



Developing a High Quality Assessment System

North Dakota Assessment Task Force

Ed Roeber
Barry Topol
John Olson

October 6, 2015



Overview of the Presentation

- ASG Overview
- Introduction to Assessment Design
 - Users and Uses of Assessment
 - Types of Assessment
- Assessment Systems
 - Summative, Interim, and Formative Assessment
- Questions about Balance in an Assessment System
- Characteristics of a High Quality State Assessment
- USED *Peer Review of State Assessment Systems*
- Process for Creating ND's Assessment System and Next Steps

www.assessmentgroup.org

2



ASG Overview

- Mission: Assist state departments of education and others to maximize value throughout the assessment design, procurement and implementation process
- ASG Senior Consultants have over 100 years of assessment industry experience at the SEA, vendor, university, and education non-profit levels
- Saved states over ***\$100 million*** since 2009

October 6, 2015 www.assessmentgroup.org 3



ASG Overview (cont.)

- ASG and its Senior Consultants have:
 - Written extensively on assessment design and implementation
 - Evaluated numerous assessment systems
 - Written dozens of Requests for Proposals (RFPs)
 - Serve(d) on state Technical Advisory Committees (TACs)
 - Helped several states conduct successful procurements

October 6, 2015 www.assessmentgroup.org 4



ASG Overview (cont.)

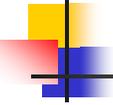
- ASG has developed an industry unique *Assessment Cost Model* that:
 - Determines the appropriate price for any assessment
 - Contains hundreds of variables and over 70 databases
 - Builds up the cost of an assessment from the ground up, function by function, activity by activity
 - Has been proven accurate to within 5%-10% of vendor costs
 - Useful in predicting assessment cost before procurement and in evaluating vendor RFP bids

October 6, 2015
www.assessmentgroup.org
5



ASG Overview - Partial List of Clients/Services

State/ Consortia	Pre-RFP Costing	Assmt. Design	RFP Devel.	RFP Tech Eval.	RFP Cost Eval.	Other
Kentucky (5)	√	√	√	√	√	√
Mississippi (4)			√	√	√	√
PARCC (8)	√	√	√		√	√
SBAC (2)	√	√			√	
South Carolina	√	√				√
WIDA (2) World Class Instructional Design and Assmt	√	√	√		√	√
Pennsylvania	√	√	√		√	
Colorado	√	√	√	√	√	√
Na'tl Center & State Collaborative	√	√				√
Maine			√		√	√
CCSSO/Stanford (2)	√					√
Arizona			√			6



Introduction to Assessment Design

7



Introduction to Assessment Design

- There are different users of assessment information (students/parents; educators; policymakers; the public)
- There are different uses of results at the state, district, school, classroom, and student levels

Question: Have the various audience and purposes for assessment been clearly defined and agreed upon?

October 6, 2015

www.assessmentgroup.org

8



Uses of Assessment

- Guide individual student learning
 - Determine extent of student learning – in the classroom, school, and district levels
 - Improve the instructional program
 - Evaluate the educational program
 - Hold students, educators, or the educational system accountable for performance
 - Certify learning
 - Predict future performance
 - Monitor growth and progress in achievement
 - Determine college and career readiness

October 6, 2015 www.assessmentgroup.org 9



Basic Types of Assessment

- There are two basic types of assessment:
 - Assessments *for* learning – These are assessment processes that occur during daily instruction and help teachers adjust instruction and students adjust their learning (formative in nature)
 - Assessments *of* learning – These are assessments occur at the conclusion of a unit of instruction (summative in nature) or at the end of a school year
- Both assessment types are important but are different

October 6, 2015 www.assessmentgroup.org 10

Assessment Users and Purposes

Assessment User	Assessment <u>for</u> Learning	Assessment <u>of</u> Learning
Students	<ul style="list-style-type: none"> • What am I supposed to learn? • Am I learning enough? • What should I do next? • What help do I need? 	<ul style="list-style-type: none"> • Am I succeeding at the level that I need to be? • Am I capable of success? • Is learning worth the effort?
Teachers	<ul style="list-style-type: none"> • What does each student need? • What strengths can I build on? • How should I group my students? • Am I going too fast? Too slow? 	<ul style="list-style-type: none"> • What grades do I give each student? • What do I tell parents? • Do any students need to be referred for special services? • Will each student be ready for the next grade?

www.assessmentgroup.org

11

Assessment Users and Purposes

Assessment User	Assessment <u>for</u> Learning	Assessment <u>of</u> Learning
Parents	<ul style="list-style-type: none"> • Is my child keeping up? • What can we do at home to support learning? • How can I help my child if he/she is struggling? 	<ul style="list-style-type: none"> • What grades did my child receive? • Is this teacher doing a good job? • Is this a good school and district?
Principal	<ul style="list-style-type: none"> • Are students being taught and learning what they should be? • Are teachers using formative feedback during instruction to guide their teaching? 	<ul style="list-style-type: none"> • How are all students doing on the state tests? • How do I allocate building resources? • Are students ready for the next level or college/workplace?

www.assessmentgroup.org

12

Assessment Users and Purposes

Assessment User	Assessment <u>for</u> Learning	Assessment <u>of</u> Learning
Curriculum Director	<ul style="list-style-type: none"> • Are teachers addressing the learning targets? • How actively are students engaged in learning? • Are the teachers using formative assessment evidence to make instructional decisions? 	<ul style="list-style-type: none"> • How did our district do? • How did each school do? • On skills that students did poorly on, where should they have been taught? • How can we improve horizontal and vertical articulation? • What professional development opportunities do we need to provide to teachers?
Superintendent	<ul style="list-style-type: none"> • Does the district have policies promoting balanced assessment? 	<ul style="list-style-type: none"> • Did our district do well? • How did each school do? • How do we compare? • How can we improve • Where do I need to allocate more funding? • What policy changes do we need to make?

www.assessmentgroup.org

13



Assessment Users and Purposes



Assessment User	Assessment <u>for</u> Learning	Assessment <u>of</u> Learning
Citizens		<ul style="list-style-type: none"> • Are our schools doing better/worse this year? • Are our students competitive with others? • Are taxpayer dollars being used well?
State Department of Education	<ul style="list-style-type: none"> • Will the use of the formative assessment processes help more students succeed? • Can more teachers be taught to use formative assessment? 	<ul style="list-style-type: none"> • Which schools are doing well or not so well? • How can we help them? • What resources are needed?

www.assessmentgroup.org

14



assessment solutions group

Assessment User	Assessment <u>for</u> Learning	Assessment <u>of</u> Learning
Legislature		<ul style="list-style-type: none"> • What policies should be put in place to improve student achievement? • What funding is necessary to improve student achievement? • Are our schools providing a high return on investments

Adapted from *Student Involved Assessment FOR Learning*, 4th edition, R.J. Stiggins, Prentice-Hall, copyright 2005

www.assessmentgroup.org 15



assessment solutions group

- One single assessment cannot meet all of the purposes of all stakeholders equally and effectively
- Some assessment purposes are mutually exclusive with other purposes
 - A school accountability (summative) test may not provide much information to guide individual student learning
- **This means that a *system* of assessment is needed to meet different purposes**

October 6, 2015 www.assessmentgroup.org 16

Balanced Assessment System



- A *balanced* system of assessment is needed to meet all stakeholder assessment purposes equally well
- A balance between summative, interim, and formative approaches to assessment is needed
 - Summative assessment
 - Annual assessment given at the end of the school year
 - Interim or benchmark assessments (mini-summative assessments) used one or more times during the year
 - Formative assessment strategies
 - These are used daily, embedded in classroom instruction, to assure students are learning what teachers are teaching.

www.assessmentgroup.org

17

Summative Assessment



- Annual End-of-Year Tests
 - These are assessments *of* student learning
 - All students tested on the same material using the same assessments
 - Typically given to all students at the same time,
 - Same grades and/or same content areas
 - Multiple-choice and constructed-response items most often used, though many states are now using technology-enhanced items and performance assessments
 - Standardized administration; can be used to compare student, school, district performance
 - Have limited instructional/remedial uses
 - Primary Stakeholders—Superintendents, Principals, State Education Agency, Legislature, and the Public

18




Summative Assessment

- Interim or Benchmark Assessments
 - These are short-cycle summative assessments *of* student learning; they are not “formative” tests
 - These are assessments *of* student learning, used for different summative purposes
 - Given periodically during the school year
 - Measure chunks of the school’s or district’s curricula in important content areas
 - Provide evidence that all students have learned key skills
 - Serve to reinforce a common curriculum and a common set of learning experiences
 - Primary Stakeholders—Principals, Curriculum Directors, Teachers, Parents, and Students

October 6, 2015 www.assessmentgroup.org 19




Types of Interim or Benchmark Assessments

- Quarterly – Have the students learned the targets for this marking period?
- Instructional Unit - Have the students learned the targets for this unit of instruction?
- Standards-Based – Have students learned the instruction on this content standard?
- On-Going – Did students learn what I am teaching them this week? Do any need additional instruction on any content?
- May be school developed, state developed or purchased from various vendors

October 6, 2015 www.assessmentgroup.org 20

The Formative Assessment Process



- “Formative assessment is a planned process in which assessment-elicited evidence of students’ status is used by teachers to adjust their ongoing instructional procedures or by students to adjust their current learning tactics.” (Popham, 2007)
- “Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students’ achievement of intended instructional outcomes. (CCSSO SCASS FAST Project, 2007)



October 6, 2015

www.assessmentgroup.org

21

Key Features of the Definitions



- Formative assessment is a *process*, not a “thing”
- Formative assessment literally takes place *during instruction*
- It is a *planned* process – It is carefully planned by the teacher in advance—when and how evidence of student learning will be collected, as well as how it will be used
- The function of this feedback is to help teachers and students make *immediate adjustments* that will improve students’ achievement of the intended curricular aims
- Teachers’ instructional adjustments - Teachers will *modify* their *current* instructional activities
- Student learning tactics - Students will change how *they* learn

October 6, 2015

www.assessmentgroup.org

22

Questions about Balance in an Assessment System



- Are all elements of the assessment system *coordinated* – do all assessments measure the same skills?
- Are standards, curriculum, instruction, learning, assessment, and professional development all aligned with one another?
- Are both summative and formative approaches to assessment used?
- Is each type of assessment *equally important*?
- Is assessment *continuous* - do the assessments measure student progress over time (daily, weekly, monthly, and each school year)?
 - Are a *variety* of formal and less formal assessment methods used on an on-going basis?

October 6, 2015

www.assessmentgroup.org

23

Questions about Balance in an Assessment System



- Is assessment integrated - does the assessment system fit well into the larger educational system?
- Are a variety of high quality assessments being used?
- Do assessments meet various purposes for assessment for the different audiences?
- How proficient are the educators in each school, each district, and the state with classroom-based formative assessment and summative assessments?

24



Characteristics of High Quality Assessment



- Reliability – The results observed/obtained can be repeated; they are stable (unless new learning has occurred)
- Validity – There is evidence to support that the assessment can be used for its intended purpose(s)
- Fairness – There is no bias in this assessment that advantages or disadvantages any student; appropriate accommodations are provided to students who need them
- Technically Sound – the assessment is psychometrically of high quality and meets all standards for technical adequacy

October 6, 2015 www.assessmentgroup.org 25



Characteristics of High Quality Assessment



- Feasible – Students know how to respond to the test and educators know how to administer it
- Practical – The assessment does not take undue time or effort to administer
- Cost Effective – The assessment provides good value for what it costs
- Useful – The assessment results are informative to a stakeholder group (but not necessarily equally to all stakeholders)

October 6, 2015 www.assessmentgroup.org 26



Select an Existing Assessment or Develop One?

- This is a key issue for states who are examining their statewide summative assessment
- The characteristics of a high quality assessment are important to consider either way
- One way to select an existing assessment for statewide use is to issue a Request for Proposals in which potential vendors need to address these characteristics in their proposal, and the state selects the best assessment at the lowest cost
- Or, the state could ask the vendor to develop a customized assessment for the state

27 www.assessmentgroup.org 27



Select or Develop Criteria

Develop State Owned or Purchase an Existing Assessment - Criteria									
Assessment Option	Cost*	Control			Online Requirement	Comparability	Quality+	Other Tools & Test Support	Internal Requirements
		Structure	Content	Timing					
Alternative 1									
Alternative 2									
Alternative 3									
Alternative 4									
Alternative 5									

* Consider technology upgrade costs; Costs to replicate a CAT for an individual state are prohibitive
 + Quality and depth of alignment to College and Career Ready standards; Depth of Knowledge (DOK) of test items

October 6, 2015 www.assessmentgroup.org 28



Developing High Quality Assessments on Your Own



- Start with clearly-stated student outcomes that will serve as the basis for the assessment
- Develop a clear blueprint for the assessment – what will be assessed, with what types of items, how many of each type will be used, and how results are to be reported
 - Requires input from curriculum specialists and assessment specialists
- Select qualified item writers or have the vendor develop the items
 - If the state elects to write the items, it provides training to item writers in how to write good items that are content-related and free of bias
- Write the needed items
- Conduct reviews of the items
 - Content experts review the items
 - Bias and sensitivity experts review the items

www.assessmentgroup.org

29



Developing High Quality Assessments



- Field testing of newly developed items is conducted to determine:
 - Item worked okay as written
 - Item needs to be revised
 - Item does not work and should be deleted from the item pool
- Final changes are made to the items
- Test forms are constructed to measure what is in the blueprints and content specifications
- Administration procedures developed and test is administered
- Psychometric analyses (equating, scaling, etc.) are conducted to make sure the test and test data are valid and reliable
- Results are reported

www.assessmentgroup.org

30



USED *Peer Review of State Assessment Systems*



- USED released to states the *Peer Review of State Assessment Systems* guidance document on September 25, 2015
- It outlines the required assessments that states use that will need to be peer reviewed
- Assessments that may need review include the grades 3-8 and high school assessments used for Title I accountability (general education and special education assessments)

October 6, 2015
www.assessmentgroup.org
31



USED *Peer Review of State Assessment Systems*



- Assessments that need to be reviewed include those that were new to the state or that were substantially revised for the 2014-15 school year
- ND was previously approved but each change in assessment requires a new peer review
- Thus, North Dakota will need to submit peer review materials for the assessment used in 2014-15, as well as any new one selected or developed for use in the future

October 6, 2015
www.assessmentgroup.org
32



USED Peer Review of State Assessment Systems



- The guidance document from USED outlines a number of Critical Elements and examples of the types of evidence that states must provide
 - Section 1: Statewide System of Standards and Assessments
 - Section 2: Assessment System Operations
 - Section 3: Technical Quality – Validity
 - Section 4: Technical Quality – Other
 - Section 5: Inclusion of All Students
 - Section 6: Academic Achievement Standards and Reporting

October 6, 2015

www.assessmentgroup.org

33



USED State Assessment Systems Peer Review



Map of the Critical Elements for the State Assessment System Peer Review

1. Statewide system of standards & assessments	2. Assessment system operations	3. Technical quality—validity	4. Technical quality—other	5. Inclusion of all students	6. Academic achievement standards & reporting
1.1 State adoption of academic content standards for all students	2.1 Test design & development	3.1 Overall Validity, including validity based on content	4.1 Reliability	5.1 Procedures for including SWDs	6.1 State adoption of academic achievement standards for all students
1.2 Coherent & rigorous academic content standards	2.2 Item development	3.2 Validity based on cognitive processes	4.2 Fairness & accessibility	5.2 Procedures for including ELs	6.2 Achievement standards setting
1.3 Required assessments	2.3 Test administration	3.3 Validity based on internal structure	4.3 Full performance continuum	5.3 Accommodations	6.3 Challenging & aligned academic achievement standards
1.4 Policies for including all students in assessments	2.4 Monitoring test admin.	3.4 Validity based on relations to other variables	4.4 Scoring	5.4 Monitoring test admin. for special populations	6.4 Reporting
1.5 Participation data	2.5 Test security		4.5 Multiple assessment forms		
	2.6 Systems for protecting data integrity & privacy		4.6 Multiple versions of an assessment		
			4.7 Technical analyses & ongoing maintenance		

34



USED Peer Review of State Assessment Systems

- Peers (experts in assessment and other areas) review the state application and provide feedback on it to the USED, which makes the final determination
- Peers will review both the completeness of each description as well as its technical quality
- There are some new areas for peer review – consortia assessments, test security, and online assessments
- States will have the opportunity to correct any deficiencies and to work towards approval
- Failure to be approved could lead to financial sanctions on the state education agency

October 6, 2015
www.assessmentgroup.org
35



USED Peer Review of State Assessment Systems

Assessment Peer Review Guidance
U.S. Department of Education

Exhibit 3: Examples of a Prepared State Index for Selected Critical Elements

Critical Element 4.2 – Fairness and Accessibility (EXAMPLE)

Evidence	Notes
<p>The State has taken reasonable and appropriate steps to ensure that its assessments are accessible to all students and fair across student groups in the design, development and analysis of its assessments.</p>	<p>General assessments in reading/language arts and mathematics:</p> <ul style="list-style-type: none"> • DIF analyses showed differences by gender for several items in reading/language arts assessments for the grades 3 and 4. Examination of the items showed they all involved reading informational text. To address this for the next test administration, a sensitivity review of all grade 3 and 4 reading/language passages involving informational text will undergo an additional bias review. Instructions for item development in future years will be revised to address this as well. <p>Alternate assessments in reading/language arts and mathematics:</p> <p>No notes.</p>

Evidence

General assessments in reading/language arts and mathematics:

Evidence #24: Technical Manual (2015). The technical manual for the State assessments documents steps taken to ensure fairness:

- Pp. 30-37 discuss steps taken during design and development.
- Pp. 86-92 discuss analyses of assessment data.

Evidence #25: Summary of follow-up to differential item functioning (DIF) analysis.

Evidence #26: Amendment to assessment contract requesting additional bias review for items and added instructions for future item development.

Alternate assessments in reading/language arts and mathematics:

The State's alternate assessments were developed by the ABC assessment consortium. Evidence for the assessments was submitted on this State's behalf by State X. (See State Assessment Peer Review Submission Cover Sheet)

Notes

General assessments in reading/language arts and mathematics:

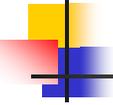
• DIF analyses showed differences by gender for several items in reading/language arts assessments for the grades 3 and 4. Examination of the items showed they all involved reading informational text. To address this for the next test administration, a sensitivity review of all grade 3 and 4 reading/language passages involving informational text will undergo an additional bias review. Instructions for item development in future years will be revised to address this as well.

Alternate assessments in reading/language arts and mathematics:

No notes.

Why this works:

- Concise and clearly written
- Evidence, including page numbers, clearly identified
- Content areas addressed and clearly identified
- Both general and alternate assessments addressed, as appropriate
- Where evidence identified shortcoming, notes discuss how State is addressing
- Cross references submission for assessment consortium



Process for Creating North Dakota's Assessment System



- Create the overall design and framework for the state's high quality assessment program
- Specify the purposes and uses of each testing component
- Consider different options for assessments at the state and local levels – statewide assessments and those to be used at the school and classroom levels – in order to keep overall testing time reasonable
- Conduct build vs. buy analysis
- If build, develop a blueprint for the new assessment program specifying the number of items, types of items, testing times, technical requirements etc.

37



Process for Creating North Dakota's Assessment System



- Decide on the assessment mode to be used (paper based, computer based, or both)
- Estimate the budget for the assessment system
 - Obtain legislative funding, if necessary
- Develop a RFP and use it in a competitive bidding process, based on the design and requirements of a high quality assessment program
- Evaluate the technical and cost proposals submitted in response to the RFP. Conduct an analysis of costs.

38



Next Steps

- North Dakota has much work to do to design and implement its new assessment system
- The state will need to decide whether to select an existing assessment or custom-develop one
- Assessment vendors can offer either type of solution
- A Request for Proposals will help the state describe its proposed solution (i.e., select or develop), provide its specifications, and obtain bids from different vendors
- The state will then be in a position to select the most effective solution to its needs
- Allow enough time (two years is desirable) and funding to develop or select and implement the assessments to be used