A School Science Assessment System to Enhance Learning
A White Paper with Recommendations for New York State

New York State Science Education Consortium
A Collaborative Association of Professional Science Education Organizations in NYS

October 1, 2018

Contents

Introduction

The purpose of this paper is to provide perspectives and recommendations pertaining to the development and implementation of a system for assessing students' achievement of the New York State P-12 Science Learning Standards (NYSP-12SLS), adopted by the New York State Board of Regents in December, 2016 and published by the New York State Education Department (2017). As stated in the Statewide Strategic Plan for Science, adopted by the New York State Education Department (2018), the goal for assessment is to "Support the development of assessments at the state, regional, and local levels that measure student achievement of all new P-12 NYS science learning standards, and use the data resulting from these assessments to enhance teaching and learning."

Attainment of this goal is an essential element in achieving the mission of the Strategic Plan, which calls for us to: "create a Statewide learning community to enhance science education and improve student achievement of the New York State science learning standards leading to career and college readiness and a scientifically literate population capable of addressing the needs of society, participating in a global economy, and sustaining the physical and living environment."
The importance of assessment is further reinforced in the Strategic Plan's vision, which describes how the mission will be achieved:

> Ensure the teaching and learning of science for all P-12 students by providing equitable access to exemplary teachers, science curriculum programming, instructional practices, and standards-based assessments that are reflective of research and best practices, along with quality resources and support from stakeholders at large.

Together, the mission and vision of the Strategic Plan convey the belief that standards-based assessments that are reflective of research and best practices will play a critical role in supporting students' achievement of the new science learning standards leading to career and college readiness and a more scientifically literate population. The vision also suggests that reforming New York State's science assessments is a necessary but insufficient goal in achieving the mission of the Strategic Plan. Improvements in science teacher education, curriculum programs, instructional practices and resources and stakeholder support are also required.

New York State has a long, rich history in periodically revising its school science education standards and redesigning state assessments to evaluate students' achievement of those standards, as described by Ormiston, K. D. (1987). The recently adopted NYSP-12SLS represent an adaptation of the Next Generation Science Standards (NGSS), developed by NGSS Lead States (2013). The NGSS were internationally bench-marked, aligned with science learning research, and were based upon "A Framework for K-12 Science Education", published by the Committee on Conceptual Framework for the New K-12 Science Education Standards, Sciences, and Education (2012).

Effective implementation of the NYSP-12SLS will require significant changes in curriculum design, instructional practices, and assessment. The new standards require three-dimensional learning, where students utilize science and engineering practices (SEPs), disciplinary core ideas (DCIs), and cross-cutting concepts (CCCs) to explain natural phenomena and solve real-world problems. As the "Assessment Framework for Next Generation Science Standards" from the Council of Chief State School Science Officers (2015) notes,

> "explicit goals for science learning are outlined in the NGSS in the form of performance expectations (PEs). PEs are statements about what students should know and be able to do at the end of instruction (for specific grades in K–5 and for grade bands in middle school and high school), and the PEs prominently incorporate all three dimensions."

Currently, these practices, core ideas, and cross-cutting concepts are typically taught in isolation, if at all, and there has been no expectation in prior standards that these dimensions should be utilized in concert to help students achieve and demonstrate a more complex and challenging performance. But their integrated use by students in making sense of phenomena and problems better reflects what scientists do, and promotes richer and deeper conceptual understanding and greater proficiency in many aspects of scientific reasoning. To evaluate students' achievement of the new standards, assessment methods and instruments will need to change in significant ways in order to promote and measure three-dimensional performance.

A recent document, "Transforming Science Assessment: Challenges and Recommendations for States", developed by Achieve (2018b), discusses current impediments to revising science assessments in many states and suggests actions and strategies to address them. Some states have moved too quickly in revising their science assessments, resulting in instruments that fail to capture the essence of what students are now being asked to do. Other states have moved too slowly in revising their assessments and the momentum for change built after standards adoption has lessened.

This document is based upon research and discussion among science education leaders within the New York State Science Education Consortium, a cooperative organization of the following science educator professional associations and networks within New York State:

- Science Teachers Association of New York State
- Science Council of New York City
- Long Island STEM Education Leadership Association
- Capital Area Science Supervisors Association
- BOCES Science Representatives Network
- Biology and Chemistry Professional Development Network
- New York State Science Education Leadership Association
- Rochester Area Science Education Leadership Team

The recommendations in this paper were developed by 27 science education leaders at the Consortium's Seventeenth Statewide Science Education Summit, held at Clarkson University's Capital Region Campus in Schenectady, New York on August 6-8, 2018. Most of the participants are active science teachers, science supervisors, BOCES science specialists, or science teacher educators. Summit participants are listed in Appendix A of this paper.

In preparation for and during the Summit, participants reviewed many documents focused on developing new state and classroom assessments aligned with the NGSS. Some of these are included in the References provided at the end of this paper. While the ideas and recommendations in these documents may have informed and influenced the discussion of participants at the Summit, a primary intention was to develop recommendations that recognized the unique context, opportunities, and history of science education reform in New York State. Since participants brought a vast amount of experience as NYS science educators to the table, their collective lens was focused and used to develop important and credible recommendations most germane to this state.

Organization of the Paper

Perspectives and recommendations in this paper are organized into four sections:

- The Purposes and Priority Recommendations for a Science Assessment System
- Ensuring Quality in the Development of State Science Assessments
- Logistics for Development and Administration of State Science Assessments
- Supporting Development and Use of Classroom-Based Science Assessments

At the Summit, the topics identified above were explored and discussed by different working groups who reported out their recommendations to all participants for feedback and suggestions in plenary sessions. Each working group was provided with a set of focus questions to elicit and guide discussion while limiting redundancy across working groups. Similar focus questions will often be used in this paper to organize and present the recommendations from the Summit.

This white paper reflects the earnest and heartfelt work of the entire Consortium; the document was composed by a writing committee following the conclusion of the summit. The Consortium would like to thank this committee, which included Mary Loesing, Ellen Mandell, Helen Pashley, Bruce Tulloch and Joseph Zawicki, for their adept writing and final editing of the overall document. A special thank you to Bruce for his facilitating the writing and editing and to Joe for assisting with the final editing and document preparation.

The Purposes and Priority Recommendations for a Science Assessment System

What should be the purpose and priorities related to state science assessments within an assessment system?

State science assessments provide a means of monitoring the science learning of all New York State P-12 students and determining how well they are achieving the science standards in place. Whether a state assessment measures learning across several grades or within a particular subject, it can provide a valid and reliable benchmark of student achievement. How well students across the state perform on state assessments informs the work of state education specialists and policy makers who are charged with the responsibility of developing valid, reliable, and equitable assessments and using student performance data from those assessments in making judgments and setting policy. The monitoring function of state assessments is, or perhaps, should be, their overarching purpose. At the same time, student performance data on state science assessments can be aggregated at regional, district, school, and classroom levels and analyzed to reveal program strengths and needs that can help guide the efforts of administrators and teachers in improving local curricula, instruction, and assessment.

In order for state assessments to be effective instruments in evaluating programs and student learning, they must be valid and reliable measures of the expectations that the standards hold for students at the end of a grade, grade span, or subject. At best, educators can only make inferences about students' "true" levels of understanding and proficiency from observing how well students perform on assessments that are carefully aligned with the learning outcomes students are expected to achieve. If items within an assessment are not measuring what they are purported to measure and/or do not give consistent and reliable results, inferences drawn from student performance on such items will be spurious and decisions based upon those inferences will be tenuous at best.

While state science assessments can help to fulfill the purpose of monitoring how well students are achieving state science standards, other purposes can be cited and some may be especially important as the state and its schools transition to new science standards. Each of these purposes will be briefly described below and priority recommendations to support the fulfillment of each purpose will be provided.

1. First, state science assessments can serve as examples of high quality assessment practices and design that could inform classroom-based assessments, as noted by Achieve (2018a). Many science teachers currently design and/or use local assessments that closely align to the content emphasis and format of existing state science assessments. While this practice can limit opportunities that students have to demonstrate their knowledge and proficiencies in local settings, new and innovative designs for state assessments aligned with new standards can help guide and support the development and use of both formative and summative three-dimensional classroom-based assessments by science teachers.

   Priority Recommendations

   (a) Develop an assessment system that integrates state and classroom-embedded components of assessment.

      i. Classroom-embedded components should be aligned with state assessment measures (implemented locally but facilitated statewide).

      ii. Student scores should reflect student performance on both summative statewide assessments and classroom-embedded assessments.

   (b) Ensure that all state assessments are composed largely of multidimensional tasks. Tasks must include:

      i. Students' use of the science and engineering practices, crosscutting concepts and disciplinary core ideas to make sense of phenomena and/or solve problems presented through real world scenarios.

    ii. Multi-modal formats (computer simulations, performance-based elements, paper-pencil items, etc.)

   (c) Develop and release sample tasks to serve as exemplars of high quality assessment practices/design that could inform the development and use of new classroom-embedded assessments (Achieve, 2018a).

   (d) Ensure that state assessment writers are well-versed in three-dimensional instruction and assessment development (have demonstrated understanding and experience with the instructional shifts necessary).

   (e) Establish time-lines that allow sufficient time in the assessment development processes to engage in rigorous and iterative design, evaluation, and refinement, as noted by Achieve (2018b).

2. Second, state science assessments can leverage opportunities for teacher and administrator professional development. Science teachers and administrators are acutely interested in learning how state science assessments will change in order to reflect the expectations of the new science standards. They will also be interested in revising their local curricula and instructional programs and strategies to support student achievement of the new standards and successful performance on the new state assessments. Given the instructional and assessment-related shifts required, significant amounts of professional development will be needed to assist local educators in making these transitions. In addition, science teacher educators who prepare pre-service science teachers for service in the schools will need to change their instructional focus and practices regarding the new standards and their aligned assessments. These individuals can also benefit from targeted professional development opportunities.

   Priority Recommendations

   (a) A coordinated and systemic program of professional development on assessment should be provided for all P-12 teachers and administrators National Research Council (2012).

   (b) Quality professional development on the new P-12 science standards and assessments should be provided to all science teacher educators who work with pre-service and in-service science teachers in undergraduate and graduate programs of science teacher preparation at colleges and universities.

   (c) Awareness sessions to enhance parent and community members' understanding of the goals and purposes of various assessments (classroom-based, interim, state-level) should be provided in face-to-face and/or online venues.

3. Third, the student performance data from state science assessments can inform district-level program design and revision at both elementary and secondary levels. In this era of data-driven decision making at all levels of the education hierarchy, it is imperative that the inferences drawn from the results of state science assessments be sound and useful. Priority Recommendations

   (a) Assessment tasks and items must be robust enough to support programmatic decision-making

    i. Psychometricians should coordinate work with test development teams during the test development process to ensure validity and reliability of tasks on state science assessments.

    ii. Analytical reports of student results should be reported in alignment with the DCI, SEP and CCC dimensions in the standards and provide data on item difficulty, response pattern, and discrimination.

   (b) District-wide assessment data should be provided to support district and department-level data-driven decisions in a timely manner.

4. Fourth, the state science assessments must ensure equity across diverse student populations (socioeconomic status, new language proficiency, special needs, etc.). The performance of students with diverse

backgrounds and abilities on new assessments should be an accurate measure of their achievement of the science standards and not be largely determined by social, language and cultural deficits.

Priority Recommendations

(a) Assessments should avoid unnecessarily large amounts of text, ambiguous language, complex terminology, and culturally-restrictive scenarios and tasks.

(b) Assessments should provide rubrics that are appropriate for the accurate scoring of open-ended responses from students with diverse backgrounds and abilities.

(c) Districts should be encouraged to monitor and evaluate opportunities to learn among their student populations.

5. Finally, state science assessments should be designed and utilized for the purposes for which they are intended (i.e. support and monitor students' achievement of standards and contribute towards fulfilling graduation requirements). Additional applications, such as use in teacher evaluation, are inappropriate.

The implementation of new state science standards provides an opportunity to review state, regional, and local education requirements, policies, and procedures and make revisions that can help ensure that all students achieve the new life, earth and space, and physical science standards at elementary, intermediate, and commencement levels. The scope, organization, and emphasis of new state science assessments can further encourage schools to provide science programs that balance breadth and depth in the experiences provided to all students. As was said during the implementation of the National Science Education Standards 20 years ago, standards and their assessments should promote programs that "pump" up all students to higher levels of achievement, not serve as "filters" to sort students into course sequences that ultimately restrict their learning and future opportunities.

This is a legitimate purpose for new assessments while their use in evaluating teachers and administrators is problematic on the basis of educational measurement theory and can narrow the scope and focus of future curriculum development and instruction. Linking student performance on new state assessments with professional evaluation will also limit educators' willingness to make the instructional shifts necessary to help students engage in the three-dimensional learning envisioned by the new state science standards.

Priority Recommendations

(a) The state, districts, and schools should ensure that all students at all grade levels achieve the life, earth and space, and physical science standards, as delineated by the new NYSP-12 science learning standards.

(b) District's Annual Professional Performance Reviews of teachers and administrators should be uncoupled from student performance on the new state science assessments.

Ensuring Quality in the Development of State Science Assessments

What processes can help ensure the development of equitable and meaningful new state science assessments?

If new state science assessments are to measure and monitor students' achievement of the NYSP-12SLS accurately and effectively, then these assessments must be designed to lead to valid inferences about what students know and can do in reference to the performance expectations and dimensions specified in the new standards. The National Research Council (U.S.) Committee on a Conceptual Framework for New K-12 Science Education Standards (2012) has noted the following as necessary conditions for developing and administering such assessments:

- "Assessment development employs theories and data about content-based cognition that indicate the knowledge and practices that should be tested" (NRC, 2012, p.264).
- "Developed tasks use student assessment data to provide information on whether the student has mastered the knowledge and practices of interest" (NRC, 2012, p.264)
- "Both qualitative and quantitative techniques are included for scoring student performance that captures fairly the differences in knowledge and practice" (NRC, 2012, p.264)

There are many processes that must be undertaken if equitable and meaningful new state science assessments are to be developed. Each of these processes will be described below and includes recommendations as appropriate.

1. First, course maps must be developed that clearly communicate which Performance Expectations (PEs) in the NYSP-12SLS are to be assessed in grades K-5 and in each of the four high school science courses. A similar course map for science courses provided in grades 6-8 is strongly advised. Such course maps will also guide the local development of science curricula, courses, instructional plans, and assessments.

2. Second, those who are responsible for the development, administration, and scoring of new state science assessments (i.e., New York State, Regional Education Agencies, School Districts and Schools) should agree on a set of essential attributes or expectations that will guide statewide implementation of the assessments.

   Essential Attributes or Expectations

   - Fidelity to three-dimensional teaching and learning is evident.
   - The overarching purpose of the assessments is communicated to stakeholders.
   - Assessments align to PEs and all components of the PEs at the appropriate grade level. Each assessment clearly allows students to demonstrate their learning in the three dimensions. and application within each dimension.
   - Equity across all subgroups and regions of the state is evident in each assessment.
   - State assessment is part of a classroom-based through state-level assessment system.
   - The purpose of each type of assessment (classroom, district, regional, state) is defined and communicated. ("An assessment designed to provide information about students' difficulties with a single concept so that it can be addressed with instruction would be designed differently from an assessment meant to provide information to policymakers for evaluating the effectiveness of the overall education system" (NRC, 2012, p. 26)).

- Every assessment is accessible to students who have had the opportunity to achieve the learning standards related to the assessment.
- Adequate teacher and student preparation, resources, time and learning opportunities are provided to support successful implementation.
- Scoring guidance documents ensure inter-rater reliability across the dimensions throughout the state.

3. Third, those who are responsible for the development, administration, and scoring of the new state science assessments must agree on the definition of a task and its criteria and constraints. Since the NYSP-12SLS require students to apply disciplinary core ideas, science and engineering practices, and cross-cutting concepts to discover and explain phenomena and to solve real-world problems, new state science assessments aligned to these standards will need to be substantially different in form and substance than previous state assessments. National and state efforts to develop new assessments that show fidelity to three-dimensional learning have honed in on the "task" as the most promising format for use in evaluating students' attainment of PEs and their dimensions. If new state science assessments employ tasks as the primary means to evaluate student learning, there must be common agreement on the essential attributes of such a task on the part of all those who develop, administer, and/or score the new assessments. One definition of a task is provided: "A task is any learning activity or assessment that asks students to perform to demonstrate their knowledge, understanding, and proficiency." Performance tasks yield a tangible product or performance that serves as evidence of learning" as noted by Defined Learning (2018), among others.

   Achieve, NSTA (2016) have created a "Science Task Screener" for use in determining if a task meets criteria needed for it to be a valid and reliable indicator of student achievement of one or more performance expectations. These criteria include:

   - Tasks are driven by high-quality scenarios that focus on phenomena or problems
   - Tasks require sense-making using the three dimensions
   - Tasks are fair and equitable
   - Tasks support their intended targets and purposes

   The National Research Council (U.S.) Committee on a Conceptual Framework for New K-12 Science Education Standards (2012) has stated that "the more precisely a design task's criteria and constraints can be defined, the more likely it is that the designated solution will be successful." Turning to the construction of three-dimensional tasks, there is a growing body of literature making recommendations about the elements that such tasks should contain and what the tasks should require students to do (Achieve (2017), Council of Chief State School Science Officers (2015), National Research Council (U.S.) Committee on a Conceptual Framework for New K-12 Science Education Standards (2012), SRI (2018). Achieve (2018b) describes the development and use of high-quality tasks as follows:

   > The assessment tasks students engage with should be grounded in rich and meaningful scenarios that are driven by problems and phenomena, and they should elicit student thinking and reasoning via the science and engineering practices, disciplinary core ideas, and crosscutting concepts. Ensuring that these tasks are of high quality and truly embody the standards should be a focal point of assessment development processes, with careful and intentional design, appropriate research and cognitive labs, and regular quality control embedded.

While sources vary to some extent in the kinds of multidimensional tasks they describe, recent work by Achieve (2018a) recommends that each task present a phenomenon or problem through the use of a real- world scenario and employ one or more prompts (questions or items) to elicit the targeted knowledge and skills. A stimulus can be thought of as the means through which a scenario or a prompt based upon a scenario is framed and communicated. Scenarios, stimuli and prompts can be defined as follows:

- Scenario: Phenomenon or problem situation used to elicit students' capabilities in a meaningful setting. This could include language intended to describe the phenomenon, datasets and observations, interactive stimuli, etc. Achieve (2017).
- Stimulus: A component of an item cluster (or task) that does not directly require a student response. A stimulus can include one or more of the following: text, audio, video, animation/simulation, experimentation, discussion, activity, and/or demonstration. Council of Chief State School Science Officers (2015)
- Prompt: A question (or item) posed to students to elicit targeted knowledge and skills. Prompts may comprise multiple parts, but they represent the lowest common unit of analysis—score points are assigned for sub scores at the prompt level Achieve (2018a).

It is recommended that the phenomenon or problem, stimulus, and prompts for a multi-dimensional task have the attributes specified below:

- Phenomenon or Problem: Essential Attributes
  - The scenario presents real-world observations
  - The phenomenon or problem is presented as puzzling or intriguing
  - The phenomenon or problem creates a need to know
  - The phenomenon or problem is explainable using grade appropriate SEPs, CCCs, and DCIs
  - Local, global or universal relevance is made clear to students in the scenario
  - The phenomenon or problem is comprehensible to a wide range of students at grade level
  - The phenomenon or problem uses as many words as needed, no more
  - The phenomenon or problem is sufficiently rich to drive the task
- Stimulus: Essential Attributes
  - The stimulus in a scenario or series of prompts effectively uses at least two modalities, enabling the content to be presented so that students can use different senses and different skills while completing the task
  - The stimulus contains real, well-crafted data or data sets as needed
- Prompts: Essential Attributes

- The majority of the prompts require students to use reasoning to complete the task successfully The prompts require students to use at least two of the following dimensions in successfully completing the task, with the goal of using all three dimensions.
  - Elements of the DCIs
  - Science and Engineering Practices
  - Crosscutting Concepts
- The prompts require students to figure something out in relation to the phenomenon or problem posed
- The prompts require students to engage in sense-making
- The prompts present questions in a variety of formats (multiple-choice, constructed response, etc.)

4. Fourth, test blueprints must be developed that show fidelity with three-dimensional science assessment while addressing the practical issues involved in administering large-scale summative assessments to statewide student populations. It is not possible for a single state science assessment, administered in two-three hours, to be designed and used to evaluate students' achievement of all the performance expectations specified for a given course of study. Consequently, a sampling of student performance on assessments with tasks aligned to a smaller set of relevant performance expectations will be necessary. Test blueprints can provide test developers, school administrators, curriculum supervisors, and teachers with necessary and appropriate information to inform their practice and help ensure that assessments are developed in as transparent a process as possible. Recommended guidelines for the development of test blueprints are provided below.

   Guidelines for Developing Test Blueprints for State Science Assessments

   - To the extent possible, blueprints for all state science assessments should be consistent in the types of information provided and how this information is formatted.

   - To provide valid evidence to support the claims for a given state science assessment, attention must be given to choosing a representative set of PEs that adequately sample the breadth and depth of course expectations for student learning. This set of PEs must be communicated by the test blueprint.

   - To determine how many PEs can be evaluated in a given state science assessment, attention must be given to the amount of time it takes students to complete tasks.

   - The standards, elements, competencies, knowledge, and/or skills being assessed should be defined specifically enough in test blueprints to allow differentiation from other likely interpretations by intended users, and specifically enough to guide test development Achieve (2017).

   - Test blueprints for a given state science assessment should be modified annually over a span of 3-5 years to ensure that student achievement of all PEs for a grade band or specific discipline are evaluated in this time frame.

5. Fifth, the State Education Department must prepare its item writers and editors to write and develop high-quality tasks for new state science assessments that are three-dimensional and provide valid and reliable evidence of students' achievement of the NYSP-12SLS. New policies and procedures will need to be considered and adopted to help ensure that new science assessments are aligned and show fidelity with the P-12 science standards. The following recommendations will help NYSED achieve this goal.

Recommendations for Task Development, Evaluation, and Use

1. Item writers will need explicit professional development focused on task development. Further professional development must be provided focused on evaluating the fidelity of a task to the SEP, DCI, and CCC that are targeted and at the appropriate level.

2. Item writers should be instructed to use Task Development Templates created by NYSED's Office of State Assessment (OSA). Each Template should include the PE and its SEP, DCI, and CCC explicitly documented at the appropriate grade level. Thus each Template should be standardized for each PE. (Note the templates developed by Paul Andersen on "The Wonder of Science" website: https://thewonderofscience.com/)

3. Multidimensional tasks should be written by individual item writers from across the state, representing all demographics.

4. Multidimensional tasks should be evaluated by item writers collaboratively, using a Three- Dimensional Fidelity Rubric (Achieve) and revised as necessary.

5. Revised multidimensional tasks should be forwarded to item editors whose further analysis of a task should include these steps:

   (a) Use of a "quick screener" to evaluate the task for fidelity to the three dimensions of the relevant PE.

      i. If the task does not screen accurately to the three dimensions, it will be eliminated from the pool.
      ii. If the task does screen well, the scientific information provided by the task must be evaluated for accuracy and relevance and revised as needed.

   (b) The task should be reviewed to determine where the three dimensions (SEPs, DCIs, and CCCs) need to be utilized by students in understanding and responding to the phenomenon or problem presented, the stimuli used to present information and data, and the prompts employed to generate student responses required by the task.

6. Multidimensional tasks should be field tested with diverse populations of students at the appropriate grade level to measure their difficulty, discrimination, and to generate student responses that can be used in the development of scoring guides and rubrics.

7. Multidimensional tasks should be used to generate a range of student responses that reflect different levels of understanding and proficiency in using SEPs, DCIs, and CCCs to explain phenomena and solve real-world problems while demonstrating achievement of a Performance Expectation. If the tasks are shown to be valid and reliable measures of student achievement of PEs, then they should be retained and included in state science assessments. Range finding and standards setting processes should be employed to ensure that the new assessments challenge but do not frustrate students in their attempts to demonstrate what they know and can do.

8. As multidimensional tasks are developed, revised, and administered to students in field-test protocols, some of the tasks along with their student performance data should be shared with teachers, curriculum administrators/supervisors (Task Samplers) in the schools to facilitate the local changes necessary in curriculum development, instructional practices, and assessment that support a transition to the NYSP-12SLS.

Logistics for Development and Administration of State Science Assessments

Unless they have been involved in the development of state science assessments, few educators can fully appreciate the amounts of time, effort, and expertise that are put into every state assessment by the testing specialists in the State Education Department and their educational specialists who write and edit test items, construct field tests, and design, review and revise operational tests before they are administered statewide. Unlike many states, New York State has drawn upon the talents and experience of classroom teachers and other science educators in all phases of the development of its science assessments rather than contract with commercial test development firms.

In the context of the NYS MST learning standards (which were adopted in 1996) and the derivative Curriculum Core Documents, current state science assessments are reasonably valid and reliable measures of student achievement of those standards, although the emphasis of these assessments has been on knowledge and basic understanding and not on the science inquiry skills or common themes that are also included in the 1996 standards.

With the approval and implementation of new standards that seek a balance in learning and using disciplinary core ideas, science and engineering practices, and cross-cutting concepts, student achievement of these standards will need to be assessed, evaluated, and reported in new ways. While it seems clear that designing the kinds of new state science assessments described in the earlier section will be challenging, it is also apparent that the mission and vision of the state's Strategic Plan for Science will not be achieved unless such assessments are developed and implemented during the next few years.

As was discussed in the first section of this paper, a state assessment system and state-developed assessments serve a number of important purposes such as monitoring student achievement, leveraging professional development and supporting local program design and modification. Perhaps most important, science instruction in many classrooms will not change if the efforts of students, teachers, and administrators to transition to new standards are not recognized and rewarded by assessments that validly and reliably measure student achievement of those standards.

While a strong argument can be made for using new standards first to revise local science curricula, instructional practices, and classroom-based formative and summative assessment before following through with new state assessments, this argument fails to recognize the climate of teacher and administrator accountability that persists in the state. Student performance on current state science assessments that are aligned with dated learning standards are still used to evaluate teachers and administrators, either as part of the Annual Professional Performance Review (APPR) or more informally. By this measure, most teachers and administrators have been judged as effective or highly effective.

So until there is evidence that significant changes in the state science assessments are imminent and until the APPR better rewards teachers and administrators for developing new programs and trying new instructional strategies aligned with the NYSP-12SLS, there is little incentive for school personnel to make the transitions necessary for fundamental reform in P-12 science education. When the state sets time-lines for implementing new sciences assessments and when guidance and support such as professional development opportunities or test samplers are provided for schools and teachers, school administrators at the central office and building levels will more likely allocate the necessary funding, time, and other resources to help teachers undertake the transitions needed to prepare for new assessments.

Given the above discussion, this section presents recommended time-lines for the development of new state science assessments so that the momentum created by the approval of the Strategic Plan for Science and the adoption of the NYSP-12SLS by the NYS Board of Regents can be sustained and schools can synchronize their own local planning with the planning taking place at NYSED.

Year	Tasks	Details
2018-2020	transitional test development of all exams, including training of item writers	item writing for all exams
	field tests of aligned Grade 5 & Grade 8 Exams, transitional exams in high school sciences	continue item writing for all exams
2020-2021	administration of transitional exams in grade 5, grade 8, Earth Science & Life Science	ongoing item writing for all exams
2021-2022	administration of new, aligned grade 5 & grade 8 exams and transitional exams in Earth Science, Life Science, Chemistry & Physics	ongoing item writing for all exams
2022-2023	administration of aligned grade 5 & grade 8 exams, new, aligned Earth and Life Science exams, transitional exams in chemistry & physics	ongoing item writing for all exams
2023-2024	administration of aligned fifth & eighth grade exams, Earth and Life Science exams, and new, aligned chemistry & physics exams	ongoing item writing for all exams

Specifically, a transitional exam could include a set number of multidimensional tasks as described above in addition to the typical test items and formats of the current state examination. This would allow both students and teachers to gain experience with multidimensional tasks and help maintain the momentum to align instruction to the new science standards.

What is a reasonable and practical time-line for the development of new state science assessments?

The Strategic Plan for Science calls for new science assessments to be administered at Grade 5, Grade 8, and at the end of each high school Regents course of study. The state testing and reporting requirements of the federal Every Student Succeeds Act (ESSA) make it difficult to establish a moratorium on current state science assessments for one or two years in order to help focus teachers' attention on making the instructional shifts called for by the new standards. However, Transitional Exams at Grades 5 and 8 and in the Regents science courses would be beneficial and ease anxieties of both teachers and students.
The New York State Standards Implementation Roadmap provides proposed time-frames for the three phases related to standard implementation:

- Phase I: Raise Awareness and Build Capacity during 07/2017 – 08/2019
- Phase II: Transition and Implementation during 09/2019 – 08/2021
- Phase III: Implementation and Sustainability during 09/2021 – 08/2014

The following chart provides a time-line for test administration activities that are consistent with the time-frames above and achieves the development and administration of new state science assessments aligned to the NYSP-12SLS for Grades 5 and 8 and for high school by June, 2024.

In what time-line and sequence should the six new state science assessments be rolled out?

Aligned Assessment	Transitional Exams	First Aligned Assessment
Grade 5 (See Note 1)	Spring 2021	Spring 2022
Grade 8 (See Note 2)	Spring 2021	Spring 2022
Earth Science	June 2021, June 2022	June 2023
Life Science	June 2021, June 2022	June 2023
Chemistry	June 2022, June 2023	June 2024
Physics	June 2022, June 2023	June 2024

Note 1: Research needs to be done to determine whether a cohort of elementary students could not take the 4th grade science assessment in 2020 and instead take the Transitional 5th grade science assessment in 2021. Doing this would still enable school districts to meet the requirement of testing students during the elementary grades. We would strongly suggest that elementary students not be tested in both 4th and 5th grades using two different sets of standards.

Note 2: Consider creation and implementation of statewide classroom embedded performance tasks in grades 2 through 8 to be administered throughout the school year. Data from these would serve as program evaluation in grades 2, 3, 4, 6, and 7, and pupil evaluation as part of summative scores for grades 5 and 8. If these tasks are required, then the grade 5 exam should assess only the 3-5 band, and the grade 8 exam should assess only the 6-8 band.

What steps could NYSED take to develop pools of item writers, item editors, and test developers needed to develop the new state science assessments?

It was noted in the preceding section that the State Education Department must prepare its current and new item writers and editors to write and develop high-quality tasks for new state science assessments that are three-dimensional and provide valid and reliable evidence of students' achievement of the NYSP-12SLS. Beyond the writing and editing of the tasks themselves, field tests will need to be designed and the data collected from field testing will need to be used in the development of the assessments. Range finding and standards setting will also be necessary. All of these activities will require the efforts of teachers and other science educators working individually and/or in committees to prepare the new state science assessments and to assure that they are valid, reliable, and equitable measures of student achievement of the NYSP-12SLS. Both new and current test consultants will need professional development and training to fulfill their responsibilities. The recommendations provided below will help to ensure that current and future test consultants will be prepared to develop the high-quality state science assessments required by the new science standards.

Recommendations to help build the expertise needed in developing new state science assessments

1. Links for teachers to apply for service as test developers and on committees needs to be communicated through BOCES, professional organizations, and list-serves (superintendents, district coordinators, science teachers, etc.)

2. Funds will need to be budgeted to pay for additional item writers because of the enormity of the task of creating three-dimensional assessment items.

3. Consultant training should be held regionally and/or concurrently with other functions to reduce or eliminate travel-related difficulties.

4. Preference in selecting candidates for item writing and standards setting should be given to current classroom teachers who have the appropriate training.

The following chart provides a general time-line for many of the test development steps that will be necessary to take in order for the NYSED and its schools to transition to new state science assessments over the next several years.

Step	Timeline	Responsibility
1	Create Committees of Item Writers, Test Developers, Editors, New Reference Table Developers for ALL grades/subjects Fall 2018	NYSED
2	Train Committee Members Spring/Summer 2019	NYSED/Experts
3	Decide on Test Format/Blueprint Summer 2019	Test Developers
4	Create Pool of Multidimensional Tasks Summer 2019 and ongoing	Item Writers
5	Field Test Development for NYSP-12SLS-aligned Grade 5, Grade 8, and Transitional Regents Exams Fall 2019	Item Writers
6	Field Test Administration for NYSP-12SLS-aligned Gr 5, 8, and Transitional Regents Exams Spring 2020	NYSED/School Districts
7	Analysis of Field Test Data Summer 2020	Test Developers
8	Release Test Samplers for NYSP-12SLS-aligned Grade 5, Grade 8, and Transitional Regents Exams in Earth Science and Life Science Fall 2020	NYSED
9	Field Test Development for NYSP-12SLS-aligned Earth Science and Life Science Assessments Fall 2020	Item Writers
10	Field Test Administration for NYSP-12SLS-aligned Earth Science and Life Science Assessments Spring 2021	NYSED/School Districts
11	Analysis of Field Test Data Summer 2021	Test Developers
12	Release Test Samplers for NYSP-12SLS Earth Science and Life Science Assessments Fall 2021	NYSED/School Districts
13	Field Test Development for NYSP-12SLS-aligned Chemistry and Physics Assessments Fall 2021	Item Writers
14	Field Test Administration forNYSP-12SLS-aligned Chemistry and Physics Assessments Spring 2022	NYSED/School Districts
15	Release Test Samplers for NYSP-12SLS Chemistry and Physics Assessments Fall 2022	NYSED/School Districts

What steps could be taken to synergize the state and classroom-level development of new tasks and item formats? (i.e. draft state test samplers, field testing of tasks, etc.)

As was noted earlier in this paper, the NYSP-12SLS can only be achieved if a science assessment system is designed, including both state-level and classroom-based components. If the completion of multidimensional tasks will be the primary means by which students demonstrate their three-dimensional learning and achievement of the standards, then it will be important for NYSED test developers, regional education specialists and classroom teachers to work together in the design, development, pilot testing, and refinement of such tasks. Taking the steps listed below will help the state and its schools build the capacity to develop multidimensional tasks that can support the transition to new science standards in our classrooms while expanding the pool of tasks that can be evaluated for potential use on the new state science assessments. With respect to these steps, it will be necessary to determine who has the authority and capacity to initiate a specific initiative (NYSED, BOCES, Large City, Professional Association, etc.) and who needs to be involved in sustaining the initiative.

Steps to create synergy between the state and classroom-level development of new tasks and item formats

- Task all BOCES/Big Five City, educational organizations to run workshops for teachers and administrators on three-dimensional task/assessment development. These workshops must be planned and presented by science teachers or other science educators. Where a BOCES or Big Five City lacks the capacity to provide such workshops, other mechanisms to build local capacity relating to three-dimensional assessment must be designed and implemented.

- Agree upon a statewide three-dimensional assessment screening tool/rubric with agreed upon criteria based on NYSP-12SLS priorities (this could be taken on by STANYS and other Consortium members if given authority from NYSED to perform these tasks).

- Require BOCES/Big Five City to coordinate task development regionally and submit exemplar tasks for creation of an item bank that all teachers can access and utilize as needed. Ideally, this item bank will be maintained at NYSED but the bank could also be maintained at a BOCES or Large City site where staff are charged with uploading, organizing, and administering the bank so that teachers can access and use tasks from a designated website.

- Create a number of (at least 5) statewide classroom embedded tasks to be administered throughout the school year in appropriate grades, grade bands, or science disciplines from which student performance data can be used to inform the improvement of state summative assessments and contribute to a state summative assessment score. (possibly these classroom embedded tasks could be used to replace performance assessment at the end of year)

- Compile and make lists/links of high quality instructional/assessment resources available at all grade levels (usable data sets, unit plans, lesson plans, formative/interim/summative three-dimensional assessment tasks)

Supporting the Development and Use of Classroom-Based Science Assessments

What should be the purpose and priorities related to classroom-based science assessment within a science assessment system?

In "21st Century Science Assessment: The Future is Now" SRI (2018), James Pellegrino argues that a balanced system of science assessments is needed to serve the different purposes and needs of state education officials, district and school administrators, and classroom teachers. One form of assessment does not and cannot serve all the appropriate purposes and needs of various actors in the educational system. Thus, it is inevitable that multiple assessments will be required to serve the varying science assessment needs of different audiences, ranging from classroom teachers to state and national policy makers (p.18).

The recent NRC report Developing Assessments for the Next Generation Science Standards Pellegrino, J, Wilson, R., Koenig, J. and Beatty, A. Editors. (2014) recommended a systems approach to science assessment that included a balance among three components designed to support and complement one another:

- assessment designed to support classroom instruction
- assessments designed to monitor science learning on a broader scale; and
- a series of indicators to ensure that students are provided with adequate opportunity to learn science in the ways laid out in the Framework for K-12 Science Education (2012) (p.19)

So while this paper has thus far focused largely on the development and administration of new state science assessments, we now turn to a discussion of the role of classroom-based assessments in a balanced system of assessments. Formative and summative assessments in the classroom may share the student monitoring function with state assessments, but they also serve other purposes. Classroom-based assessments provide teachers with immediate feedback on the progress being made by students in achieving the performance expectations that frame and guide the instructional program. This information should inform the frequent and ongoing decision-making of teachers regarding the scope of the curriculum, the choice of instructional strategies, the pacing of instruction, the need for remediation or review, the efficacy of the assessments being used, and other considerations. Local assessments can also inform students and their families about their progress in meeting the standards and point to areas of strength and identify needs for improvement.

This "metacognitive" function of assessments can be very powerful in promoting more focused and enhanced learning. Finally, classroom-based assessments can communicate not only the expectations of the course and the teacher but also the expectations of the district and state if they are coordinated with state assessments in a balanced system that is comprehensive, coherent, and continuous (Pellegrino, SRI Education White Paper, 2016). Clearly, this purpose cannot be fulfilled if local assessments and state assessments are not closely aligned and coordinated in regards to the standards that students are to achieve at the end of a grade, grade band, or science subject.

As the NYSP-12SLS are implemented statewide, transforming classroom-based assessments from their current emphasis and form to strategies, tools, and instruments that assess three-dimensional learning will be challenging. If state science assessments focus on using multidimensional tasks where students are presented with scientific phenomena or problems and required to use DCIs, SEPs, and CCCs to explain the phenomena or solve the problems, then it is appropriate for classroom teachers to also shift to the use of such "phenomena-based" assessments or tasks as they transition from the MST core standards to the NYSP-12SLS. Whether teachers individually design and use such tasks themselves or implement tasks developed by external sources, they will need substantial levels of ongoing professional support to evaluate if the tasks are valid and reliable indicators of three-dimensional learning.

Teachers must also learn how the tasks can be optimally integrated into their instructional programs to achieve the major purposes of classroom-based assessment noted above. If students throughout the state are to achieve the NYSP-12-SLS and demonstrate this achievement through successful performance on state science assessments, then they must be prepared by teachers who have learned how to both teach and assess students three-dimensionally. Such learning cannot be accomplished unless professional development programs are targeted, well-crafted, readily available, and provide teachers with the opportunity to apply their learning in their classrooms and receive ongoing feedback on the results of their applications.

How can we help build local capacity to develop the needed classroom science assessments?

Teachers need to see an entity in a position of leadership provide a vision of what new classroom-based assessments should look like as they consider how to implement phenomena-based, multidisciplinary assessments in their classroom. Clearly, the development and distribution of "Test Samplers" by the State Education Department, containing examples of multidisciplinary tasks and other test formats, can help inform the development and use of three-dimensional, classroom-based assessments in the schools. However, the samplers will likely present examples of the type of tasks that will be found on statewide summative assessments while the type of tasks needed for formative, interim, and summative assessments in a classroom, school, and district may require a different level of specificity and form to be successfully embedded in the instructional program and take account of the learning progressions that occur during the school year or across several years. Consequently, professional organizations of science educators, such as STANYS and the NYS Science Education Consortium, might play significant roles in guiding teachers during their transition to three-dimensional instruction and assessment. Such organizations have members with the expertise and experience to design prototype classroom-based assessments, pilot and field test them in the schools, and revise and disseminate them to science teachers through their professional development initiatives and other venues.

STANYS and its sister associations and networks within the NYS Science Education Consortium might also take additional steps to help support teachers in the development and use of three-dimensional assessments at the local level.

Steps to help teachers develop and/or implement classroom-based assessments aligned with the NYSP-12SLS

- Provide "one-stop" shopping for support materials by referring teachers to websites such as the NYS Science Education Consortium websiteNYS Science Education Consortium (2018), "The Wonder of Science" website Anderson, P. (2017), the LISTEMELA PD Resources website/citeLISTEMELA, and websites from Achieve and the National Science Teachers Association National Science Teachers Association (2016).

- Provide a vehicle to post exemplars and locally-developed assessments for teacher reference. Provide resources such as data sets or possible anchoring phenomena for teachers to use while developing their classroom-based tasks and other assessments (Having access to this material will reduce the time needed for teachers to independently search for rigorous tasks and allow them to focus on developing stimuli and prompts to direct students to utilize DCIs, SEPs, and CCCs in interpreting and explaining phenomena or solving problems.)

- Provide assistance to help teachers recognize how to integrate the CCCs and SEPs into DCI scaffolding. Teachers need to recognize the shift from MST-based assessments to phenomena-based assessments.

- Urge that school districts provide professional development or calendar time for Professional Learning Community/Common Planning among teachers charged with developing and/or evaluating three-dimensional assessments.

- Offer introductory professional development to provide teachers with a preliminary exposure to phenomena-based classroom assessments.
- Provide continuous professional development opportunities throughout the year that enable the new assessments being developed and used in classrooms to be evaluated for proper alignment and richness and revised as necessary.
- Provide teachers access to exemplars of what classroom-based assessments could look like after instruction while maintaining the rigorous nature of the assessments at each grade band.

This paper has stressed the importance of providing carefully planned and systemic professional development and sources of exemplary tasks in preparing science teachers and other educators so they understand why new state and local three-dimensional assessments are necessary and how teachers can begin to develop and/or evaluate phenomena-based tasks and other formats to ensure these assessments are valid and reliable measures of three-dimensional learning aligned with the NYSP-12SLS. While these resources are viewed as necessary in transitioning to a new science assessment system, they are not sufficient to achieve the mission and vision of the Strategic Plan for Science. Science teachers, science supervisors, and building administrators will also need to know that their efforts in transitioning to three-dimensional classroom-based assessments are encouraged and supported by their central office administrators, Boards of Education, and the State Education Department. Any change or innovation carries the risk of failure and the growing pressures upon teachers, administrators, and schools for transparency and accountability have made many educators risk-adverse. “Top-down leadership for bottom-up reform” was the state mantra for reform a generation ago. As teachers and their supervisory colleagues begin the process of developing new three-dimensional classroom-based assessments, they will need assurances from all levels of the educational hierarchy that their efforts are valued and that the work they are doing is complex, demanding, and important.

In their classrooms and schools, teachers have many options for assessment; formative, interim, summative. During the first year of their transitional work, a reasonable expectation would be for teachers to each develop/select and administer at least two phenomena-based tasks encompassing one or more DCIs supported with at least one CCC and one SEP. At the teacher’s discretion, the phenomena-based task could be used as a component of a benchmark assessment throughout the year or used on a formative, interim, or summative assessment of an instructional storyline. This local expectation would be encouraged by a decision from NYSED to develop transitional state science assessments as described earlier in the paper. This said, teachers may still need further assurance that they should continue to balance phenomena-based and other more familiar forms of assessments within the daily/weekly course of instruction as this process of reforming assessment evolves.

What kinds of professional development opportunities regarding the new classroom-based science assessments do science teachers need?

The Strategic Plan for Science and the NYS Science Standards Implementation Roadmap recognize that professional development is needed in all three phases of the Plan. In Phase I, the emphasis is on building teachers’ awareness of the new NYSP-12SLS and developing their capacity to make the instructional shifts needed to help students achieve performance expectations that are three-dimensional. As noted earlier, achieving each performance expectation requires the integration and use of disciplinary core ideas, science and engineering practices, and cross cutting concepts to interpret, analyze, and ultimately explain a science phenomenon or solve a real-world problem.

So in order to promote and support this kind of student learning, teachers will need to learn and implement new teaching strategies that can complement and enrich the student-centered, “minds-on”, and investigative strategies they already use in their classrooms and laboratories. In doing so, it is good to remember that both formative and summative assessment strategies, tools, and instruments can and should be seamlessly embedded in lesson and unit plans to provide teachers and students with opportunities to assess their growth in understanding and their capacity to develop and support their explanations of phenomena and solutions to problems.

Fortunately, there are new and promising instructional strategies being developed and refined that can support the three-dimensional teaching and learning required by the NYSP-12SLS. Two of them are labeled the Question Formulation Technique (QFT) and the Claim, Evidence, Reasoning (CER) approach. These instructional strategies can help guide students in applying the Science and Engineering Practices to phenomena. After observing a phenomenon, QFT engages students by having them ask questions. QFT asks students in groups to brainstorm questions with all participating, convert open/closed questions, identify priority questions and think about the next steps. The priority questions can become the driving or essential lessons as students figure out the phenomenon. CER asks students to make claims (answers to questions) using evidence (these can be a few science-based artifacts) and reasoning (this connects the claims and evidence and includes the use of science principles). The value of the CER approach is that it pulls together the SEPs, DCIs, and CCCs all in one integrated package. Getting students to incorporate the science in their reasoning takes the claims (answers) beyond the evidence collected and requires deeper thinking with greater justification for the claims.

Beyond the use of QFT and CER, there are other strategies that can help students achieve the performance expectations in the NYSP-12SLS, including modeling, problem-based learning, and the use of the 5E Learning Cycle. The key is to ensure that these strategies contribute partially or fully to three-dimensional learning and to know when they can be effectively employed.

In order to provide professional development to teachers and other educators that builds upon prior knowledge and provides a foundation for understanding what phenomena-based assessments are and why they are necessary, it is recommended that the topics for emphasis in professional development venues be addressed in the order listed below.

Recommended sequence of topics to be emphasized in professional development venues

1. Fundamentals of three-dimensional learning to unpack the NYSP-12SLS

2. Research-based instructional strategies such as QFT, modeling, and CER and training related to how such strategies support three-dimensional teaching and learning

3. The development and/or evaluation and use of multidimensional, phenomena-based classroom tasks and related assessments, for both formative and summative evaluative purposes

Leadership provided by State, regional, and local education authorities and the provision of systemic and coordinated professional development and the necessary resources will go far in supporting classroom teachers and science supervisors as they transition to the use of new assessments aligned with the NYSP-12SLS. However, the Strategic Plan for Science recognizes that many other stakeholders in science education reform must understand the reasons for the reform and how science classrooms, instruction and assessments will change over the next several years if they are to support these transitions. There is strong and recent evidence that administrators, parents, school board members, community members, and others can effectively resist what are presented as educational reforms if they are not engaged early in the cycle and the limitations of the status quo are not clearly delineated and contrasted with the opportunities afforded by the proposed reform. To this end, the following recommendations are provided to broaden the base of support for a new science assessment system that includes classroom-based components aligned with the new standards and consistent with new state science assessments.

Recommendations to broaden the support for local reform in science assessment:

1. Provide venues to parents, administrators, and community members to increase their understanding of:

    - three-dimensional learning and what three-dimensional instruction and assessment looks like in the classroom
    - the changes needed in districts and schools in order to transition from elementary, intermediate, and commencement level assessments based upon the 1996 MST standards to three-dimensional, phenomena-based assessments based on the NYSP-12SLS
    - the implications of the needed changes for local program development, resource allocations, grade determination, academic advising, and other school functions related to student learning.

2. Provide professional development opportunities to school supervisors and administrators to enable them to evaluate teachers using criteria referenced to three-dimensional teaching, learning, and assessment.

Summary

The Statewide Strategic Plan for Science provides a mission and vision that, if achieved, will provide all students in New York State with the opportunity to obtain a P-12 science education that will be recognized both nationally and internationally as a program of unqualified excellence. Our state can enhance its position as a leader in science education reform if stakeholders have the collective will and determination to marshal the necessary human and fiscal resources and work together to achieve the goals and objectives of the Strategic Plan. Significant, pervasive, and lasting improvements in science curriculum, professional development, materials support and resources, and administrative and community support must occur if the NYSP-12SLS are to be successfully implemented. However, the key to cementing such reforms historically and now is the capacity of new state and local science assessments to measure accurately and with fidelity the improvements to student learning that result from these changes. The New York State has one of the largest economies in the world and few other states or nations possess the concentration of talent, innovation, capability and social concern found within the borders of this state. Surely, there is no reason for us to falter in our plans to ensure that all children receive the science education that they deserve and need in our advanced, technological, and globally-connected society.

Future scientists, others whose careers depend upon strong grounding in science, and those who pursue other paths are not inspired by how many facts or ideas they have learned about the natural world. They are inspired by confronting engaging phenomena or problems that require them to apply their knowledge and reasoning to make sense of what they observe or to solve the problem at hand. Assessments, whether used in the service of classroom teaching and learning or in monitoring the progress of students across the state, can convey what we believe about and aspire for our children; that they have minds of their own and should be empowered to use them.

APPENDIX A

PARTICIPANTS AT THE SEVENTEENTH STATEWIDE SCIENCE EDUCATION SUMMIT

Participant	Professional Organization
Alan Ascher	Science Council of New York City
Joyce Barry	Long Island STEM Education Leadership Association
Keith Bogert	Capital Area Science Supervisors Association
Lisa Brosnick	Science Teachers Association of New York State
Matthew Christiansen	Science Teachers Association of New York State
Glen Cochrane	Science Teachers Association of New York State
John Cunningham	Science Council of New York City
Karin Cyganovich	Science Teachers Association of New York State
Sandra George	Science Teachers Association of New York State
Michelle Hinchliffe	Science Teachers Association of New York State
Kenneth Huff	Science Teachers Association of New York State
David Jacob	BOCES Science Representatives Network
Elaine Jetty	NYS Biology and Chemistry Professional Development Network
Emily Kang	Invited Participant
Mary Lobello	Science Teachers Association of New York State
Mary Loesing	Long Island STEM Education Leadership Association
Edel Maeder	Rochester Area Science Education Leadership Team
Ellen Mandell	Science Council of New York City
Helen Pashley	Science Teachers Association of New York State
Fred Pidgeon	NYS Science Education Consortium
Rose Sanders	Science Teachers Association of New York State
Arnie Serotsky	NYS Science Education Consortium
Anchala Sobrin	Invited Participant
Mary Thomas	BOCES Science Representatives Network
Bruce Tulloch	NYS Science Education Consortium
Samuel Washington	Science Teachers Association of New York State
Joseph Zawicki	New York State Science Education Leadership Association

References

Achieve. (2017). *Instructional materials + teacher support = student success.* (Available at: https://www.achieve.org/im)

Achieve. (2018a). *Criteria for procuring and evaluating high-quality and aligned summative science assessments.*

Achieve. (2018b). *Transforming science assessment: challenges and recommendations for states.*

Anderson, P. (2017). *The wonder of science.*

Committee on Conceptual Framework for the New K-12 Science Education Standards, D. o. B., Sciences, S., & Education, . B. o. S. E., National Research Council. (2012). *A framework for k-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

Council of Chief State School Science Officers. (2015). *Science assessment item collaborative assessment framework for the next generation science standards.* (Available at: https://www.ccsso.org/sites/default/files/2017-12/SAICAssessmentFramework$_F$*INAL.pdf*)

National Research Council. (2012). *A framework for k-12 science education: Practices, cross-cutting concepts, and core ideas*. Washington, D.C.: National Academies Press.

National Science Teachers Association. (2016). *Ngss@nsta.*

New York State Education Department. (2017). *New york state p-12 science learning standards.* On-line. (Available at: http://www.p12.nysed.gov/ciai/mst/sci/documents/p-12-science-learning-standards.pdf)

New York State Education Department. (2018). *New york state strategic plan for science.* On-line. (Available at: http://www.p12.nysed.gov/ciai/mst/sci/documents/Final-Statewide-Strategic-Plan-for-ScienceRev.pdf)

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, D.C.: National Academies Press.

NYS Science Education Consortium. (2018). *Transition resources for the implementation of the nys p-12sls.*

Ormiston, K. D. (1987). *History of regents examinations, 1865-1987.* Albany: The State Education Department, Bureau of Elementary and Secondary Testing Programs, On-line. (Available at: http://www.p12.nysed.gov/assessment/hsgen/archive/rehistory.htm)

Pellegrino, J, Wilson, R., Koenig, J. and Beatty, A. Editors. (2014). *Developing assessments for the next generation science standards.* (Available at NAP.edu)

SRI. (2018). *Education and learning.* (Available at: https://www.sri.com/research-development/education-learning/)