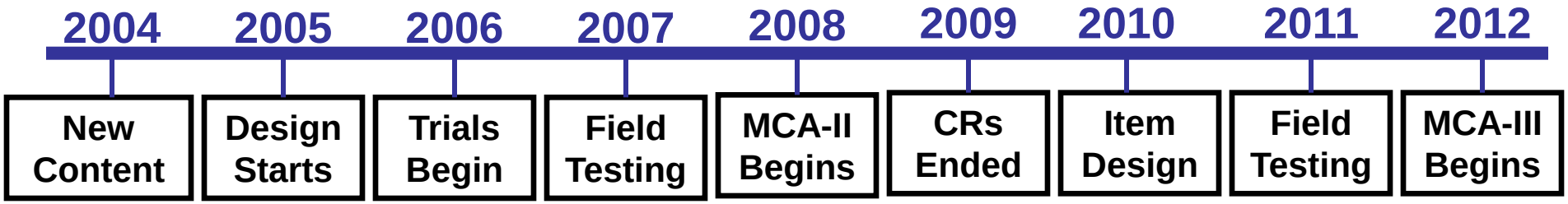
2008: First operational assessment and standard setting PLUS Revised standards under development

2009: MN Legislature forbids use of state funds for use in human scoring of constructed-response items

2010: Item design for MCA-III

2011: Field-testing of MCA-III

2012: Operational MCA-III



Revision Timeline (M.S. 120B.023)

	Mathematics	Science	Reading/Language Arts
2006-2007	**Standards Revision**		
2007-2008	MCA-III Test Specifications		
2008-2009	MCA-III Item Development	**Standards Revision**	
2009-2010	MCA-III Items Field Tested	MCA-III Test Specifications & Item Development	**Standards Revision**
2010-2011	**MCA-III Operational (Grades 3-8)**	MCA-III Items Field Tested	MCA-III Test Specifications & Item Development
2011-2012		**MCA-III Operational**	MCA-III Items Field Tested
2012-2013			**MCA-III Operational**
2013-2014	**Something Operational High School (?)**		
	Standards revision in 2015-2016	Standards revision in 2017-2018	Standards revision in 2018-2019

Pending ...
1. Common Core Standards ?
2. Race to the Top Grant Application ?
3. Race to the Top Assessment Consortia ?
4. ESEA Reauthorization ?

The Journey of a Thousand Miles …

- Some initial challenges:
  - How do you craft the Request for Proposals? Will test publishers recognize what we want? (Will we?)
  - How do we determine what fair cost for services is?
  - How do we fund the initial research & development efforts?

- Some initial supports:
  - Chief's office
  - Science teacher community in Minnesota
  - Grant funding

Designing the Unknown

- What is this supposed to look like?
  - Nobody had done this before.
- What were the design constraints when we first started?
  - Shoot the moon: very sophisticated animation
  - Content specialists could often out-think the computer designers
- Where did we start?
  - Developed templates using universal design to facilitate programming and minimize “play factor”
  - Involved programmers early to make sure the item can be created
- Which platform are we designing this for?
  - PC native program, but Minnesota is Mac heavy

Preparing for Administration

- "Not blinking" in the rumor that we really weren't going to computer
- The Department of Corrections not allowing any adjudicated juvenile to use the computer
- A checkered history of computerized testing up to spring of 2008 (i.e., spring 2007)
- Preparing schools to load 34 MB – 53 MB times 9 forms (total of approximately 300-475 MB) of test content to their servers
- Getting every school to demonstrate they were ready to administer these tests in a state with no infrastructure standards for technology
- Designing a test security protocol for a limited number of operational forms

Looking Back

- Still encountered issues in Year 1
  - Adobe released a new update of Flash approximately one week before the testing window
  - Kids click *everywhere*: There is no maximum of QC review that is satisfactory.
- Statewide Assessment Technology Work Group (SATWG)
  - Helped us define the issues
  - Helped us pilot software fixes in the field
- Student engagement
  - Anecdotal comments of high student engagement (even high school)
  - Very few omits

Lessons Learned

1. Sometimes no one has learned the lesson before you: “Sin Boldly”
2. Shoot the moon, but “play where I can see you”
3. Predict accurately the next technology that is coming
4. Establish your SATWG early: they can be your closest allies
5. Seek out “standardized flexibility”
6. Scenario-based assessment is a blessing and a curse

Looking to the Future

- Stakeholders are demanding that the computer assessment deliver results immediately: computer-scored CRs
- Revisioning scenarios: stand-alone items or all scenario-based?
- Implementing STEM requirements in the science tests: technology lets us go there more quickly
- Developing the “2.0” version of these assessments using simulations
  - How do we allow them to simulate yet make it summative?
  - Is the simulation the “new scenario”? Or is it a part of the scenario (i.e., the “sim-erio”)

Simulation Thoughts for Future

- Investigation
  - Student is involved in an experiment that requires manipulation of materials/variables
  - During the process of the investigation, quantitative data is generated and recorded
  - Questions tend to focus on experimental design, data analysis, tools, developing questions, conclusions.
- Phenomena
  - Building something (e.g., ecosystem)
  - Watching/interacting with scene
  - Qualitative data may be generated
  - Observation of natural event
  - Items lend themselves to engineering, content strands, Inquiry