

Assessment Reform in the United States: Implications for NH, Your District, and School



SCOTT MARION

**NATIONAL CENTER FOR THE IMPROVEMENT OF
EDUCATIONAL ASSESSMENT (DOVER, NH)**

**PRESENTATION TO THE NHSAA/NHASEA BEST
PRACTICES CONFERENCE ON EDUCATION FOR ALL
CHILDREN**

SEPTEMBER 24, 2010

Overview of Today's Talk

2

- Introduction to the Race to the Top (RTTT) Assessment Program
 - Intentions, Requirements, Process, Consortia
 - ✦ Many slides, but we'll go fast and you can read later
- The Consortia
 - Similarities and differences
 - Challenges and opportunities
 - Choices for NH
- Implications for NH Schools
- Questions and Answers (I hope)



Disclosure

3

- My colleague, Brian Gong, and I served as measurement experts for USED during the process
- The Center for Assessment worked (*pro bono*) on all three consortium proposals



The RTTT Assessment Program

4

- The RTTT Assessment Program is a \$350 million set aside of the \$4.3 billion RTTT program
 - Part of the 2009 American Reinvestment and Recovery Act (ARRA) of 2009
- The program was authorized to fund two types of awards:
- Comprehensive assessment systems (Category A)
 - Up to two awards of up to \$160 million each
- High school course assessments (Category B)
 - Up to one award of up to \$30 million



Goals and Intentions of the Program

5

- Broadly speaking, the intentions of the U.S. Department of Education (USED) were to:
 - Increase the rigor/complexity of state assessments as a driver to increase student learning
 - Avoid “reinventing the wheel” by having 50 states have to go through very similar processes and to try to capitalize on economies of scale (some naiveté)
 - Increase the comparability of assessment results across states to increase “transparency” and to increase expectations across all states



Purposes and Uses

6

- Assessments are validated for specific purposes and uses (i.e., assessments are not valid for all potential uses just because they are valid for one use)
- One of the truisms in educational measurement is that tests cannot serve more than a limited number of purposes (e.g., 1-2) well
 - Be very suspicious of anyone promising tests that serve multiple purposes well
- In their initial statements about the RTTT assessment program issued a “laundry list” of purposes, but they listened well (sort of)....



Comprehensive Assessment System Priorities

7

- *Absolute Priority: Comprehensive Assessment Systems Measuring Student Achievement Against Common College- and Career-Ready Standards*
- *Competitive Preference Priority: Collaboration and Alignment with Higher Education*



Category A Absolute Priority details...

8

- (a) Measures student knowledge and skills against a common set of college- and career-ready standards in mathematics and English language arts in a way that—
- (i) Covers the full range of those standards, including standards against which student achievement **has traditionally been difficult to measure**;
 - (ii) As appropriate, **elicits complex student demonstrations or applications** of knowledge and skills;
 - (iii) Provides an accurate measure of student achievement **across the full performance continuum**, including for high- and low-achieving students; and
 - (iv) Provides an **accurate measure of student growth** over a full academic year or course;



Category A Absolute Priority details...(continued)

9

- (b) Consists of assessment components in mathematics and in English language arts that include, for each subject, one or more summative assessment components that—
- (i) Are administered at least once during the academic year in grades 3 through 8 and at least once in high school; and
 - (ii) Produce student achievement data and student growth data (both as defined in the NIA) that can be used to determine whether individual students **are college- and career-ready or on track to being college- and career-ready**
- (c) Assesses all students, including English learners and students with disabilities; and
- (d) Produces data, including student achievement data and student growth data, that can be used to inform—
- (i) Determinations of school effectiveness for purposes of **accountability under Title I** of the ESEA;
 - (ii) Determinations of **individual principal and teacher effectiveness** for purposes of evaluation;
 - (iii) Determinations of **principal and teacher professional development and support needs**; and
 - (iv) **Teaching, learning, and program improvement.**



Category A Competitive (IHE) Priority details...

10

- (a) Commits the IHE or IHE system to participate with the consortium in the design and development of the consortium's final high school summative assessments in mathematics and English language arts in order to ensure that the assessments measure college readiness;
- (b) Commits the IHE or IHE system to implement policies, once the final high school summative assessments are implemented, that exempt from remedial courses and place into credit-bearing college courses any student who meets the consortium-adopted achievement standard for each assessment and any other placement requirement established by the IHE or IHE system; and
- (c) Is signed by the State's higher education executive officer (if the State has one) and the president or head of each participating IHE or IHE system.



Category B Absolute Priority details...

- (a) For each course in the assessment program—
 - (i) Measures student knowledge and skills against standards from a common set of college- and career-ready standards (as defined in the NIA) in subjects for which such a set of standards exists, or otherwise against State or other rigorous standards;
 - (ii) As appropriate, elicits complex student demonstrations or applications of knowledge and skills;
 - (iii) Produces student achievement data (as defined in the NIA) and student growth data (as defined in the NIA) over a full academic year or course that can be used to inform—
 - (A) Determinations of individual principal and teacher effectiveness and development and support needs; and
 - (B) Teaching, learning, and program improvement; and
 - (iv) Is designed to assess the broadest possible range of students, including English learners (as defined in the NIA) and students with disabilities (as defined in the NIA);
- (b) Includes assessments for multiple courses that will be implemented in each member State at a scale that will enable significant improvements in student achievement outcomes statewide; and
- (c) Includes a process for certifying the rigor of each assessment in the assessment program and for ensuring that assessments of courses covering similar content have common expectations for rigor.



Category B Competitive Priority 1: Focus on Preparing Students for Study in STEM-Related Fields

12

An eligible applicant addressing this priority must, in addition to addressing the priority throughout the application narrative, provide a separate plan that describes —

- (a) The courses for which assessments will be developed;
- (b) How the courses comprise a rigorous course of study that is designed to prepare high school students for postsecondary study and careers in the STEM fields; and
- (c) How input from one or more four-year degree-granting IHEs will be obtained in developing assessments for the courses.



Category B Competitive Priority 2: Focus on Career Readiness and Placement

13

An eligible applicant addressing this priority must, in addition to addressing the priority throughout the application narrative, provide a separate plan that describes—

- (a) The courses for which assessments will be developed;
- (b) How the courses comprise a rigorous course of study in career and technical education that is designed to prepare high school students for success on technical certification examinations or for postsecondary education or employment; and
- (c) How relevant business community participation and support will be obtained in developing assessments for the courses.



The USED Process

14

- Because of the unusual circumstances and the compressed timeline, USED solicited comments before the release of a draft Notice Inviting Applications (NIA) instead of more standard comment period
- Solicited expert testimony from multiple assessment professionals in Boston, Atlanta, Denver, and D.C.
- Listened and incorporated much of the testimony in the final NIA, but did not include many important suggestions, such as funding many (e.g., 5-7) consortia to allow for more learning opportunities
- NIA released on April 6, 2010, applications due on June 23, 2010—not a fun time!
- Reviews in August, awards announced September 2nd



The State Process

15

- States starting organizing into consortia last fall as soon as the intention for consortium funding was made known by USED
- Many consortia formed, dissolved, and merged before settling on the final two category A consortia
 - At times the process felt more like a junior high dance than assessment professionals doing their work
- I'm still not completely clear on the reasons, but the chiefs and the governors were pushing hard for a single consortium
 - Comparability?
 - Making sure all states get a piece of the pie?
 - Efficiency?
- Unfortunately, almost all of them have never dealt with the challenges of being in a consortium



The Consortia

16

- **Category A:**
- Partnership for Assessment of Readiness of College and Careers (PARCC)
- Smarter-Balanced Assessment Consortium (SBAC)

- **Category B:**
- State Consortium on Board Examination Systems (SCOBES)



Category A Consortium Membership

17

Consortium Name	Lead State	Governing States	Non-Governing Member States
SMARTER Balanced Assessment Consortium (31)	WA	CT, HI, ID, KS, ME, MI, MO, MT, NC, NM, NV, OR, UT, VT, WA, WI, WV	AL, CO, DE, GA, IA, KY, ND, NH, NJ, OH, OK, PA, SC, SD
Partnership for Assessment of Readiness of College and Careers (27)	FL	AZ, DC, FL, IL, IN, LA, MD, MA, NY, RI, TN	AL, AR, CA, CO, DE, GA, KY, MS, NH, NJ, ND, OH, OK, PA, SC

Category B Consortium Membership

18

Consortium Name	Governing States
State Consortium on Board Examination Systems (12)	AZ, CT, KY, MA, ME, MS, NH , NM, NY, PA, RI, VT

Awards

19

- PARCC--\$170 million
- SBAC--\$160 million
 - PARCC asked for \$10 million more
- SCOBES—Not funded (questionable review process!)



The Category A Consortia

20

- **PARCC**
 - Major design goals:
 - ✦ Report achievement based on a clear definition of college and career readiness
 - ✦ Support high stakes accountability decisions
 - Computer-**based** summative assessment
 - Innovative item types
 - “Through-course” performance assessments
- **SBAC**
 - Major design goal:
 - ✦ Inform/improve teaching and learning to help prepare students for CCR
 - Computer-**adaptive** summative assessment
 - Computer-adaptive optional interim assessments
 - Considerable teacher involvement in item development and scoring open-ended items



Category B Consortium

21

- **SCOBES:**
 - Major design goals
 - ✦ Raise teaching and learning in U.S. high schools to internationally competitive levels
 - ✦ Use “best in the world” systems instead of reinventing the wheel
 - Participating schools would select from a set of “qualified” Board Examination Systems (BES)
 - Rely on comprehensive instructional and assessment systems
 - Comprehensive, multi-disciplinary approaches (i.e., a lot more than reading and math)
- Since SCOBES did not get funded, the rest of the presentation focuses on the Category A winners



Both Category A Proposals Promise To....

22

- Measure the full range and depth of the Common Core State Standards (CCSS)
- Ensure that results are comparable across states within each consortium
 - Strong policy interest in trying to ensure comparability across consortia
- Establish achievement standards at the end of high school that are linked to empirically validated definitions of college and career readiness
- Support instruction and learning by providing actionable feedback to educators and others
- Ensure that the assessments are accessible as possible for all learners
- *And to cure all diseases and feed the world* 😊



PARCC Design (3-8)

23

- Designed using “Evidence-centered Design” (ECD)
- “Streamlined” computer-based summative assessments (administered after 90% of instruction) for fast reporting of results
 - Will include, to the extent feasible and supported by research, “innovative item types” and automated scoring of open-ended items
- “Through-course performance” assessments, which are part of the summative score, but administered during several test windows throughout the school year
 - Administered after 25, 50, and 75% of instruction, respectively
 - Math will use three of the through-course assessments
 - ELA will use four, the 4th being a listening/speaking assessment tied to the 3rd ELA performance assessment
 - Will rely on AI scoring, in part, to the extent it is supported by research



PARCC Design—High School

24

- Summative assessments in each grade, 9, 10, and 11
- The high school assessments are structured, i.e., through-course, similarly to grades 3-8 assessments
- Modular-based so that theoretically could be split apart and reassembled differently



Smarter-Balanced Design (3-8)

25

- **Computer-adaptive summative assessments**
 - “adaptive” assessments rely on strong Item Response Theory (IRT) assumptions whereby the computer selects items for each student based on their responses to previous items
- **Intention to develop computer adaptive interim assessments based on well-defined learning progressions—optional for state use**



SBAC-Grades 3-8 adaptive comprehensive summative assessment

26

- ELA and mathematics (separate assessments)
- Adaptive portion includes multiple-choice, technology-enhanced and more traditional constructed-response items*
 - Designed to measure “grade-level” and “instructional level”
- Used for measuring achievement and growth for Title 1 accountability purposes (on track to being college- and career-ready)
- Includes 2 performance tasks** (typically 1-2 class periods) in each content area (to be administered only one time per year contingent on budget and item bank)
- Secure high-stakes item pool

* The basic preliminary item blueprint across grades includes: for Reading 44% selected response, 21% technology enhanced constructed response, 18% traditional constructed response, 18% performance assessments. For Writing, Speaking and Listening, 9% selected response, 32% technology enhanced constructed response, 25% traditional constructed responses and 35% performance assessment. For Math 22% selected response, 41% technology enhanced constructed response, 14% traditional constructed response, 23% performance assessment.

** The proposed number and scope of performance tasks is an estimate. The actual tasks to be included within the system will be dependent on the final version of the Common Core State Standards and the number and scope of tasks necessary to assess the full range of the standards at each grade.



SBAC- Grades 3-8 adaptive interim

27

- Based on learning progressions and/or CC content clusters
- Blueprint will provide for more in-depth assessment of what students know and can do based on smaller clusters of content and learning progressions (on track to college-and career-ready)
- Will include a performance task bank mapped to the content standards for local use
- A non-secure pool of items



SBAC—HS Summative Assessment System

28

- ELA and mathematics (separate assessments)
- Adaptive portion includes multiple-choice, technology enhanced and more traditional constructed-response items
 - Parts will be modular so students can be assessed when ready
- Used for measuring achievement and growth for Title 1 accountability purposes (college- and career-ready)
- Includes up to 6 performance tasks (typically 1-2 class periods) in each content by the end of grade 11--half of the tasks will assess the ELA or math content in the context of science or social studies (per the expectations of the CCSS)
- Secure high-stakes item pool
- States must administer the summative assessment in grade 11 until research shows that scores from assessments administered earlier (grade 9 or 10) are comparable to those from the grade 11 assessment.
- An adaptive grade 9/10 summative assessment will be available for States whose selected high school growth model requires a comparable score between the grade 8 and grade 11 administrations.



SBAC-Grades 9 - 12 adaptive interim assessments

29

- Blueprint will provide for more in-depth assessment of what students know and can do based on smaller clusters of content and learning progressions (on track to college-and career-ready)
- Will include a performance task bank mapped to the standards for local use
- A non-secure pool of items



Key Differences

30

Smarter-Balanced

- Adaptive testing
- Constructed response included in adaptive summative
- Optional interim adaptive assessment modules
- Single HS summative assessment

PARCC

- Fixed form—CBT
- Through-course performance tasks (including extended tasks)
- Significant release of items after each administration
- HS summative in grades 9, 10, & 11

Common Challenges

31

- **Design/Implementation**
 - Designing innovative, computer-administered items
 - Valid AI scoring
 - High school assessment
 - Ensuring comparability
- **Policy/Practical**
 - Testing windows
 - Common accommodations policies
 - Working with vendors
 - Professional development and support
 - Working and playing well together!!



Specific Key Challenges

32

Smarter-Balanced

- Using adaptive assessments for content-based interpretations
- Incorporating open-ended items into CAT (never been done)
- Transparency

PARCC

- Using through course components in summative score
- Determining who takes which HS assessment
- Test security
- Valid equating

Implications for school and districts

33

- Provide opportunities for students to learn to the depth and complexity called for in the CCSS
- Working with the state, design appropriate plans to transition from current standards to CCSS
 - Gap analysis, materials alignment, etc (state will help here)
- Depending on consortium, figure out how best to incorporate and use the through-course assessments
- Depending on consortium, incorporate interim/benchmark assessments
 - Could allow districts to give up currently used commercial products
- Take advantage of PD opportunities provided by consortium/state



Questions and Answers

34

- For any questions we don't get to here or if you think of something later....
- smarion@nciea.org

