

North Dakota Assessment Task Force



Summary of Fordham Report on Content and Quality of Next Generation Assessments

John Olson
Barry Topol
Ed Roeber

April 11, 2016

Overview of Presentation



- ASG presents brief summary of Fordham report and highlights key issues of importance to the ND Assessment Task Force
- Presentation based on report published February 2016 by the Fordham Institute entitled "*Evaluating the Content and Quality of Next Generation Assessments*"
- Our focus is on the findings for Grades 5 and 8. HumRRO also did a similar study but focused only on the high school grades.

April 11, 2016

www.assessmentgroup.org

2



Study Overview

- Fordham experts evaluated content and quality of Grades 5 and 8 (“capstone grades” for elementary and middle school) assessments for ELA/Literacy and Math
- Information from study useful to educators, parents, policymakers and state officials on strengths and weaknesses of next-generation assessments:
 - ACT Aspire
 - PARCC
 - Smarter Balanced
 - MCAS (state test for MA)
- Evaluation criteria based on CCSSO’s report “Criteria for Procuring and Evaluating High Quality Assessments”

3



Study Criteria and Ratings Levels

Test items and forms reviewed using specific criteria from CCSSO report:

- Alignment to standards – ELA and Math
- Cognitive demands of test items -- ELA and Math
- Best practices in testing and technical quality

Ratings given by evaluators (large panels of reviewers):

- Weak match– little to no evidence found in the tests
- Limited match – some evidence for the criteria found
- Good match – many examples of evidence found
- Excellent match – strong evidence

4



Study Overview (cont'd)

- Focus on three key areas/questions:
 - Content – do assessments place strong emphasis on the most important content for CCR standards and/or CCSS?
 - Depth – do tests require all students to demonstrate range of thinking skills, including higher-order, as called for by the standards?
 - Overall Strengths/Weaknesses – what are these for each assessment in ELA and Math?

CCR = college and career readiness
 CCSS = common core state standards



Findings for Evaluation of ELA/Literacy Content

Test assesses the Reading, Writing, and other content most needed for CCR

- ACT Aspire – mostly limited ratings
- MCAS – mostly weak and some limited ratings
- PARCC – mostly excellent ratings
- SBAC – mostly excellent and good ratings

Note – all tests were weak/limited in assessing Speaking/Listening

6



Findings for Evaluation of Math Content



- Test assesses the Math content and skills most needed for CCR
- ACT Aspire – limited and weak ratings
 - MCAS – all limited ratings
 - PARCC – all good ratings
 - SBAC – all good ratings

7

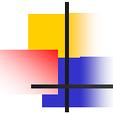


Findings for Evaluation of ELA/Literacy Depth



- Test assesses the depth that reflects the demands of CCR in ELA/Literacy
- ACT Aspire – mostly good ratings but weak on measuring cognitive demand
 - MCAS – mostly good ratings but limited on measuring cognitive demand
 - PARCC – mostly excellent and some good ratings
 - SBAC – mostly good and one excellent rating

8



Findings for Evaluation of Math Depth



Test assesses the depth that reflects the demands of CCR in mathematics

- ACT Aspire – excellent ratings but limited on cognitive demand
- MCAS – all excellent ratings
- PARCC – mostly good or excellent ratings
- SBAC – mix of excellent and good ratings but limited on use of high quality items

9



Summary Ratings



Overall Content and Depth Ratings for ELA/Literacy and Mathematics

	ACT Aspire	MCAS	PARCC	Smarter Balanced
ELA/Literacy CONTENT	L	L	E	E
ELA/Literacy DEPTH	G	G	E	G
Mathematics CONTENT	L	L	G	G
Mathematics DEPTH	G	E	G	G

LEGEND
 Excellent Match
 Good Match
 Limited/Uneven Match
 Weak Match

April 11, 2016

www.assessmentgroup.org

10



Overall Strengths and Weaknesses of Program



- ACT Aspire – items are generally high quality but emphasis is mostly on DOK-3 level items and not enough at DOK levels 1 and 2, especially for grade 8
- MCAS – items are high in technical and editorial quality but need more higher-order thinking skills to be assessed, especially for ELA/Literacy

DOK = Depth of Knowledge

11



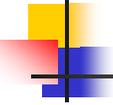
Overall Strengths and Weaknesses of Program



- PARCC – good measures of CCR and cognitive demands; could use some more attention to the accuracy of items, both editorial and mathematical
- SBAC – good measures of CCR; could use more items at DOK level 1, especially at grade 8

DOK 1 – rote or basic skills like identifying obvious detail in text
 DOK 2 – multi-step operations or comprehension across one or more sentences
 DOK 3 – strategic thinking, problem solving, identifying complex themes
 DOK 4 -- extended thinking/problem solving, synthesis and analysis across texts

12



Summary of Findings

- The new next generation assessments (PARCC and SBAC) are superior in
 - Measuring content per CCR standards
 - Assessing depth of knowledge and thinking skills
 - Use of a variety of innovative item types, including technology enhanced items
 - Technical quality and validity

April 11, 2016 www.assessmentgroup.org 13



Issues for ND Assessment Task Force

April 11, 2016 www.assessmentgroup.org 14



Issues for ND Assessment Task Force



- Main issues to consider
 - Quality
 - Time
 - Costs
 - Selecting the best assessment for ND
- Assessment quality
 - Alignment to state's content standards and depth of knowledge assessed by test
 - Variety of item types used
 - Overall technical quality and rigor of the assessment
 - Do not sacrifice quality – it is paramount!

www.assessmentgroup.org

15



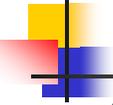
Issues for ND Assessment Task Force



- Total time required to complete the assessment
 - Higher quality assessments with variety of open ended item types usually take more time
 - May want to consider minimizing use of performance tasks
 - If possible, get rid of any redundant or non-aligned tests in the state and districts
- Costs
 - Consortia assessments are likely less expensive to implement than a custom developed, high quality state assessment
 - Use of existing high-quality item pools is an option
 - Use of AI/machine scoring in the future may also save on costs

www.assessmentgroup.org

16



Issues for ND Assessment Task Force



- Selecting the best high-quality assessment
 - Older state approaches like MCAS are not really good enough anymore; most don't fully assess the CCR standards adequately or use innovative item types
 - ACT Aspire needs some improvements made, especially in alignment to content standards and measuring depth
 - New consortium assessments meet many of the important criteria, but may also need some improvements to be made
 - State should choose what is best to meet its needs and demand from the vendor that improvements are made to increase quality of the assessment
 - A clearly written RFP that spells out the state's requirements is essential to getting the right assessment for the state

www.assessmentgroup.org 17



Questions?

January 27, 2016 www.assessmentgroup.org 18