Design Study Report and Recommendations

Developing a System for Tracking State Assessment Policies in Science and Mathematics Education

MARCH 2016

PRESENTED TO:
National Science Foundation
4201 Wilson Boulevard
Arlington, VA 22230

GRANT #1445610

PRESENTED BY:
Rolf K. Blank
Principal Investigator
NORC at University of Chicago
4350 East-West Highway
Bethesda, MD 20814-4499

NORC at the UNIVERSITY of CHICAGO

Developing a System for Tracking State Assessment Policies in Science and Mathematics Education

Design Study Report and Recommendations

An NORC Project under a Grant from the National Science Foundation
(#1445610)

March 2016

Rolf K. Blank, Principal Investigator
Vi-Nhuan Le, Senior Research Scientist
NORC at University of Chicago
Education and Child Development Department
Contact: Blank-Rolf@norc.org
301 634 9325

Table of Contents

List of Tables

Introduction

The NORC design study was initiated in 2014 under a grant from the National Science Foundation (NSF) as one of the STEM education indicators development projects. The primary goal of the project is to develop and test a model for collecting and reporting comparable information on state policies regarding student assessments in mathematics and science education. The outcome of the project will be recommendations regarding an ongoing state assessment tracking system. This report is intended to inform the expert project advisers, state participants in the pilot study, other state education leaders, as well as the National Science Foundation, about the viability of the initial model for collecting and reporting policy information across the states.

The longer-term goal is to develop an online system for reporting state assessment policies across states, and a new grant has been funded to extend the pilot work into a three-year implementation project. The design study is intended to inform mathematics and science educators and leaders at state, local, and national levels about which state assessment policies can be effectively collected and reported and how the information can provide comparable useful information.

The proposal to NSF for developing a tracking system on states math and science assessments is based on the recommendations of a National Research Council (NRC) committee: *Monitoring Progress toward Successful K–12 STEM Education: A Nation Advancing?* (2013). The NRC committee of leading science and mathematics educators and education researchers recommended the development and implementation of a set of indicators to improve tracking and evaluation of the quality of STEM education in our nation's schools. Recommended indicator #12 focuses on reporting and analysis of state policies on student assessment in science and mathematics for all states. The NRC committee reported on the need to initiate collection and reporting of state assessments in math and science, and the relationship of assessments to state content standards.

Rationale for the System of Tracking State Assessment Policies

Research and analysis of the content of state achievement tests has yielded concerns about their academic rigor, as the items tend to assess lower-level cognitive skills (such as recall, recognition, and application of procedures) as opposed to higher-level cognitive skills (such as analysis, evaluation, and synthesis of ideas) (Darling-Hammond, et al., 2013; Resnick, et al., 2003; Webb, 2002). Even in the NCLB era, with its emphasis on accountability, researchers have found state tests generally assess lower-level cognitive skills. In a recent study of tests in 17 states that were selected because the tests were reputed to be more cognitively demanding, Yuan and Le (2012) found that fewer than 2 percent of mathematics items assessed a higher level of cognitive demand. Using a different methodology on 19 state tests, Polikoff, Porter, and Smithson (2011) found that only 7 percent of mathematics items required students to use higher-order cognitive skills, and fully 80 percent of mathematics items assessed memorization, recall, and use of routine procedures.

The lack of test items assessing more advanced analysis and knowledge has resulted in reforms to state content standards and assessments, with the most prominent reform being the development of the Common Core State Standards (CCSS-M) for mathematics and the Next Generation Science Standards (NGSS) for science (NGA/CCSSO, 2010; Achieve & Lead States, 2013). The CCSS-M and NGSS have the potential to improve the degree to which deeper learning is assessed through state achievement tests. The CCSS-M establishes a single set of educational standards for kindergarten through 12th grade that identifies the concepts and knowledge that students should acquire to show that they have attained the skills necessary for college and career success (Yuan & Le, 2014). Similarly, the NGSS initiative focuses on coherency and progression of core conceptual science skills from early grades through high school, with an emphasis on preparing students for college, career, and citizenship. A recent study report and recommendations from the National Research Council is informing the development of new designs and systems for science education in the states to support the goals of the NGSS (NRC, 2013).

The NORC plan for a design and pilot study was intended to test a model for collecting information on states' policies for state-administered student assessments to meet several needs.

The pilot study (2014–15 school year) tested a design for collection of state policies information from three sources—state websites, survey of state assessment staff (focusing on design and methods of reporting), and state content specialists in math and science (focusing on content and alignment of assessments and standards). The design options and issues of focus in the pilot study are summarized in a paper, developed from the recommendations of a project expert panel (NORC, 2015).

A key issue for the design study was how to track states use of innovative features of the mathematics assessments being provided through the consortia--Smarter Balanced (SB) and PARCC, including performance tasks (with multiple steps and explanation of work), balance of items across the different levels of depth of knowledge, use of computer-based testing to improve turnaround time for scoring and reporting, measures of learning content and skills, use of adaptive testing to measure a full range of student knowledge, benchmark testing to track progress through the year, and a digital assessment library for formative classroom assessment. To work toward an indicator of state assessments alignment to standards especially relative to content coverage and levels of cognitive demand, NORC and the advisers considered the Criteria for High-Quality Assessments defined by the chief state school officers (CCSSO, 2014).

Reporting of state policies and trends over time is equally important for science education as part of the proposed state policies tracking system because of the recent shifts in science education standards and the strong interest in new forms of student assessment to match the direction of the standards. Some states are following the recommendations of the NRC committee on NGSS science assessments (NRC, 2014), including moving toward a system of science assessment with varied methods, assessing the three dimensions of science instruction, and maintaining annual reporting of student progress.

Objectives of Design Study

The design study and pilot project addressed two questions about state assessment policies for mathematics and science and development of a system for tracking and reporting across the states:

1) What core information on state assessment programs and policies should be collected and reported across states, i.e., what information on types of assessments and characteristics of assessment programs are important to have available in a 50-state report or online resource?
2) What is an effective method for reporting comparable information on the extent to which student assessments in math and science are aligned to state-adopted content standards, and what is the relationship to the Common Core State Standards-Mathematics and Next Generation Science Standards?

The design pilot study was carried out with support and participation by staff of ten state departments of education (ID, KY, KS, MA, MI, MN, NC, RI, UT, WA). Each of the states voluntarily participated in the study after a request was sent by NORC to state assessment directors. Information was collected during May and June 2015, with reporting covering assessments used with schools and students during the 2014–15 school year (NORC, 2015).

Organization of Report

A) Selection of State Assessment Policies for Reporting System

The project expert panel (see Exhibit 1) reviewed the existing sources of information on state assessment policies and characteristics of the assessments as well as the uses and potential uses of assessment results by decision-makers, educators, researchers, and parents. NORC conducted interviews with state specialists in assessment and science and math education, and presented the initial designs for the reporting system at meetings to gain input from state specialists on their needs and interests. The panel members with support of NORC staff also considered the recent research and recommendations on needs for improved indicators of K–12 science and mathematics education.

A key decision in planning for a state assessment policies indicator is the degree to which reporting and analysis across states should be comparable and quantifiable. For purposes of cross-state reporting, a common metric and organization of information in tabular format facilitates comparisons; however, with greater depth of information, key differences in state policies and design of state assessments could be highlighted and shared among users in different states. Although school accountability has been the primary use of state assessments during the

past 15 years, the panel recommended that the reporting system include identifying the range of intended uses and applications of state assessments in mathematics and science as well as the state role in supporting different types of assessment instruments and designs.

The panel recommended collecting and reporting information for the following categories of state assessment policies:

- Types of state assessments in mathematics and science
- Intended uses of assessment data
- Item or task design
- Timing and methods of administration
- Cost of assessments
- Methods of assessment reporting and dissemination.

The major findings from the pilot study that tested the collection and reporting for these categories are outlined in a following section.

B) Relationship of State Assessments to Standards

The expert panel discussed and reviewed several options with NORC staff for reporting on degree of alignment of math and science assessments in relation to content standards. The central question in studying alignment is the degree to which assessments are consistent with the content and depth of teaching and learning called for in the standards (Webb, 2002; Polikoff, Porter, and Smithson, 2011). Many states have conducted alignment studies for the assessment instruments used for accountability reporting as part of the requirements under federal No Child Left Behind funding. Very recently, a new study has reported on the quality of the state consortia assessments based on the Common Core State Standards and two other assessments used by states, based on the criteria defined by CCSSO (Doorey & Polikoff, 2016).

Under the scope of the NSF grant, it would not be possible to conduct a new study of the alignment of each state's assessments in math and science. However, the panel recommended that NORC report indicators of the degree of alignment by reporting on assessments *content* and *depth* relative to standards – that is, reporting on grade-specific content topics, cognitive levels of

assessment items, and the range of methods used in assessing content knowledge and practices. The indicators would be selected to provide analysis and comparison of state assessments' content rigor and breadth specific to a grade, and the degree to which the items/tasks address the depth of knowledge and cognitive demand of the standards. Thus, regardless of the standards defined by state policy, and the assessments that are selected by states, the alignment indicators should provide a common basis for reporting across states.

The following indicators of content alignment were identified for testing in the pilot study:

- State policy on content standards, and relationship to CCSS and NGSS
- When standards were developed and approved
- Source of state assessment framework or blueprint (consortia, state, other)
- Percentage of grade 5 math assessments on selected content topics
- Cognitive complexity/demand categories represented in assessment items/tasks
- Elementary grades at which selected topics were assessed
- Percentage of grade 11 math assessments on selected content topics
- Timeline and methods of development of new science assessments
- Major content topics for new science assessments.

Pilot Study Results: Findings by State Policy Reporting Category

The results of the pilot study for the key reporting categories are summarized across the ten states that participated. The sources of information for the planned categories were state websites and survey questions provided to state education staff. After the information by category was collected, reviewed and clarified, NORC conducted follow-up interviews with pilot study participants to assess the process and these findings are reported.

Types of Assessments

Results from the pilot study data reporting and analysis are summarized across the states. Following are selected summary findings on types of assessments:

- All pilot states administered required mathematics assessments at grades 3–8, states varied in the elementary grades tested in science. Pilot states required one high school comprehensive assessment in math and science or end of course assessment for high school courses.
- The pilot states reported on math and science assessments that are supported by the state in addition to the assessments designed for accountability reporting. The types of assessments include: End of Course (EOC) assessments, ACT, SAT, EXPLORE, Compass, NAEP, and TIMSS.
- Four of the ten pilot states used math assessments in 2014–15 from the assessment consortia, Smarter Balanced or PARCC, and one state used the NECAP science assessments. Several states combined consortia assessments with state-developed assessments.
- Seven states administered mathematics and science assessments at a single high school grade—four states assessed at the tenth grade, while three states assessed students at the eleventh grade. Three states administered math and science tests at multiple grades, including grades 9 and 11, and two states at all grades. Recommended improvements based on the pilot results are to report the state policy on retaking high school assessments, and to identify which assessments are required to be passed for graduation.
- Two of the pilot states support K–2 assessments in mathematics and science, which are optional for use by districts and schools.

Purposes/Uses of Assessments

The pilot design study asked for reporting on four categories of uses of the assessment results in mathematics and science. In addition, within the accountability use, states were asked to report on the level of accountability reporting – district, school, teacher, and student levels. All of the pilot states reported that mathematics assessments were used for state and federal accountability reporting at the school level.

The table below provides a summary of the reported uses of assessments in science, at grades 3–8 and high school, across the four categories. Eight states used the science assessment results for school accountability at the elementary, middle, and high school levels. One state did not use the assessment results in 2014–15 for school accountability because the assessment is in a transition year. All of the states reported the science assessments were used to inform curriculum and

instruction. Only one state reported the assessment results in science were used to determine college readiness and one state used the results for student placement. Similar results were reported for uses of mathematics assessments.

Table 1. Uses of Assessment Results—Science, (Grades 3–8, 9–12)

State	Accountability	Inform Instruction	College Ready	Student Placement
1	Yes	Yes		Yes
2	Yes	Yes		
3	Yes	Yes		
4	Yes	Yes		
5	Transition yr.	Yes		
6	No	Yes		
7	Yes	Yes	Yes	
8	Yes	Yes		
9	Yes	Yes		
10	Yes	Yes		

Within the accountability reporting use, seven states used the assessment results in mathematics and science for district-level accountability, and five states used the assessment scores for teacher accountability. Only one of the states reported using the elementary/middle grade results for student-level accountability, whereas five states reported using the high school level test level results for student-level accountability. For these states, the math assessments were part of an exit exam, where students needed to pass the tests as part of their graduation requirements.

Item/Task Design

In the pilot study, states reported on the percentage of the state assessment by type of item or task. Seven states were able to report the percentages for mathematics, grades 3–8. Three of the states used Smarter Balanced consortia items in grades 3–8 (states 1, 4, 7), and, due to the adaptive assessment design, the types of items students take vary within a range. Two states did not report item types because the assessment framework does not specify item types, focusing instead on content topics and depth of knowledge; and, one state chose not to report on math items.

Table 2. Mathematics: Types of Assessment Items Used in State Assessments, 2014–15, Grades 3–8

	Percentage of Items					
State	**Perform. Tasks**	**Extend Resp.**	**Brief Resp.**	**Tech. Enhance**	**Grid Response**	**Select Response**
1	5–10	0–5	0–10	10–20	0	50–75
2	0	4	8	0	0	88
3	0	30	10	0	0	60
4	5–10	0–5	0–10	10–20	0	50–75
5	0	0	7	9.5	6	78
6	0	0	0	0	20	80
7	5–10	0–5	0–10	10–20	0	50 to 75

A high proportion of assessment items in elementary/middle grades math assessments are select response (multiple choice) items, from 50 to 88 percent of the items. Four states have extended responses items (with one state at 30 percent of items), and six states have brief written response items in their assessments. Performance tasks are included in three states math assessments (from Smarter Balanced).

A key issue revealed by the pilot results is improving the definitions for reporting across states so that percentages are comparable between states. First, the definition of technology enhanced items could be further delineated. Second, the types of items used on an assessment vary by grade level, and in a revised version, it was suggested that a specific grade (e.g., 5th) should be examined.

Table 3. Science: Types of Assessment Items Used in State Assessments, 2014–15, Grades 3–8

	Percentage of Items					
State	Perform. Tasks	Extend Constr.	Brief Resp.	Tech. Enhance	Grid Response	Select Response
1	0	0	0	0	0	100
2	0	0	0	0	0	100
3	0	30	0	0	0	70
4	0	0	0	0	0	100
5	0	0	2	48	0	50
6	0	0	0	5	0	95
7	33	10	57	0	0	0
8	0	13–14	3–6	0	0	80–84

In three states, all of the science items in elementary and middle grades are select response items, and three more states have over 70 percent of items being select response. Three states include some extended response items in science and one state has over half brief response items. One state has 48 percent of items in a technology enhanced format.

Assessment Administration Policies

Assessment Platform, Calculators, Timing

Increasingly, states are moving towards computer technology to administer the assessments. In mathematics, five states use a computer-adaptive format. In science, only one state uses computer-adaptive format. Only two of the pilot states administer the math assessments solely in a paper-and-pencil format, and one state administers science in paper and pencil format. Three states administer assessments with computer technology and two of these states also offer the paper-and-pencil format.

One question not resolved by the pilot study was the extent to which methods of administration are actually used, since in some states both paper-pencil and computer-based platforms are in place. It would be useful to track the percentage of students assessed with each method.

All of the states in the pilot study reported on the month in which assessments are administered through the state. Currently most states are administering required assessments close to the end of the school year (e.g., late May), with one state in the pilot study conducting their assessments in March. The information should be reported and tracked over time by state. States all reported on the state policies regarding testing time for the assessments, and while most states provided guidance about expected time for completion, only a few states had state requirements limiting time for assessments. A revised version should address the question of what state guidance is provided to districts and schools with respect to the expected time needed for most students to complete the assessment.

Table 4. Percentage of Items with Calculators Allowed on Mathematics Assessments

State	Gr 3–5	Gr 6–8	Gr 9–12
1	None	45–60	45–60
2	82	82	100
3	None	50	50
4	None	50	50
5	90	90	100
6	50	70	70
7	None	(7–8: 100)	100
8	None	45–60	45–60

Across the eight pilot study states there is wide variation in the extent to which calculators were allowed for students completing math assessment items. Three states allow calculator use on elementary assessments, and while all eight states allow use of calculators at grades 6–8 and 9–12, they differ widely on the percentage of items allowing use of calculators.

Cost of Assessments

Only three of the states in the pilot study reported any data on the average per student cost for assessments in math and science, with the average cost varying from $7 to $26 per student. Several factors impeded state assessment directors from reporting on assessment costs, including the problem of disaggregating costs by subject, problem of state-level comparability given differences in how assessments are developed and administered, and the problem of

interpretation of assessment costs in a political environment. The recommendation of the advisory panel is not to pursue reporting of assessment costs in an online reporting system.

Methods of Reporting and Dissemination

C) Public Release of Assessment Items

States were also asked about the issue of releasing assessment items to educators and the public. Half of the states in the pilot study reported that the state reported that they did release items, usually from the prior year. For example, the list below indicates that from five to 50 percent of science assessment items were released for broader review and use.

Table 5. Science Assessment Items Release

State	Release	Percent Items	Recent year
1	Yes	5	2013
2	No		
3	No		
4	Yes	50	2014
5	No		
6	Yes	8	2013–14
7	Yes	18	2014–15
8	No		
9	No		
10	Yes	33	2013

The state that is releasing 50 percent of assessment items in science and math reported that districts and schools use the items and results to guide instructional improvement, and the state views this step as important for improving student results. One recommended improvement in this item in new version of the survey is to ask about other ways that assessment examples are provided, such as through a web site that is Accessible for educators in the state.

D) Timing of Reporting on Assessment Results

All of the pilot study states reported on time period for reporting assessment results for three different purposes: 1) use of data by teachers, 2) review by students and parents, and 3) school

accountability. The month when the assessment results are reported for these purposes are shown below across the pilot study states.

Table 6. Month of Reporting Annual Results

State	Teachers	Students/ Parents	School Accountability
1	May – June	June	October 2015
2	August	August	September
3	June/August	October	September
4	[new assessment]		
5	Immediately	August	Early August
6	June	June	September
7	October/Nov.	November	December
8	June	September	August

The advisory panel found these results to be informative and the format for collection and reporting should be continued.

Intended Uses of Assessment Data –Recommendations for Revision

The pilot study instrument asked that states report on the purposes or uses of specific assessments. All of the states reported the use of requirement mathematics for school accountability under federal and state policies, and eight states reported science assessments were used for school accountability. Four of the pilot states indicated that data were used for teacher accountability and five states reported data were used for student accountability, primarily for graduation requirements. All the states reported data were used to inform instruction and curriculum. In future revisions the advisers recommended asking for an example of local use of assessment data with instruction and curriculum.

In future survey development, information should also be sought about intended uses of state-supported assessments, and the questions about college readiness, placement, and graduation requirements should be asked of other assessments supported by the state, such as end of course assessments, college entrance exams, benchmark assessments, and formative assessments. The reviewers identified a need to simplify the reporting on accountability uses of assessment data,

but increase the information reported on uses in curriculum and instruction improvement, student college and career readiness, and uses for high school graduation.

The pilot study asked states about use of the assessment data in growth models for the purpose of accountability, and whether the state growth model for analyzing student achievement change is reported at the student, teacher, school or district levels. In mathematics grades 3–8, nine of 10 pilot states did use assessment data to track student-level growth and reporting growth at the school-level. Four states used scores to compute district-level growth and three states were using math scores for 2014–15 as part of teacher evaluation reporting. In reviewing the data from the pilot, the study advisers viewed the questions on student growth models as lower priority for an online reporting system on math and science assessment policies, and indicated that additional information that would make the item useful would be too detailed for this reporting system.

Content of State Mathematics Assessments & Relation to Standards

One important objective of this design study is to test a method of collecting, reporting, and analyzing information on the relationship of state student assessments in mathematics to the state-adopted K–12 math content standards. The NSF-supported EAGER grant for the state assessment policy indicator was awarded at a time of strong interest in how states are responding to the new Common Core State Standards in mathematics and how student achievement will be assessed in relation to the new content standards. Thus, an important objective for the instrument design was to develop a method of collecting information by state on the content of state mathematics assessments, and then to analyze how the assessments are aligned to the Common Core Standards for Mathematics.

Two strategic decisions were made in the instrument design through work with the expert advisory panel. First, we decided it is important to identify the official policy position of each state concerning the CCSS-M and the relationship of the state standards to CCSS-M. Then, further questions would provide reporting on the degree to which the content of the state assessment reflects the state standards and the CCSS-M.

Table 7. State Policy Position on CCSS Math

State	Same as State	Minor Revisions	Major Revisions	Unrelated
1	Yes			
2	Yes			
3				Yes
4	Yes			
5	Yes			
6		Yes		
7		Yes		
8	Yes			

Five of the eight states reporting on this question are using the same standards as the CCSS-M, while two states made minor revisions, and one state developed their own math standards. Two states did not report on the mathematics standards. This information is useful for analyzing the data on relationship between state assessment blueprints and the content of assessment items.

Our hypothesis is that many state assessments are likely to have some consistency with the content of the CCSS-M. A second strategic decision was that the policies reporting instrument should be designed to be completed in a reasonable amount of time through cooperation with the staff of a state education agency. With this assumption, it is likely that information can be collected on indicators of assessment content, but a complete content alignment study covering all standards and grade levels would be prohibitive under the scope of work for this grant.

Mathematics Assessment Framework/Blueprint and Content Topics

The pilot study identified the source of the math assessment frameworks or blueprints. The results show that three of the pilot study states were using the Smarter Balanced blueprint for mathematics assessment, two of the states were using the PARCC blueprint, and five states were using a state-specific blueprint or framework. Two states have a state framework used in conjunction with the consortium blueprint for specific grades.

To analyze assessments across states, the pilot study asked for states to report on the percentage of the grade 5 assessment that focuses on each of six topics. The topics are consistent with

content for grade 5 recommended in the Common Core Standards. These topics provide a way to provide comparable data across states.

Table 8. Percentage of Grade 5 Math Assessment Items by Selected Content Topics

State	Write/ Interpret	Add/ Subtract Fractions	Multiply/ Divide	Multi-digit Numbers	Volume/ Multiply	Graph in 1st Quad.
1	5	8	13	10	8	5
2	8	8	15	15	15	8
3* (est.)	n/a	25	20	25	20	10
4	0–5	5–15	10–20	5–15	5–15	5–15
5	4	7	0	12	9	3
6	5–10	20–25	20–25	20	10	5
7	15	25	10	30	5	5
8	0–10	5–15	10–20	5–15	5–15	5–15

As a method of analyzing assessment content by topic, Grade 5 was recommended by the expert panel because it is a key transition grade between math instruction in the elementary grades K–4 and middle grades mathematics instruction. The instrument was designed to target two grades (5, 11) to obtain information on the content of assessments. The plan was to request and report information on content of mathematics assessments for only two grades to provide indicators of how mathematics content is presented in the state assessments in relation to a standard set of reporting categories. The categories are based on the mathematics content expected in the CCSS-M. The panel advised that requesting SEAs to report on all grades of state mathematics assessments would place too large a response burden. SEA contacts were asked to provide the source of the information reported.

Cognitive Expectations Dimension for Mathematics Assessment Items

Analysis of content alignment of standards and state assessments has included content topics and cognitive demand expectations or complexity (sometimes called “depth of knowledge”). Under NCLB, states have conducted alignment studies and submitted content alignment analysis reports to the US ED. Several different models for alignment analysis have been used with state assessments. More recently the Smarter Balanced and PARCC state consortia have incorporated this dimension in their development of assessment blueprints. In planning for the State

Assessment reporting instrument a challenge is how to provide a method of consistent, comparable reporting on the level of cognitive demand or complexity of assessment items across states. That is, what is a common metric for reporting when states may have different methods of aligning assessments to standards? The expert panel recommended reporting the percentage of items in an assessment on three general categories: procedural fluency, reasoning or conceptual understanding, and real world modeling.

The results of the pilot study on methods of reporting on cognitive demand/complexity for grade 5 mathematics assessments indicated that four of eight states did not have an alignment analysis that was used to prepare a report on cognitive demand/complexity of the math assessment items—either before the assessment or with the results. The comments from respondents on the pilot study indicate that this cognitive expectations dimension is typically included in the development of items but is not reported in the blueprint or in reporting of assessment results. Several states had their staff review the assessment items and place them in categories, and some found the grade 5 assessment items not be readily placed in mutually exclusive categories. Even if states do have reporting and categorization in their assessment blueprint or framework by cognitive demand or complexity, they may not have found a close match to the pilot study common categories. Based on the analysis of pilot study results, the expert advisers recommended that information be collected from states on the cognitive dimension categories that each state uses and how the categories are used by the state, e.g., assessment development and reporting results.

Elementary Mathematics Progression of Content Topics

The intent of this question regarding mathematics assessment content is to identify when specific topics (drawn from the CCSS-M) are first assessed and by viewing several key topics to analyze topic progression through the elementary grades, and to compare the progressions across the states.

Table 9. First Elementary Grade when Math Topic Assessed

	Topic					
State	**a**	**b**	**c**	**d**	**e**	**f**
1	3	3	5	5	5	6
2	3	3	4	5	6	7
3	3	3	4	4	6	7
4	3	3	5	5	6	6
5	3	3	6	5	6	7
6	3	3	4	5	6	6
7	3	3	4	5	6	7
8	3	3	5	5	6	7

Topic definitions:
a. Understand a fraction as a number on a number line
b. Multiply and divide to solve word problems
c. Multiply a fraction or a whole number by a fraction
d. Relate volume to the operation of multiplication
e. Divide fractions by fractions
f. Represent proportional relationships between quantities

The responses show high consistency across the states in three of the topics, including a, b, e (divide fractions), and variation by one or two grades in assessing topic c (multiply fractions) and variation by one grade in assessing topic d (relate volume to multiplication) and topic f (represent proportional relationships). Overall, the pilot study states are assessing each of these selected topics between grades 3 and 7, and there is a high degree of consistency across the states.

Content of High School Mathematics Assessments

The expert panel recommended that NORC request assessment content information for mathematics assessments that most students take at grade 11, and they recommended reporting content topics for two types of state high school assessments--End of Course Assessments (EOC) and Comprehensive Assessments (CA). The topics for reporting were selected from the recommended mathematics content in the high school mathematics Common Core Standards.

Table 10. Percentage of HS Math Assessment Items by Selected Content Topics

	Topic							
State	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**
1 (CA)	5	26	5	12	14	17	5	29
2 (CA)	5 –15	5- 15	0–10	0–10	15–25	5–15	0–5	15–25
3 (CA)	5	7	0	2	27	11	3	8
4 (EOC)	9	6	3	6	16	6	3	19
5 (EOC)	13	8–13	13	10	15	10	10	15
6 (EOC)	5–15	10–15	15–25	10–20	20–30	5–10	0	5–10
7 (CA)	5–15	5–15	0–10	0–10	15–25	5–15	0–5	15–25

Topics definitions:

a) Applying trigonometry to solve problems
b) Summarizing & interpreting data and making inferences
c) Transformational geometry
d) Setting up and solving systems of equations and inequalities
e) Building, interpreting, and analyzing functions in different representations
f) Interpreting and working with structure in equations and expressions
g) Working with complex numbers
h) Applying mathematics to solve real world problems

The reported percentages on the HS grade 11 selected mathematics topics, and our follow-up interview results, indicate that the SEAs used different approaches to reporting the information. Seven states were able to compute a percentage or percentage range for the portion of the grade 11 assessment for each of the eight topics selected for this study. In interviews, state respondents indicated that the reported percentages may include some assessment of related topics. (The percentages are typically based on either an existing state analysis or an analysis completed for this request). State respondents also noted that the EOC assessments are administered to students completing the course, but the reported course assessment may not have been at grade 11. One state reported the content categories for their state comprehensive assessment was for administration to students at grade 10.

Summarizing across the eight topics, several states reported close to 100 percent of the assessment items, while others (states 3, 4 in table) reported just over half of the assessment items. The HS assessment topic with the highest average portion of the math assessment across states is *Building, interpreting, and analyzing functions in different representations*. The average percentage of the assessments on this topic is 18 percent. The topic with the greatest variation

across states is *Applying mathematics to solve real world problems,* with percentages varying from 29 percent of the assessment to eight percent of the assessment. The two grade 11 assessment topics with the smallest portions of the assessment are*: Transformational geometry*, and *Working with complex numbers*, with each topic covering an average of eight percent of the assessment.

On reviewing the reporting of assessment content topics, the expert advisors recommended that in the next version of the policies reporting instrument, further information be collected on the assessments being reported. The end of course assessments in mathematics are not given for a specific grade, but rather at completion of a high school course. They recommended that states report on content of the most advanced high school math assessment that most students take before graduating, either a comprehensive assessment (e.g., the HS assessment offered by the Smarter Balanced consortium) or the highest level end-of-course assessment most students take in high school.

State Science Assessment Content and Relation to Standards

The science assessment content section in the design pilot study focused on how and when states are transitioning standards and assessments to the Next Generation Science Standards.

The first set of items identified the official policy position of each state concerning the NGSS. We are interested in if states have changed their science standards, and then, the relationship of state science assessments to NGSS. Second, the data collection instrument should be designed to be completed in a reasonable amount of time through cooperation with the staff of a state education agency. With this assumption, it is likely that information can be collected on indicators of assessment content, but a complete content alignment study covering all standards and grade levels would be prohibitive. From input of the panel, it was decided that content alignment analysis of assessments to NGSS will need to be delayed until after more states have made this transition. For this year, the focus is on the timing and nature of the transition to new assessments.

The questions addressed: What is your state's official position about the Next Generation Science Standards in relation to your state-adopted K–12 content standards in science? Were your state standards adopted since 2013? As of 2015, four of the ten pilot study states reported they have adopted state science standards that are the same as NGSS, and two were adopted with minor revisions. Six states adopted standards since 2013, when NGSS was released.

Table 11. State Policy Position on Science Standards

		State Position on Next Generation Science Standards			
State	**Adopt since 2013**	**Standards Same as NGSS**	**Minor Revisions**	**Major Revisions**	**Unrelated**
1	Y	Yes			
2	Y	Yes			
3	N				Yes
4	N				
5	N				
6	Y	Yes			
7	Y		Yes		
8	N				
9	Y		Yes		
10	Y	Yes			

State Planning for New State Science Assessments

In responses to questions on Plans for New Science Assessments, states identified whether the state is planning new science assessments in relation to the NGSS and, if so, the target operational year for new assessments. The responses shown in Table 12 show that six states are planning new assessments, and five of the pilot states reported on the types of science assessment being planned as part of a new science assessment system. Four states are planning a combination of summative, benchmark, and formative assessments, and one state is planning for summative assessments only.

Table 12. New State Science Assessments by Year

State	New Assessments	Year Operational	With Other States
1	Yes	2017	Yes- 2
2	Yes	2016	
3	Yes	2017	CCSSO Collab.
4	Yes	2018	
5	Yes	2018 (gr. 6–8)	
6	Yes	2018	Yes- 2

The questions on state plans for new assessments, as reported in Table 13, provide a model for the full online policies reporting system to collect information from a broader set of states regarding their plans for new science assessments. The categories of reporting appear to be inclusive of options that states are planning, and the information is likely to be valuable to other states.

Table 13. Plan for State Science Assessments

State	Summative	Interim/ Benchmark	Formative
1	Yes	Yes	Yes
2	Yes	Yes	Yes
3	Yes	No	No
4	Yes	Yes	Yes
5	Yes	Yes	Yes
6	Not decided		

The pilot states were asked about use of the three major content dimensions of the structure for the NRC Framework for K–12 Education (same dimensions also in NGSS) in planning for their new assessments. Results reported in Table 14 indicate whether the three dimensions are part of the planning for new summative science assessments. Four states that reported on assessment content for new science assessments will include all three dimensions and all four intend to plan the assessment so that all three dimensions are addressed in the same sets of items.

Table 14. New Science Summative Assessments—Content Dimensions Planned

State*	Disciplinary Core	Science/Eng. Practices	Cross-cutting Concepts	Assess Together
1	Yes	Yes	Yes	Yes
2	Yes	Yes	Yes	Yes
3	Yes	Yes	Yes	Yes
4	Yes	Yes	Yes	Yes

* Pilot study states reporting plans for new science assessments by 2017–18.

Summary and Recommendations

In addition to reporting on state assessment policies in the pilot study, NORC staff asked several follow-up questions of the state education agency respondents about the comprehensibility of the survey items, whether the information needed to respond was available, whether there was assessment policy information that would be helpful but was not asked in the pilot survey, and the total amount of time needed to complete the survey.

States varied between two and six hours to complete the survey, with a median response time of four hours. For the most part, states found the intent and wording of the items to be clear and indicated that they had sufficient access to information to respond to the items. State respondents provided specific feedback about ways in which to improve the survey, and below we discuss the states' insights that will help inform the development of an improved instrument for reporting on assessment policies.

States Want More Information about the Development of the Science Assessments

Several states indicated that they are in the process of adopting new standards related to science, so it would be helpful for them to know more about the science assessment policies and practices of other states. For example, states wanted to know which companies would be involved with the development of the science assessments, whether there would be a system of assessments, and whether states expect the testing format to change. To the extent possible, states also

requested information about the development costs for the science assessments, while acknowledging that it may be difficult for other states to report or evaluate those costs.

Caution is Warranted When States Are Undergoing a Transitional Period

A few states cautioned that because they were implementing new assessments, their responses may reflect idiosyncrasies associated with the transitional period. For example, some states indicated that the reporting of results was scheduled for a later date than what was typically the case, and that they did not release test items from the current operational form because they had yet to build an available pool of releasable test items. This feedback suggests that future surveys may need to ask states that are undergoing a transitional period whether their responses are indicative of their typical policies and practices, or whether their responses were anomalous and influenced by the transitional nature of their assessment system.

For states undergoing a transitional period, it may be better to include a "don't know/yet to be decided" option so that the responses reflect policies that will be implemented as opposed to policies that are still under consideration or in flux.

Information to be Solicited Directly from the Consortiums

States administering the CCSS-aligned tests indicated that they could not answer certain items (such as those relating to item types or the percent of items requiring a calculator) because the information was available only to the consortium. The assessment content topics in consortium-developed assessments are incorporated into claims reporting categories which also include cognitive levels, and the content topics of certain items are not specifically reported in assessment blueprints. The adaptive nature of new assessments also limits reporting on information regarding item types used with students. To the extent possible, future surveys should solicit this information directly from the consortiums.

References

Achieve & Lead States (2013). *Next Generation Science Standards: By States, For States.* Washington, D.C.: National Academies Press. http://www.nextgenscience.org/next-generation-science-standards

CCSSO's (2014) Criteria for Procuring and Evaluating High-Quality Assessments http://www.ccsso.org/Documents/2014/CCSSO%20Criteria%20for%20High%20Quality%20Assessments%2003242014.pdf

CCSSO/ATP (2013) Operational Best Practices for Statewide Large-Scale Assessment Programs. http://www.ccsso.org/Resources/Publications/Operational_Best_Practices_for_Statewide_Large-Scale_Assessment_Programs.html

Darling-Hammond, L., Herman, J., Pellegrino, J., Abedi, J., Baker, E., Bennett, R. et al. (2013). *Criteria for high-quality assessment*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

Doorey, N. & Polikoff, M. (2016) Evaluating the Content and Quality of Next Generation Assessments. Washington, DC: Fordham Institute. http://edexcellence.net/publications/evaluating-the-content-and-quality-of-next-generation-assessments

National Governors Association Center for Best Practices and Council for Chief State School Officers. (2010). *Common Core State Standards*. Washington, DC: National Governors Association Center for Best Practices and Council of Chief State School Officers. http://www.corestandards.org/

National Research Council (2013). *Monitoring Progress toward Successful K–12 STEM Education: A Nation Advancing?* Washington, DC: National Academy Press.

National Research Council. (2013). *Developing Assessments for the Next Generation Science Standards.* James W. Pellegrino, Mark R. Wilson, Judith A. Koenig, and Alexandra S. Beatty, Editors; Committee on Developing Assessments of Science Proficiency in K–12.

Board on Testing and Assessment. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

NORC (2015) Study Design Paper: Developing a System for Tracking State Assessment Policies in Science and Mathematics Education Instrument Design and Pilot Study Plans, Bethesda, MD. March.

Polikoff, Porter, and Smithson (2011) How well aligned are state assessments of student achievement with state content standards? *American Educational Research Journal, 48*(4), 965–995.

Resnick, L.B., Rothman, R., Slattery, J.B., & Vranek, J.L. (2003). Benchmark and alignment of standards and testing. *Educational Assessment*, *9*, 1–27.

Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H. C., Wiley, D. E., Cogan, L. S., et al. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco: Jossey-Bass.

Webb, N. L. (2002). *Alignment study of language arts, mathematics, science, and social studies of state standards and assessments in four states.* Washington, DC: Council of Chief State School Officers.

U.S. Dept. of Education's Peer Review Guidance (issued 9/25/15) http://www2.ed.gov/policy/elsec/guid/assessguid15.pdf

Yuan, K. & Le, V. (2012). *Estimating the percentage of students who were tested on cognitively demanding items through the state achievement tests*. Santa Monica, CA: RAND Corporation.

Exhibit 1: Expert Advisors – Fall 2014 and Fall 2015

Catherine Bowler, Science & Engineering Assessment, Massachusetts Department of Education

Pete Goldschmidt, CSU-Northridge (formerly New Mexico State Assessment Director)

Brian Gong, Executive Director, Center for Assessment, Dover, NH

Robin Hill, Mathematics, Kentucky Department of Education

Tammy Howard, State Accountability Services Director, North Carolina

Peter McLaren, State Science Specialist, Achieve, Inc. (formerly Rhode Island)

Hillary Michaels, HumRRO, Education Research & Evaluation, Alexandria, VA

Jessica Mislevy, Education Research Scientist, SRI International

Brett Moulding, Science Education Professional Development Director, Utah

Tiffany Neill, Science Education Specialist, Oklahoma State Department of Education

Scott Norton, CCSSO Strategic Initiative Director for Standards, Assessment, & Accountability

Cinda Parton, Assessment Specialist, Washington State Office of Education

Jim Pellegrino, University of Illinois at Chicago, Learning Sciences Institute

William Schmidt, Center for the Study of Curriculum, Michigan State University

Ellen Whitesides, Mathematics Educator & Math SCASS, Washington, DC