



North Dakota Assessment Task Force

Small Group Exercise Observations

Barry Topol
Ed Roeber
John Olson

January 27, 2016



Overview/Content

- Provide thoughts and comments on ND Task Force small-group exercise of December 21, 2015
 - Comparison of Computer-Based vs. Computer Adaptive Tests
 - Pros and cons of purchasing an off-the-shelf assessment
 - Discussion of assessment quality and alignment to state standards
 - Thoughts on Grade 11 assessment/college entrance exam




General Observations

- The Task Force is on the right track in recommending changes in the structure of the North Dakota system of assessments
- Purpose, content, and item types recommended to be used in grades 3-8 make sense
 - A few clarifications and observations will be presented to the Task Force for consideration
- Purpose, content and item types recommended to be used in high school is clear
 - A few observations are noted to make sure the decision on high school assessments is not made too hastily
- Alignment to state standards/quality should be integrated into the thinking for both grades 3-8 and HS

3




Computer-Based vs. Computer-Adaptive Tests

January 27, 2016

www.assessmentgroup.org

4



Computer Based (CBT) vs. Computer Adaptive (CAT) Tests

- Attributes common to both CBT and CAT
 - Delivery method/elimination of paper
 - A valid/accepted method of testing students
 - Ability to place students on a common scale
 - Ability to handle many types of accommodations
 - Students tend to be more engaged with the test
 - Capability of using unique item types
 - Results for multiple choice items can be returned quickly*

* Human scored items will take roughly the same time to score

5



CBT vs. CAT - Differences

Attribute	CBT	CAT
Assessment Type	Fixed Form	Adaptive
Implication	All students take the same exam	Assessment items/exam adapt to the ability of the student
Number of items required	Same as fixed form paper and pencil (PPT) exam	Several multiples of fixed form requirement
Implication	Lower cost than PPT and potentially lower cost than CAT	Custom Development can be cost prohibitive for a single state. Check for available item banks.
Test length (time)	Same as PPT	Same as PPT but measurement can be more precise.
Student Engagement	Better than PPT	Theoretically the highest level of student engagement can be reached
Scoring Scale	Common Scale	Common Scale. Proponents of CAT believe the abilities of students at the higher and lower levels of the scale are identified more accurately.

January 27, 2016
www.assessmentgroup.org
6



CBT vs. CAT

Considerations for ND



- Few off-the-shelf CATs are available
 - SBAC
 - AIR – Using Utah or other shared items
 - Other CATS (i.e., NWEA MAP) are not used for summative assessment
- Custom-developed CAT assessments are expensive due to the number of items required (not the technology) – need item banks in the thousands

January 27, 2016 www.assessmentgroup.org 7



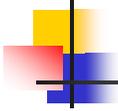
CBT vs. CAT

Considerations for ND



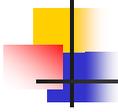
- There will likely be more options in procuring a CBT assessment
 - Build
 - Buy
 - Hybrid
- The CBT vs. CAT determination can be left open and vendors can be asked to respond to one or the other (or both) testing modes in the final RFP

8




Pros and Cons of Different Assessment Options

January 27, 2016
www.assessmentgroup.org
9

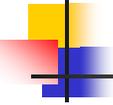



Build vs. Buy vs. Hybrid Assessment Options

- Build – Vendor develops a custom assessment specifically aligned to the state’s content standards and requirements

Pros	Cons
<ul style="list-style-type: none"> • Get what you want • Alignment to standards • High quality • High amount of control <ul style="list-style-type: none"> • Design – structure, content, item types • Timing of administration • Delivery method (CBT/PPT) 	<ul style="list-style-type: none"> • Cost – Likely most expensive option • Time to Market – 2 years to implement due to development and field testing requirements • Can not compare student results to those of other states

January 27, 2016
www.assessmentgroup.org
10



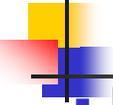
Build vs. Buy vs. Hybrid Assessment Options



- Buy – State purchases an “off the shelf” assessment from a vendor

Pros	Cons
<ul style="list-style-type: none"> • Cost – Generally the low cost option • Time to Market – Fastest (< 1 Yr.) • Alignment to North Dakota State Standards – Possible but state needs to verify • Comparability – Perhaps in the future depending on the product selected. Nationally used tests will have norms that can be used for comparability. 	<ul style="list-style-type: none"> • Control – State has no control over content, design • Quality – Items testing critical thinking skills are an issue • May need to augment test for better alignment • Detailed reporting information and sub-scores may not be available • Tests in areas such as Science or Social Studies may not be available (two different testing programs?)

11



Build vs. Buy vs. Hybrid Assessment Options



- Hybrid – A vendor combines an existing “off-the-shelf” test with custom-developed items. Custom items are field tested in Yr. 1 and available for use in Yr. 2.

Pros	Cons
<ul style="list-style-type: none"> • Time to Market – Can launch quickly in year 1, using OTS test component • Cost – less expensive than a full custom assessment • Better alignment to state standards than OTS product • Generally speaking, there is better quality and control than with an OTS product 	<ul style="list-style-type: none"> • Quality – Varying quality of products by vendor. State needs to ensure the quality and alignment of the OTS component meets its needs. • Comparability – Can only compare OTS items and only if enough other states use that component of the assessment (or national norms are available)

12



Build vs. Buy vs. Hybrid

Build Vs. Buy Vs. Hybrid - Summary

Assessment Option	Cost*	Control			Online Requirement	Comparability W/other states	Quality+	Other Tools and Support	Internal Requirements	Time To Market**
		Structure	Content	Timing						
Custom Developed Assessment	Highest	High	High	High	Up to State	Low	High	Depends	Med- High	1 year to Dev. 1 year to FT
Buy Off the Shelf Assessment	Lowest	Low	Low	High	Most Likely Up to State	TBD	?	High	Lowest	6 months
Develop Hybrid Assessment	Middle	Med	Med	High	Most Likely Up to State	TBD	Med/High	Depends	Med- High	6-9 months

* Consider technology upgrade costs; Costs to replicate a CAT is prohibitive for an individual state unless items are shared
 + Quality and depth of alignment to College and Career Ready standards; Depth of Knowledge (DOK) of items; Reporting capability
 ** From RFP award to test administration. Allow 2 to 5 months for RFP development and approval depending on state requirements/regulations

January 27, 2016
www.assessmentgroup.org
13



Assessment Quality and Alignment to North Dakota State Standards

January 27, 2016
www.assessmentgroup.org
14



Assessment Quality

- An element ASG did not see reflected in the small group exercise was assessment quality
- There is a relationship between assessment cost and quality
- Custom-developed tests can have a higher alignment to state content standards
 - Alignment is two-way – Are all the standards assessed, and are they assessed by items with the same rigor as the standards

15



Assessment Quality

- Quality – To some degree, quality is in the eye of the beholder but in the assessment world, quality is generally measured in several key ways:
 - Alignment – How well does the test align to the state's content standards?
 - Reliability – Is there evidence of the stability and internal consistency of the measures used?
 - Validity – a) Construct Validity - Is there evidence to support that the test measures what it is supposed to measure and that it can be used for its intended purposes?

16



Assessment Quality

- Validity - b) Criterion Validity – Is there evidence that the results be validly reported at various levels (e.g., standards, sub-standards and items) to policymakers, administrators, teachers, parents and students?
- Comparability – Can results of this years assessment be compared to previous and future years? If desired, can they be compared to other states' results that use the same instrument?
- Other – Does test measure higher order (21st Century) thinking skills, problem solving, and other important skills that students should have?
- *Note: USED Peer Review also evaluates the quality of a state's assessment*

17



Grade 11 Assessment Options

January 27, 2016

www.assessmentgroup.org

18

Grade 11 Assessment Options



- There are several options for high school assessments (e.g. general, EOC, CRR or CEE tests like ACT or SAT)
- The task force appears to be leaning to the ACT as the grade 11 exam
- Different assessments should be used for different purposes
- The main use of a college entrance exam (CEE) is to determine which students are college ready
- Different vendors have very different programs, and both have changed their offerings substantially in the past few years. Each has their pros and cons.
- If a CEE is desired, that element of the assessment system should be identified in the RFP

19

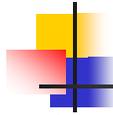
Grade 11 Assessment Options (cont.)



- An RFP will enable the state to assess the strengths and weaknesses of the various offerings AND obtain a more competitive price
- States electing to use additional assessments (beyond CEE) at the high school level do so for several reasons
 - Additional assessments allow for progress monitoring of all students
 - Additional assessments may align more closely with state standards than a CEE
 - Many states find that use of a common vertical scale to track and report student progress more precisely from grades 3 – 11 is valuable

www.assessmentgroup.org

20

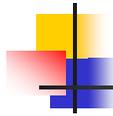



Grade 11 Test

Pros and Cons of Using a College Entrance Exam (CEE) as the Grade 11 Assessment

Pros	Cons
Simplifies the assessment system	Assessment may not be appropriate for non college-bound students
Eliminates an assessment	Different assessments are used for different purposes. A CEE is not appropriate for all purposes
Enables college-bound students to focus on the college entrance exam instead of on two tests	Will not give the state information on how high schools are performing for all students
Approximately 8 states are using a CEE for their grade 11 test in 2016-17 so a precedent has been set	Federal approval is still required- state must show alignment between the assessment and the state content standards
Reduces cost if state is already paying for a CEE	The cost of CEEs is generally more than the cost of regular assessments
Math, English language arts and Science content is included in the assessment	Alignment to state standards may be less than ideal

January 27, 2016
www.assessmentgroup.org
21

Questions?

January 27, 2016
www.assessmentgroup.org
22