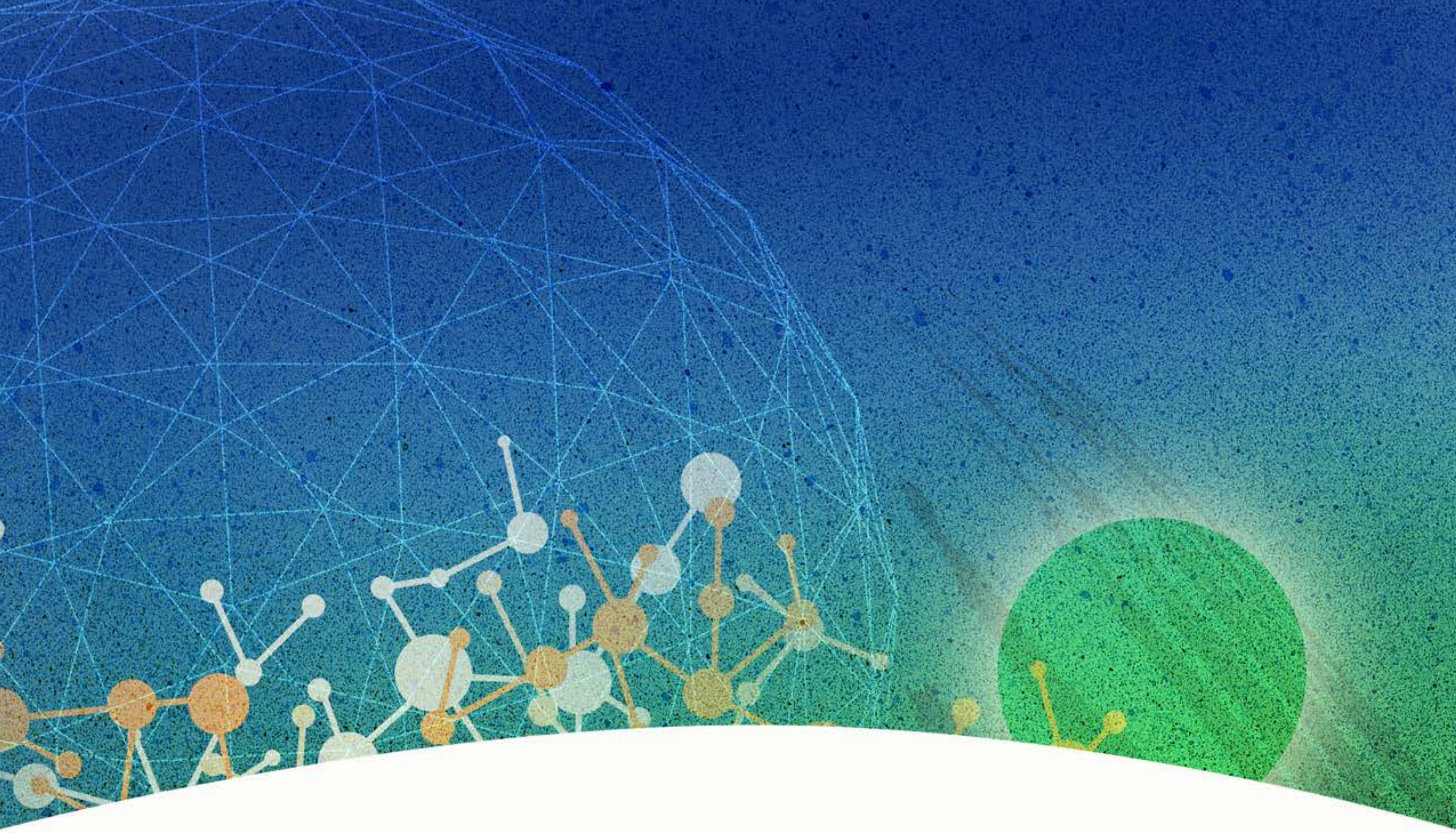


Invitational Research Symposium on Science Assessment

Measurement Challenges and Opportunities

Rodger W. Bybee

September 24-25, 2013

Washington, DC

K-12 Center at ETS

Invitational Research Symposium on Science Assessment

Statement of Purpose

This report summarizes expert presentations and discussions that took place during an Invitational Research Symposium on Next Generation Science Assessments hosted by the Center for K-12 Assessment & Performance Management at ETS on September 24–25, 2013, in Washington, DC.

The primary purpose of the symposium was to share and stimulate expert thinking regarding the design of systems of assessments—formative, interim, and summative—that will support the successful implementation of the Next Generation Science Standards (NGSS). In addition, the symposium generated a shared awareness of challenges between the constructs in the NGSS and our current capacity to assess them so that future research and development can be focused on these areas. This symposium provided a valuable opportunity for teams of state assessment professionals and science supervisors to coordinate their thinking and for assessment experts to gain deeper understandings of the needs of those in the field.

This summary report and the research papers, presentation slides, and videos of discussions at the symposium are available at www.k12center.org/events.html.

Acknowledgments

I wish to acknowledge Nancy A. Doorey and Pascal D. Forgione of the Center for K-12 Assessment and Performance Management at ETS for their contributions, recommendations, and support in the preparation of this report. In addition, I acknowledge the assistance of Charlene Tucker and Kim Fryer on this report.

Table of Contents

List of Figures

Invitational Research Symposium on Science Assessment: Measurement Challenges and Opportunities

Rodger W. Bybee[1]
Next Generation Science Standards

In April 2013, the Next Generation Science Standards (NGSS) were released. The new standards for science education have engaged a variety of opportunities, questions, challenges, and actions for the science education community. Central among those questions are ones about assessment and accountability. “Will assessments change?” “How will state and local assessments change?” The brief answer to these questions is yes. The assessments will change. The more complicated questions are the following: How will the assessments change? When will they change? What will assessments based on NGSS look like?

Recognizing these questions and associated concerns about the implementation of NGSS, the Center for K-12 Assessment & Performance Management at Educational Testing Service, along with the Council of Chief State School Officers (CCSSO) and the College Board, organized a symposium to discuss the next generation of science assessments and specifically to address the challenges associated with designing, developing, and implementing those assessments. This report is a summary of the presentations and key recommendations of the symposium.

The report is organized in three parts. Part I discusses standards in science education with particular attention to the development, noteworthy design attributes, and challenges of the NGSS, the principal focus of the symposium.

Part II uses the symposium agenda, as found in Appendix A, as its organizing framework. Discussion follows the sequence of sessions. At this point, I note that respective links to the commissioned papers and slides for this part are provided so the reader of the electronic version will

[1] Rodger W. Bybee is the chair of the Science Forum and Science Expert Group for the Next Generation Science Standards and was a questionnaire panel consultant for Program for International Student Assessment (PISA) 2006. Until 2007, Dr. Bybee was the executive director of the Biological Sciences Curriculum Study (BSCS), a nonprofit organization in Colorado Springs, CO, that develops curriculum materials, provides professional development for the science education community, and conducts research and evaluation on curriculum reform. Prior to joining BSCS, he was the executive director of the National Research Council’s (NRC) Center for Science, Mathematics, and Engineering Education (CSMEE), in Washington, DC. Between 1985 and 1995, he participated in the development of the National Science Education Standards, and between 1992 and 1995 he also chaired the content working group of that NRC project.

easily be able to go deeper into the substance of each session. Part II ends with a summary of the key learnings and major findings of the Science Assessment Symposium.

Part III focuses on moving beyond the challenges. What is ahead in the critical policy and practice work at the state and national levels in implementing the NGSS and assessments? This part draws upon a panel discussion (Session 7) and the closing remarks (Session 8) for the symposium. The final component of Part III is a concluding discussion of the major themes of the symposium's presenters.

Part I

Standards for Science Education

Although the history of science education reveals numerous committee reports, yearbooks, and other publications that served as standards, one can reasonably argue that it was the late 1980s and early 1990s that brought the term *standards* into the science education community. *Science for All Americans* (Rutherford & Algren, 1989) provided the basis for *Benchmarks for Science Literacy* (AAAS, 1993) and prepared the country for *National Science Education Standards* (NRC, 1996).

The fundamental idea behind science education standards is to describe clear, consistent, and comprehensive skills, knowledge, and abilities that would serve as the basis to reform essential components of the science education system—teacher education, programs for school science, teaching practices, and assessments at local, state, and national levels. The NGSS were designed to reflect advances in fields of science over the last decade, as well as incorporate the growing understandings concerning learning and teaching in science. As such, these standards present the most recent challenges and opportunities to improve science education in the United States.

Development of the Next Generation Science Standards (NGSS)

Development of NGSS began in 2010 and consisted of a two-step process. The first phase of development was led by the National Academy of Sciences. The National Research Council (NRC), the operational arm of the National Academy of Sciences, developed *A Framework for K-12 Science Education* (NRC, 2012). The framework provided a solid foundation based on science learning research that describes what science all K-12 students should know and be able to do. The NRC convened a committee of 18 nationally and internationally recognized experts: practicing scientists, including two Nobel laureates; cognitive scientists; science education researchers; and science education standards and policy experts.

A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas (NRC, 2012) has three parts. The first part presents a vision for science education which includes the guiding assumptions and organization. Part two provides the content for science and engineering education.

Finally, part three addresses the means to realize the vision by addressing the integration of content, implementation, equity, and guidance for the NGSS.

There are two aspects of the Framework that warrant special emphasis in this report. The first is that the Framework calls for science education to reflect science as practiced in the real world, in contrast to a set of facts, ideas, and skills to be mastered.

This foundational orientation in the Framework is operationalized as the second arguably unique attribute of the Framework, and one that was central to the discussions at the symposium: the Framework calls for all students to be actively engaged in science and engineering activities that integrate the three dimensions of the standards—science and engineering practices, crosscutting concepts, and core ideas in science disciplines. In other words, the Framework, and subsequently the NGSS, require the development of instructional and assessment tasks that go beyond the command of facts or analysis of data, requiring students to engage in well-structured investigations; develop, test and refine models; and communicate research findings.

The NRC Framework provided guidance for developing standards to ensure fidelity to the Framework. For this discussion, the following summarizes the NRC recommendations for standards development.

The standards should:

- Set rigorous goals for all students.
- Be scientifically accurate.
- Be limited in number.
- Emphasize all three dimensions.
- Include performance expectations that integrate the three dimensions.
- Be informed by research on learning and teaching.
- Meet the diverse needs of students and states.
- Have potential for a coherent progression across grades and within grades.
- Be explicit about resources, time, and teacher expertise.
- Align with other K-12 subjects, especially the Common Core State Standards (CCSS) in English Language Arts/Literacy and Mathematics.
- Take into account diversity and equity. (NRC, 2012, pp. 297–307)

The second phase was a state-led effort managed by Achieve to create the NGSS. Twenty-six states pledged their commitment to give serious consideration to adopting NGSS. Each state also committed to create a broad-based team, comprising K-12 representatives (including science teachers), scientists and engineers from the business community, employers, and education leaders. These state teams provided feedback on drafts of the standards and were critical partners. The teams also provided updates for key constituents within their states.

The final NGSS document was developed through the collaborative effort of these 26 lead states in cooperation with stakeholders in science, science education, higher education, and business and industry. Draft standards underwent multiple reviews, including two publicly released drafts, which provided all interested and involved individuals and groups with an opportunity to inform the proposed content and practices, as well as the organization of the NGSS. The result of this process is a set of rigorous, high-quality K-12 science standards that have passed a review for fidelity by the NRC. NRC reviewers used the vision and content of the Framework as a baseline to evaluate the consistency of the draft NGSS. The review panel included members of the original NRC Framework committee and other experts who were familiar with the Framework. The National Academies Press (NAP) has published the NGSS.

The Opportunities and Challenges of Accommodating Conceptual Shifts in the Next Generation Science Standards (NGSS)

The NGSS present an opportunity to improve curriculum, teacher development, assessment and accountability, and ultimately student achievement in science. In order to bring this opportunity to reality, the science education community must address conceptual shifts in the NGSS. While the conceptual shifts refer to NGSS, in reality, the shifts have direct implications for curriculum, instruction, teacher development, and assessment. The following description of conceptual shifts is based on *A Framework for K-12 Science Education* (NRC, 2012) and Appendix A of the NGSS (NGSS Lead States, 2013).

Interconnect Three Dimensions

NGSS standards have three dimensions: (a) disciplinary core ideas, (b) scientific and engineering practices, and (c) crosscutting concepts. Most state and district standards, and subsequent programs and assessments, express these dimensions as separate entities. The integration of rigorous content and application reflects how science is practiced and implies a science education that provides deeper experiences and understanding of concepts and practices.

Recognize Learning Progressions

Science concepts in NGSS build coherently from kindergarten through grade 12. The NGSS presents a progression, from grade-band to grade-band, that builds understanding and abilities throughout a student's K-12 science education.

Include Engineering

The NGSS includes both science and engineering. Science and engineering practices and crosscutting concepts are designed as an integral component of the standards.

Address Nature of Science

The nature of science was not directly addressed in *A Framework for K-12 Science Education*; however, feedback from the states and the public clearly and strongly indicated a need to include an understanding of the scientific enterprise. So, the nature of science is included in NGSS. The basic understandings about the nature of science are expressed in the categories and statements in Figures 1 and 2.

- Scientific investigations use a variety of methods.
- Scientific knowledge is based on empirical evidence.
- Scientific knowledge is open to revision in light of new evidence.
- Scientific models, laws, mechanisms, and theories explain natural phenomena.

Figure 1. Nature of science categories most closely associated with practices.

- Science is a way of knowing.
- Scientific knowledge assumes an order and consistency in natural systems.
- Science is a human endeavor.
- Science addresses questions about the natural and material world.

Figure 2. Nature of science categories most closely associated with crosscutting concepts.

The NGSS appendices elaborate these categories in the form of learning progressions for grade bands K-2, 3-5, 6-8, and high school. The content for the nature of science is stated as part of the practices and crosscutting concepts and presents a conceptual shift of instructional materials, teaching, and assessment of learning outcomes.

CoordinateScience With Common Core State Standards (CCSS) for English Language Arts (ELA) and Mathematics

Implementing the NGSS presents an opportunity for science to be an integral part of students' comprehensive education. The NGSS makes connections between the individual science standards and the components of CCSS, thereby leveraging the opportunity to support student learning of the ELA/literacy and mathematics expectations within the context of science instruction. Work at Stanford University is already underway to address the advantages to both NGSS and CCSS.

One clear implication is the need for integrated programs, including assessments. Further, the incredible complexity of the NGSS and conceptual shifts present significant challenges for those with responsibilities for assessment.

Part II

A Symposium on Science Assessment

As of September 2013, seven states had adopted the NGSS, and an even greater number had the standards under active review and consideration. In addition to the requirement within the Elementary and Secondary Education Act (ESEA) for science assessments at three grade bands, many states have within law or regulation requirements for summative science assessments at additional grades, and often with high stakes for students at the high school level. In addition, a growing number of states and districts are building the use of summative assessment results in science into teacher evaluations systems. How can states with such requirements navigate the transition to the NGSS? What are the key challenges, most promising work underway, and best starting points to support states as they begin this work? Teams of science and assessment leaders from 35 state education agencies gathered to hear from architects of the NGSS and leaders in science education and assessment at the Invitational Research Symposium on Science Assessment.

The following sections provide summaries of the main themes of the symposium from the respective sessions. Background papers and slides can be found on the website at:

www.k12center.org/events/research_meetings/science_assessment.html.

Session 1

The Next Generation Science Standards (NGSS):

What Is the Vision of Scientific Competency and Science Instruction?

In this session, leaders of the NRC committee that developed the Framework and the group of states that developed the NGSS provided an overview of the vision of instruction underlying these standards and the challenges that educators and the measurement community will confront when building instructional and assessment resources.

A Focus on Competence

Jim Pellegrino, University of Illinois at Chicago, began the session with a review of the standards and framework as guides for the reform of K-12 science education. He located the NRC Framework and NGSS central to curriculum, instruction, and assessment in contemporary science education.

One of the themes of Pellegrino's introduction was that the definition of competence is a different way of thinking about learning outcomes. He made the point that the NRC Framework and NGSS have proposed descriptions of student competence as being the intersection of knowledge involving: (a) important science practices, (b) disciplinary core ideas, and (c) crosscutting science concepts. The performance expectations within the NGSS represent the intersection of these three dimensions. Pellegrino further elaborated: “Competence should be viewed as something that develops over time and increases in sophistication and power as the product of coherent curriculum and instruction.”

To this point, Pellegrino made the connection to assessments. Very importantly, he summarized three grand challenges for the assessment community. Those challenges are stated here in their entirety (see Figure 3).

Grand Challenges

Design valid and reliable assessments reflecting the integration of practices, cross-cutting concepts, and core ideas in science. The performance expectations of the Next Generation Science Standards (NGSS) pose significant challenges for assessment design. Considerable research and development will be needed to create and evaluate assessment tasks and situations that can provide adequate evidence of the proficiencies implied in the NGSS. This research must be carried out in instructional settings where students have had an adequate opportunity to construct the integrated knowledge envisioned by the National Research Council Framework and the NGSS.

Use assessment results to establish an empirical evidence base regarding progressions in science proficiency across K-12. Much of what is assumed in the NGSS regarding learning progressions needs to be validated through empirical research. This validation requires assessment tasks and situations that can be used across multiple age and grade bands so that we can determine how proficiency changes over time with appropriate instruction. The empirical results can then be used to support the design of more effective curriculum materials and instructional practices.

Build and test tools and information systems that help teachers effectively use assessments to promote learning in the classroom. For teachers to effectively implement assessment as part of their pedagogy, they need tools for presenting tasks and collecting and scoring student performance. They also need smart systems that provide actionable information about the meaning and implications of student performance relative to instruction and student learning. Such systems will need to be designed in collaboration with learning scientists and teachers to insure their validity, usability, and utility.

Figure 3. Science assessment: grand challenges.

Pellegrino concluded with two important thoughts. The first is a variation on Voltaire’s admonition that, “The perfect is the enemy of the good.” In this context, he cautioned against seeking perfection before acting on behalf of students and teachers. His second caution that assessment should not be the driver of educational practice was expressed as—“Assessment should not be the tail wagging the science education dog.”

The Vision

The second presentation was by Helen Quinn of Stanford University (emeritus) who chaired the NRC committee that prepared *A Framework for K-12 Science Education* (NRC, 2012). There were two important themes of Quinn’s narrative. First, the NGSS introduces new perspectives for K-12 science education and subsequently assessment. Changing perspectives of science education was a consistent and important theme across speakers at the symposium. Quinn summarized the changes as a vision for science education that:

- Reflects science as practiced by scientists—practices, concepts, and ideas intertwined.
- Presents ideas that build coherently across K-12.
- Focuses on deeper understanding and application.
- Integrates science and engineering.
- Prepares *all* students for college, career, and citizenship.
- Aligns with the CCSS.

Quinn's second theme was the importance of the *Framework* as an enduring contribution to science education because it provides a holistic vision that will continue to influence the fundamental components of curriculum, instruction, and assessment.

There are several major decisions that rest on the shoulders of those with the responsibility of implementing NGSS. How can new approaches to assessment be incorporated into teachers' development? How will curriculum embed formative assessment opportunities? How can changes in instruction and assessment be coordinated over time?

The Reality

The team of Brian Reiser (Northwestern University), Joseph Krajcik (Michigan State University), and Lisa Brody (Park View School) made the vision for science teaching based on NGSS a reality. They made the translation from the *Framework* to NGSS to K-12 classrooms real through exploring the question, "What does it mean to learn that matter is made of molecules?" Designing learning activities to help students build explanatory models was at the heart of the presentation. The presentation became real with the video of Brody teaching a sixth grade unit on the particle nature of matter. Throughout the video segments, students collectively proposed, challenged, and revised an explanatory model, thus demonstrating what the integration of NGSS dimensions actually look like in a classroom. Indeed, the video showed multiple practices working together.

In summary, the presentation and video provided answers to the question, "How can you build students' understanding of disciplinary core ideas through practices?"

The answer provided by Reiser, Krajcik, and Brody is summarized in Figure 4.

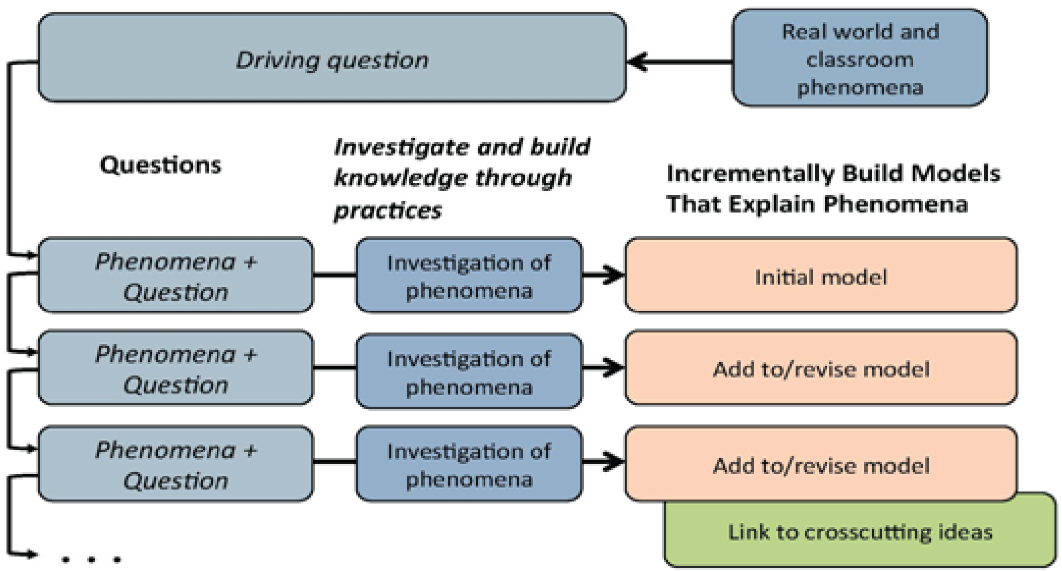


Figure 4. Building disciplinary core ideas through practices.

The clear and explicit presentation left the participants with the question, "How does this vision of classroom instruction translate to assessment?"

Next Generation Science Standards (NGSS) Perspectives

Stephen Pruitt, Achieve, Inc., opened with two important perspectives. First, he stressed that the NGSS is a set of guidelines developed for states, by states. Second, he underscored the fact that NGSS was based on the Framework and that the NGSS had passed a fidelity review by the NRC.

Pruitt closed with the following key messages for states:

- Effectively implement the CCSS.
- Identify the role that science plays in the overall education plan.
- Develop a thoughtful and deliberate implementation plan that supports the overall education plan.
- Be patient and take adequate time to implement assessments for NGSS.

The Work and Responsibility

Juan-Carlos Aguilar, the state science supervisor of Georgia, delivered a succinct and salient message. He noted that the hardest work is ahead of the entire science education community, whose responsibility it is to attend to the professional development of teachers, the adaptation and/or adoption of instructional materials, and the uses of appropriate assessments at the national, state, and local levels.

To summarize, speakers in this session addressed questions about the vision of scientific competency and science instruction. Based on NGSS, scientific competency is a different way of thinking for students and teachers. This way of thinking involves understanding and applying three dimensions: scientific and engineering practices, disciplinary core ideas of science, and crosscutting concepts of science. By extension, this new way of thinking about K-12 science education has direct implications for: (a) curriculum and instruction, (b) teachers' professional development, and (c) assessment. Further, it is especially important that these three educational components are coherent as they are implemented in educational systems.

The vision of instruction designed to provide students opportunities to learn this new perspective was made real by the video of Lisa Brody and her 6th graders and subsequent discussions.

The concluding presentations by Pruitt and Aguilar underscored the need to begin work on instructional materials, professional development, and assessments. There were repeated cautions that the science education community needs to be patient and take the time that is needed to do it right. However, given the importance of student proficiency in science for citizenship, college and careers, that patience also needs to be tempered by Jim Pellegrino's caution: "Do not let the perfect be the enemy of excellence."

Session 2

Measuring Across Multiple Dimensions: How Can We Measure the Ability to Engage in Science Practices in the Context of Rich Content?

Among the most significant new features of the NGSS is a call for the integration—in instruction as well as assessment—across three dimensions of content: science and engineering practices, disciplinary core ideas, and crosscutting concepts. This is a new level of complexity for the assessment field. This session included an overview of the discussions to date of the NRC Committee on the Assessment of Science Proficiency and then turned to a National Science Foundation (NSF) funded project on the development of science tasks designed to assess across the three dimensions.

A National Research Council (NRC) Report on Assessment

Mark Wilson, a co-chair of the NRC committee from the University of California, Berkeley, described the committee's charge. Because of the importance of their work, the committee's charge is presented here.

- Make recommendations for strategies for developing assessments that validly measure student proficiency in science as laid out in the new K-12 science education framework.

- Review recent and current, ongoing work in science assessment to determine which aspects of the necessary assessment system for the Framework's vision can be assessed with available techniques and what additional research and development is required to create an overall assessment system for science education in K-12.
- Make recommendations to state and national policy makers, research organizations, assessment developers, and study sponsors about the steps needed to develop valid, reliable, and fair assessments for the Framework's vision of science education.

Wilson continued with a summary of NGSS-based challenges that the committee and assessment community will have to address.

- To develop assessment tasks that integrate the three dimensions.
- To develop tasks that can assess where a student can be placed along a sequence of progressively more complex understandings of a given core idea, and successively more sophisticated applications of practices and crosscutting concepts.
- To develop assessment tasks that measure the connections between the different strands of the disciplinary core ideas (e.g., using understandings about chemical interactions from physical science to explain phenomena in biological contexts).

The NRC committee is exploring a systems approach to assessment of the NGSS. Essential components of the assessment system could include classroom-based assessments designed to provide feedback on activities and instruction, assessments to monitor student learning, and assessments to track factors that influence learning outcomes (e.g., opportunities to learn, appropriate resources, teaching strategies). Wilson elaborated on the three forms of assessment.

This presentation concluded with a summary of assessment challenges, including:

- Developing rich assessment tasks that evaluate the intended practices, disciplinary content, and crosscutting concepts.
- Having the platforms and resources to administer these tasks.
- Scaling the tasks in the presence of multidimensionality and linkage across dimensions.
- Scoring the tasks.
- Developing informative, useful reports and moderation strategies of test results.
- Making use of information from classroom assessments for accountability purposes.
- Making use of process indicators.
- Assembling the assessment components into a coherent system.

Wilson concluded with a concern. He noted that he supported this ETS symposium and its emphasis on assessment. However, he further noted that the challenges associated with NGSS also require attention to curriculum, instruction, and the professional development of science teachers.

The NRC committee's report is expected to be released in late 2013 and will be available on the NRC website.

Designing Assessments

To determine what these assessment challenges look like in practice, the next presentation addressed the challenge of "Designing NGSS Assessments to Evaluate the Efficacy of Curricular Interventions." This team, Angela DeBarger and Christopher Harris, SRI International, and William Penuel, University of Colorado, Boulder, were commissioned by the K-12 Center to prepare a paper describing the lessons learned to date in their NSF-funded research project to develop NGSS-aligned tasks. In their paper, they state that:

> A key initial goal for all assessment design is to design an assessment that allows for accurate judgments about what individual students know and can do. Assessments of student proficiency in meeting the performance expectations of the Next Generation Science Standards should generate evidence for claims about students' capacity to demonstrate understanding of core ideas and fundamental cross-cutting concepts through engaging in science practices. (DeBarger, Penuel, & Harris, 2013, p. 5)

The team used an evidence-centered design (ECD) approach for developing assessment tasks that address the three dimensions of NGSS. They shared examples of a task and rubric and concluded with an insightful discussion of lessons they learned.

One of the lessons addressed in their paper is that the design of tasks requires collaboration of parties with different expertise and experience.

> To develop science assessments, ECD involves science educators, science content specialists, assessment designers and psychometricians. Articulating the claims about what students can know and should be able to do must be guided by science educators and content specialists. Developing appropriate tasks and scoring guidelines requires not only domain knowledge but also experience with working with students for whom tasks are intended so that context and language are appropriate. Identifying which measurement models should be used to determine how student scores provide evidence of claims requires knowledge of statistics and psychometrics. For technology-based assessments, software engineers also may be involved in the earlier domain modeling and conceptual assessment phases so that design specifications are usable at the implementation phase. Importantly, co-designers coordinate their activity, but all interactions among designers do not take place within the full group. (DeBarger et al., 2013, p. 10)

Another lesson discussed in their paper is that an ECD approach to assessment design is an iterative process:

> Responding to these design decisions sometimes requires revisiting earlier decisions that have been made about claims and approaches guiding task design. If these models for assessment

design are to be reusable and sustainable they must be refined as new information is learned either from research about how learning occurs or from the performance of particular items and tasks. (DeBarger et al., 2013, p. 11)

Their presentation focused on an efficacy study they completed of a project-based inquiry science (PBIS) program. This study focused on core ideas in the physical and Earth sciences and the scientific practice of modeling. They had to answer a series of design questions.

How can we

- **design** tasks that integrate disciplinary core ideas and crosscutting concepts?
- **promote coherence** across two disciplinary areas?
- **ensure fairness** for both PBIS and non-PBIS contexts?
- construct **easy-to-administer tests** for a large number of students?
- **facilitate** efficiency and reliability in scoring?

As the presentation progressed, the ECD perspective became clear. The team had to be explicit about design decisions, and they had to refine their decisions. A very important perspective was their summary of building the assessment argument. That perspective is summarized in Figure 5.

• What do we expect students to know and be able to do?	**STUDENT MODEL**
• How do we know it when we see it?	**EVIDENCE MODEL**
• What are task features to elicit desired evidence and how can tasks vary?	**TASK MODEL**

Figure 5. Building the assessment argument.

Based on the framework in Figure 5, the team described focal knowledge, skills, and abilities (Focal KSA) that students were expected to know and be able to do, as a result of instruction. All Focal KSAs were based on NGSS and dimensions of performance expectations. They described the potential observations related to Focal KSAs, and they described the characteristic features and variable features of assessment tasks. At this point, the team presented an example that demonstrated very well an assessment task. The assessment task and different features are displayed in Figures 6 and 7.

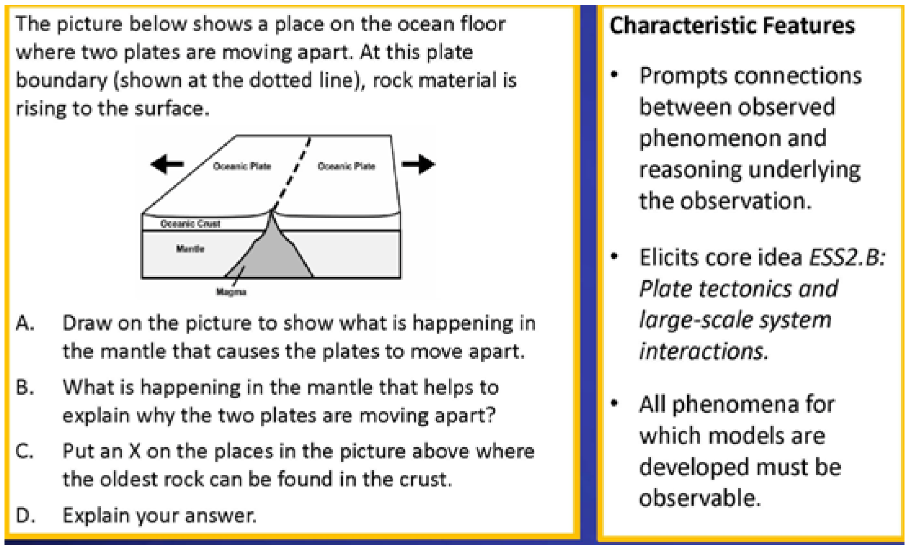


***Figure 6. Earth science task: Diverging plate boundary*, characteristic features.**

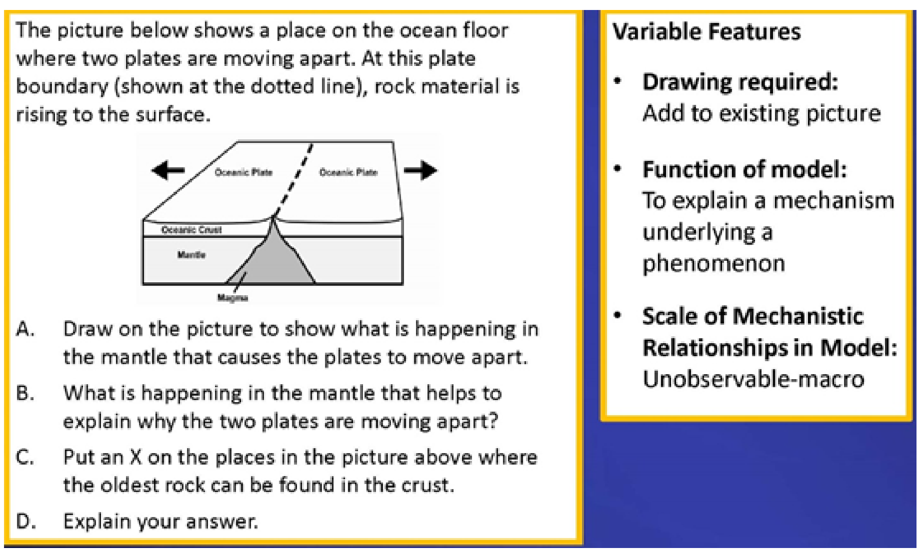


***Figure 7. Earth science task: Diverging plate boundary*, variable features.**

The presentation continued with descriptions of the rubrics and scoring of the task and concluded with implications. The implications included (a) the need to recognize and design for intended purposes of the assessment and (b) that the investment in ECD promoted coherence in the assessment system.

In conclusion, the two presentations in this session did a superb job of translating the NGSS from policies to actual assessments. The presentations provided existence proofs that it is possible to design assessments based on NGSS. Both presentations also included some challenges. Among the important observations:

- There are many complex challenges associated with designing the types of rich tasks that will be needed to address the performance expectations of the NGSS.
- These challenges extend beyond assessment design to the arenas of curriculum, instruction, and professional development.
- An ECD approach holds promise for clarifying and guiding the design of the necessary rich assessment tasks.
- Assessment design will face administration constraints, including those of cost, time, and technology. The assessment design process will require the collaboration of science educators, science content specialists, assessment designers, and psychometricians.

Session 3

What Can Be Learned From Current Large-Scale Assessments, and What Are the Implications for Assessing the Next Generation Science Standards (NGSS)?

In recent years, the frameworks for several large-scale science assessments have been updated to reflect advances in the sciences. As a result, new assessments have either been developed or are under development. How can assessing the NGSS benefit from this work? This session provided an overview of the areas of overlap in the science frameworks for National Assessment of Educational Progress (NAEP), Program for International Student Assessment (PISA) 2015, AP® Physics, and one of the New Zealand National Science Assessments. Selected tasks from these assessments were shared to illuminate how important hard-to-measure constructs within the NGSS are being assessed. The context for each assessment—its proximity to curricula and the consequences associated with the results—was discussed.

Jonathan Osborne, Stanford University, made the following points as he introduced the session. First, the examples for current large-scale assessments show that we are not starting from scratch. Second, it is important to go beyond content, that is, the disciplinary core ideas of science. Finally, we must assess a set of competencies.

An Overview of Programs

Alicia Alonzo, Michigan State University, prepared for the symposium a very thorough paper on the session's topic. While she summarized key points from her paper in her presentation for the symposium participants, a much deeper discussion can be found in her paper on the K-12 Center's website.

Because the session was on current large-scale assessments (e.g., NAEP, PISA, AP science), Alonzo presented summaries of her analysis of overlaps and gaps (see Figure 8) and examples of new assessment tasks (see Figure 9).

- Where is there strong overlap?
  - Middle of scientific inquiry process (SP2-6)
  - Outcome of engineering design (EP6)
- Where are there gaps?
  - Broad view of modeling (SP2) and argumentation (SP7)
  - Oral components of practices
  - Reflection and revision

Figure 8. Large-scale assessment frameworks and the Next Generation Science Standards (NGSS).

- NAEP Science
  - Hands-On Task
- NAEP TEL (Technology and Engineering Literacy)
  - Computer-Based Engineering Task
- PISA
  - Extended Written Response
- AP® PHYSICS
  - Extended Written Response

Figure 9. New assessment tasks and the Next Generation Science Standards (NGSS).

Along with her summary, Alonzo pointed out that many components of current NGSS practices are reflected in sample assessment items and that computer-based assessments hold great promise for evaluating science practices. She also pointed out remaining challenges—constraints of on-demand assessments and the limitations of our current knowledge base.

National Assessment of Educational Progress (NAEP)

The session continued with a presentation on (NAEP) by Peggy Carr of the National Center for Education Statistics. The theme of the talk was NAEP Innovations in Action. Carr described innovations of item types (e.g., performance-based tasks and interactive tasks) and evidence types (e.g., automatically captured action data). Since 1996, NAEP has included hands-on tasks (HOTs). In 2009, NAEP incorporated tasks that required students to solve science problems in a computerized environment (e.g., a simulated natural or laboratory setting). She demonstrated an item in this format.

Carr also shared an NAEP assessment developed for 2014 on Technology and Engineering Literacy (TEL). An example from the TEL assessment was used to demonstrate how students went about the process of problem solving, with the results indicating the degree to which students are efficient and systematic in solving a problem.

The contemporary NAEP will assess technology and engineering literacy in 2014. This is one positive example that applies to NGSS. Also, the use of computer-based hands-on performance tasks is a second constructive example. In addition, the evaluation of a student's efficiency and systematic approach to science and engineering practices holds promise for assessing the NGSS.

Program for International Student Assessment (PISA)

PISA is an international test of scientific literacy. Eric Steinhauer and Janet Koster van Groos, both from the team at Educational Testing Service that developed the PISA tasks, summarized key aspects of PISA for the symposium. Although scientific literacy—the ability to engage with science-related issues, and with the ideas of science, as a reflective citizen—is not in the foreground of NGSS, PISA still provides constructive examples for the assessment of NGSS. First, the PISA items require the application of scientific competencies and knowledge *within contexts*, some of which are static (e.g., written passages, tables, and illustrations), and some of which are more dynamic (e.g., animations and interactive simulations). The contexts for PISA are usually the introductions in units that include two to six independently scored items.

The items in PISA units may be simple selected-response (i.e., multiple-choice), constructed-response, or complex multiple-choice. The complex multiple-choice items require students to create scorable responses by, for example, constructing sentences using several drop-down selections, moving elements around a screen to complete a diagram, or using a simulation to generate data that can be used to support a scientific explanation. This presentation included released items from PISA 2006 and 2009 and a simulation that was a variant of a PISA 2005 task prepared by ETS.

PISA units demonstrate the potential of computer simulation and the use of a context to assess the different dimensions of science and engineering practices, disciplinary core ideas, and crosscutting concepts.

AP Science

What can we learn from the recently revised AP programs? Karen Lionberger, College Board, began with a general discussion of AP science curriculum frameworks and the new emphasis on science practices. AP science curriculum frameworks include big ideas, enduring understandings, essential knowledge, and learning objectives.

- *Big ideas* encompass the core scientific principles, theories, and processes of the discipline that cut across traditional content boundaries and provide students with a broad way of thinking about natural phenomena.
- *Enduring understandings* incorporate the core concepts that students should retain from the learning experience.
- *Essential knowledge* and *science practice(s)* are derived from learning objectives for what students should *know* and be able *to do*.

The AP science practices are, in effect, both learning objectives and instructional strategies. In the AP framework, practices enable students to establish lines of evidence and use them to develop and refine testable explanations and predictions of natural phenomena. The AP science practices are displayed in Figure 10.

- 1.0 The student can use representations and models to communicate scientific phenomena and solve scientific problems.
- 2.0 The student can use mathematics appropriately.
- 3.0 The student can engage in scientific questioning to extend thinking or to guide investigations within the context of the AP course.
- 4.0 The student can plan and implement data collection strategies in relation to a particular scientific question.
- 5.0 The student can perform data analysis and evaluation of evidence.
- 6.0 The student can work with scientific explanations and theories.
- 7.0 The student can connect and relate knowledge across various scales, concepts, and representations in and across domains.

Figure 10. AP science practices.

One should note the high degree of consistency between science and engineering practices in NGSS and the AP science practices. Both, for example, include models, use of mathematics, analysis of data, forming scientific explanations, and communicating scientific phenomena.

The AP learning objectives are similar to, but not the same as, NGSS performance expectations. Recall that the learning objectives pair science content and practices. Here is an example of an AP learning objective: "The student is able to use Newton's third law to make claims and predictions about the action-reaction pairs of forces when two objects interact." Lionberger concluded with examples of assessment items that align to one or more learning objectives.

Before continuing with the next speaker, Alonzo pointed out what participants should have realized at this point. First, large-scale assessment development work is ongoing and can help inform the work at hand to assess across multiple dimensions. Second, some components already exist and may be modified by states and school districts. Finally, changing current perceptions to competency-based assessments holds great promise. Still, there are constraints. The examples tended to focus on individual components of the practices and not full practice or process of scientific inquiry or engineering design. Also, there is difficulty in assessing students' collaborations and the reflective aspects of students' performances.

Insights From New Zealand

Do New Zealand National Assessments provide helpful insights? *Yes* is the response from Jennifer Mackrell of the New Zealand Qualifications Authority (NZQA). Mackrell described the National Certificates of Educational Achievement (NCEA) that are awarded at three levels aligned with, but not restricted to, the final three years of secondary school. Academic standards in New Zealand are designated as either external (to be assessed in standardized national examinations similar to the summative statewide assessments in the United States) or internal (standardized tasks that are completed in the classroom and scored by trained teachers using rubrics, with external moderation). Scores from the external and internal assessments are aggregated for accountability purposes. Mackrell went on to clarify details and provide examples of internal science assessment tasks and their scoring. Within New Zealand, this model has high trust and is viewed as recognizing the professionalism of teachers. It has the added value of helping teachers understand the assessments, the instructional expectations, and the associated accountability.

In her paper, Alonzo pointed out that "compared with teachers in many other countries, New Zealand teachers are assessment experts" (New Zealand Qualification Authority, n.d.e., p. 1, as cited in Alonzo, 2013, p. 73). This does not occur by accident. Teachers in New Zealand receive significant professional development associated with the country's assessment program. In particular, the quality assurance built into the system includes external moderation—making sure that teachers are making consistent internal assessment decisions across the country by providing feedback and professional development.

So, what can be learned from current large-scale assessments? There are several thoughts to take away from this session.

- We are not starting from scratch. Several examples were presented of rich science assessment tasks that may serve as interesting models (e.g., NAEP, PISA, AP physics, New Zealand National Science Assessments).
- The power of technology is likely to play a key role in the development of new complex science assessments, including the automatic capture of action data (see NAEP example).
- There are multiple examples of science assessments currently in place that assess across dimensions of science practices and content, though not exactly in the way that the NGSS are structured.
- Constraints including time and cost will need to be addressed as assessments are developed.
- There are models provided by NAEP and PISA for assessing engineering practices.
- A notable contrast between assessment systems in New Zealand and the United States is the strong role of teachers as developers, administrators, and scorers of assessments. The expertise of teachers has been developed over time through extensive professional development, particularly through the moderation of scoring.

Osborne concluded the session with his concerns about the three A's: (a) assessments may lack *Authenticity* for students; (b) assessments may be *Atomized* by test developers; and (c) teachers may manipulate instruction (and assessments) so they become *Algorithmic*.

Session 4

Breakout Sessions

Two opportunities were provided for participants to attend small breakout presentations/ discussions. A total of nine options were provided. A description of the breakout session content is provided in Appendix B.

Session 5

How Can Formative Assessment Be Used to Support Improved Teaching and Learning in Science?

Some empirical evidence suggests that formative assessment supports improved teaching and learning. Formative assessments provide teachers with information, during the teaching-learning process, about students' developing understanding. As a result, instruction can be adjusted and timely feedback given to the students. This session addressed the question, "What do we know about effective formative assessment in science, and how can this knowledge be used to inform states' implementation plans for the NGSS?"

Perspectives on Formative Assessment

Joan Herman, University of California, Los Angeles, was commissioned to prepare a paper for the symposium, *Formative Assessment for Next Generation Science Standards: A Proposed Model*. Herman opened the session with the CCSSO definition of formative assessment: "… a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes" (Council of Chief State School Officers, 2008).

After this introduction, Herman described a sequence for the formative assessment process. That process summarized the work of Heritage (2010) and generally followed a sequence that included: setting learning goals, obtaining and interpreting evidence of current learning, identifying the gap between the goals and current learning, and taking action to close the gap. Herman pointed out that there is a synthesis of research on this topic (Black & Wiliam, 1998; Hattie & Timperley, 2007; Kingston & Nash, 2011; Pellegrino, Chudowsky, & Glaser, 2001).

At this point, Herman introduced a very important perspective relative to the role of formative assessment in classroom instruction. Formative assessments should be viewed as coherent, coordinated sets of tests that occur in cycles of time that may be short, medium, or long (Wiliam & Thompson, 2007) and on a continuum from informal to formal (Shavelson et al., 2008). This was further explained in the Herman paper (Herman, 2013, p. 9):

> Richard Shavelson and colleagues (2008), in turn, defined a continuum of formative assessment that ranges from informal to formal and varies relative specific pre-planning. The continuum essentially focuses on short and medium cycle tools, running the gamut from unplanned "on the fly" assessments that particularly capitalize on "teachable moments," through planned interactions during the course of instruction that are specifically designed to both support learning and elicit evidence of it, to formal embedded assessments at key juncture points to evaluate what all students have learned. The informal activities are directly embedded in instruction and indecipherable from it, where teachers (and students) may draw inferences, provide feedback on student learning by observing student behavior and interactions during classroom activities, from analyzing students' work, such as class assignments, lab work, science notebook entries and/or homework, or from analyzing and responding to whole class and/or small group discussion questions (see also Bell & Cowie, 2001).

After a brief discussion of research and a model representing the critical elements of formative assessment, Herman presented the challenge for NGSS. The challenge consisted of accommodating to the progressions of fused science knowledge (i.e., the three dimensions) and assessing and adjusting to students' variable knowledge while still progressing toward learning goals. In the commissioned paper, Herman stated the need for “tasks that can accurately locate students on multiple learning trajectories and provide diagnostic feedback that can support their forward movement across the three” (Herman, 2013, p. 21). See Figure 11.

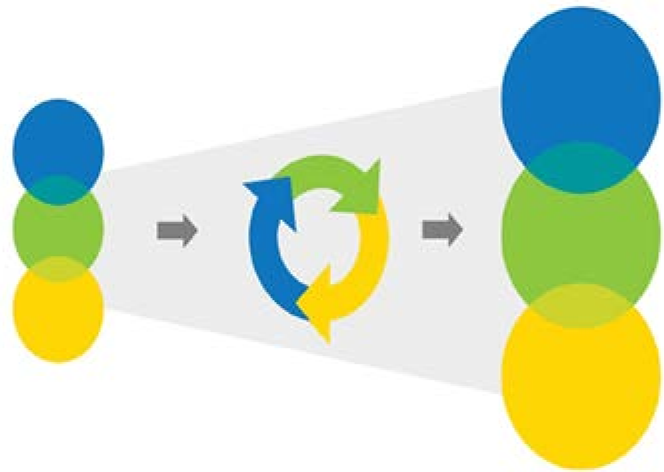


Figure 11. Challenge for Next Generation Science Standards (NGSS): Progressions of fused science knowledge.

Herman concluded with several challenges in moving forward. Those challenges included clarifying learning progressions, increasing the measurement capacity, developing curriculum materials with embedded strategies and tools for formative assessments, developing teachers’ capacity to use formative assessments, and understanding the learning culture.

A Classroom Example

The session continued with presentations of selected examples of formative science assessments. Krajcik began with a video showing a classroom where the teacher prepared students to meet a performance expectation that included the process of building a model to describe that waves are reflected, absorbed, or transmitted through various materials. The discussion by Krajcik and the video grounded the discussion in the realities of the classroom and illustrated many of the features of formative assessment introduced by Herman. Krajcik’s presentation emphasized his dream—engaging students in constructing models throughout the K-12 curriculum; but, the video also demonstrated the integration of the three dimensions of the NGSS in the context of classroom instruction.

A Game-Based Example

The session continued with a presentation on game-based learning and assessment by Malcolm Bauer, Educational Testing Service, and GlassLab. Bauer introduced a very helpful comparison of assessment practices and their analogs in games (see Figure 12).

Formative assessment practices where teachers	Analog where games present
Elicit evidence of student learning minute to minute, day by day, and week to week	Telemetrics and analytics
Identify and share learning expectations with their students	Explicit challenges and goals
Structure opportunities for students to take ownership of their learning	Choices and flexible pacing
Structure opportunities to activate students as instructional resources for one another	Structured contexts for collaboration and sharing—multiplayer platforms
Provide feedback to move learning forward and create a structure for students to act on it	Just in time feedback linked to rewards

Figure 12. **Formative assessment practices and their analogs in games.**

Based on the analogs in Figure 12, Bauer continued with a clarification of learning progressions as presented in NGSS and other constructs, such as English language arts (i.e., CCSS). At this point, Bauer introduced a game involving the reduction of air pollution with the constraint that jobs could not be lost in the process of solving the pollution problem. The game provided a comparison of final jobs and pollution levels, feedback on students' ideas, and a report on individual students' progress toward the learning goals. Bauer concluded with the following points.

Games and gaming culture match well with formative assessment practices. In the classroom, game play is part of a larger system of formative assessment and learning. Finally, game-based assessments have the potential to measure science and engineering practices, crosscutting concepts, and disciplinary core ideas.

A Computer Simulation

Edys Quellmalz and Mark Loveland, WestEd, summarized their team's work on a project titled SimScientists, an assessment system designed to achieve seven goals.

- Create simulation environments that represent dynamic science system phenomena *in action*.
- Take advantage of technology to represent causal and temporal relationships, patterns, and changes in scale.
- Develop age-appropriate science system models.
- Take advantage of technology to support active application of science practices.
- Establish the scientific and technical quality of simulation-based assessments.
- Gather evidence of feasibility of implementation of simulation-based assessments.
- Conduct research to examine the impact of formative uses of simulation-based assessments on deep science learning.

These goals have been transformed into suites of assessments in the life sciences (e.g., ecosystems, cells, human body systems) and the physical sciences (e.g., force and motion, atoms and molecules). This WestEd project is grounded in research on model-based learning (Buckley, 2012; Gobert & Buckley, 2000), evidence-centered assessment design (Mislevy et al., 2003), and the cognitive sciences.

SimScientists uses embedded assessments in a regular instructional sequence. So, for example, there is instruction on middle grade performance expectations on Ecosystems: Interactions, energy, and dynamics (MS LS2). As one part of the instructional sequence, students encounter a computer-based assessment plus a reflection activity. Instruction then continues until students take another formative assessment with reflection that addresses disciplinary core ideas and science practices. For classroom teachers, SimScientists assessment systems provide feedback on students' misconceptions and current abilities, graduated coaching, and a progress report for individual students.

The WestEd team conducted this as a research project for two assessment suites. The research was conducted in two states with 26 teachers and 3,694 students. The research findings provide a particularly positive response to the question that heads this session. The team found significantly better performance on pre/post and benchmark assessments of students who used the curriculum-embedded assessments compared to those who did not, and benchmark assessments documented the technical quality of the computer-based simulations. Therefore, the study produced evidence of the effectiveness of simulation-based formative assessments on learning and evidence of the potential of simulation-based assessments to enhance and augment state science assessment systems.

Blended Learning and Assessment

Nancy Butler Songer, University of Michigan, reminded participants that the NGSS, and the performance expectations in particular, were about blended learning. Further, the performance expectations were the assessable components of the NGSS architecture. She went on to underscore another theme of the symposium. The science education community needs good illustrations and rich examples to better understand what blended learning and assessments look like and what information those examples can provide.

Songer first described four assessment tasks that were integral to an integrated instructional unit for fifth grade on ecology and biodiversity. The tasks included different practices and the same disciplinary core ideas. The task had different forms to collect and display data and to construct a scientific explanation.

Songer's second example was from a high school unit on climate change and involved more complex data sets and the use of models to make predictions, based on evidence, about the impact of climate change on the distribution of the Red Squirrel. The scoring rubrics for these formative assessments used "claim, reasoning, and evidence" to evaluate students' responses. This example also included an excellent sample of individual student responses for core ideas and practices across items in a lesson.

Based on her work, Songer presented three conclusions for the symposium participants. First, we must continue the dialogue of what blended learning looks like, what assessment tasks look like, and how to use evidence of blended knowledge for both formative and summative purposes. Second, good assessments may consist of clusters of tasks organized around a family of blended products (e.g., multiple tasks) that pair one disciplinary core idea and crosscutting concept with a few different practices. Disciplinary core ideas and crosscutting concepts work together to provide strong evidence. Finally, coding and feedback can and should provide suggestions on errors of all kinds, including: disciplinary core ideas, crosscutting concepts, practices, and blended products.

The theme for this session was the use of formative assessments to support improved teaching and learning of NGSS. The presentations gave several different uses of formative assessments in the service of both classroom instruction and feedback about student learning.

Understanding the importance of short, medium, and long term cycles within the instructional sequence, using models, games, computer-based embedded assessments, and several assessment tasks within an instructional sequence, all contributed to a positive and constructive answer to the question, "How can formative assessment be used to support improved teaching and learning of the NGSS?"

Session 6

Next Generation Science Standards (NGSS) Assessment System Designs: What Are the Choices, Trade-Offs, and Timelines?

As the NGSS are adopted, states will need to develop their plans for implementation, which may include professional development and instructional materials, formative and benchmark assessment resources, and summative assessments. Because tasks designed to assess students' ability to engage in scientific practices in the context of rich content are expensive to develop and score and require extended time for administration, states will grapple with competing tensions as they envision their assessment programs. In this session, the challenges of assessing NGSS were presented, as were responses designed to address the challenges. Representatives of two states discussed the strategic choices they must make relative to the design choices ahead of them.

Rodger Bybee, Biological Sciences Curriculum Study, reminded the participants that it had only been five months since release of the NGSS. In this short time, seven states have adopted the new standards for science education. He noted different domains of the educational system that are influenced by standards—curriculum and instruction, teacher development, and assessment and accountability. While the latter was the theme of this symposium, the other domains were either referenced or provided a context for the presentation.

Bybee concluded the introduction with the caution to build on the examples provided by presenters throughout the symposium and not to yield to the challenges of designing and implementing assessments for the NGSS.

The Measurement Challenges

Joanna Gorin, Educational Testing Service, summarized a paper she coauthored with Robert Mislevy for this symposium. Their paper, *Inherent Measurement Challenges in the Next Generation Science Standards for Both Formative and Summative Assessment*, brought to the foreground the measurement challenges presented by the NGSS. Gorin began by reminding the participants of the conceptual shifts and sources of complexity in NGSS assessment. She made a point that performance expectations do not specify particular content and do not imply any particular instructional or assessment task.

Gorin reviewed the multiple potential uses of NGSS assessments and introduced and elaborated on three types of challenges.

- Task design and scoring (e.g., appropriate construct models, complexity of tasks)
- Psychometrics (e.g., scaling and estimation, reliability, and generalizability)
- Logistical and practical considerations (e.g., time, cost, technology requirements)

The Gorin and Mislevy paper explored in clear detail the psychometric challenges inherent in the types of assessment that will be required to assess the complex expectations of the NGSS, and it made the point that those challenges are different for different use cases (e.g., formative vs. summative). They concluded with a charge to the psychometric community:

> ...psychometricians would be well advised to broaden their perspectives to embrace alternative approaches to statistical modeling and scoring that may not be as familiar or comfortable to us, but are more likely to provide key stakeholders with the type of information that they so desperately need to improve science education. (Gorin & Mislevy, 2013, p. 24)

Returning to the different use cases for NGSS assessments, Gorin discussed specific ways the unique challenges may be reduced. She concluded by listing three general strategies that she and Mislevy recommended for overcoming challenges in NGSS assessment:

- Build coherence into an NGSS assessment system across use cases.
  - Articulate design choices.
  - Identify design imperatives for optimization in each use case.
- Use evidence-centered design
  - Make trade-offs where needed by considering relationships between design choices and assessment goals.
- *Borrow* information.
  - External information available in a particular use case.
  - Internal information available from the assessment itself.

Applying Evidence-Centered Design to Systems of Science Assessments

The next presentation served to address the challenges through the application of system design options, in particular the use of an evidence-centered design approach when considering NGSS-based assessment systems. Kathleen Scalise, University of Oregon and member of the NRC Committee on the assessment of the Framework, made the presentation. After a clear and informative discussion of evidence-centered design, Scalise introduced and described three options that states might use to implement assessments for NGSS. Because of their importance, the three models are presented here in detail. This description is adapted from the slides Scalise used.

The first option is a curriculum-based model. Here, through an adoption process, each state adopts one or more highly developed curricula that meet state specifications based on the NGSS. These would include NGSS-based instructional materials, embedded assessment, performance tasks, professional development modules, and possibly end-of-unit or end-of-course assessments. Any or all of the assessments could be technology-enhanced.

Local education authorities (LEAs) would then adopt and implement one or more of the state approved curricula. Teachers would score the embedded tasks which, after external moderation to ensure consistency of scoring, become the assessment of record. In this option, the state would monitor and report achievement based on course completion. (See Figure 13.)

The second option involves a group of states working together to develop a library of high quality, extended science performance tasks that each state can utlize within it state assessment system. The tasks involve the integrated dimensions of the NGSS and require student reflection. Participating states could choose to create a divided item bank, with some tasks being shared for use within instruction or interim assessments and others secure for trend analysis. The assessment tasks could range in time from a few days to self-contained multi-week units of instruction with embedded tasks.

This model allows for elements of choice and/or assignment of common tasks and windows of opportunity for assessment. State monitoring would include Opportunity to Learn information from surveys such as NAEP Science and/or school audits. (See Figure 14.)

The third option is a traditional model but with an inverted emphasis. To begin, states would leverage the infrastructure created by the Race to the Top Consortia. What is the inverted emphasis? In this model, the majority of funding and/or effort would go into the development of a library with a deep selection of instructional, classroom assessment, and professional development resources. Using the digital library and item development and banking systems developed with Race to the Top funding, states would develop, vet, and share classroom resources.

Participating states would also develop a set of robust extended NGSS-aligned technology-enhanced assessments (TEAs) that, due to the time required to complete each one, would be matrix sampled across each district and school. These TEA tasks would be designed to serve classroom needs as well as provide critical evidence of learning. State monitoring would involve data from a small group from the TEA standardized tasks that would be matrix sampled across each state and school. In addition, portfolios of individual student TEAs combined with reflection mechanisms such as professional learning communities would be utilized at each school. (See Figure 15.)

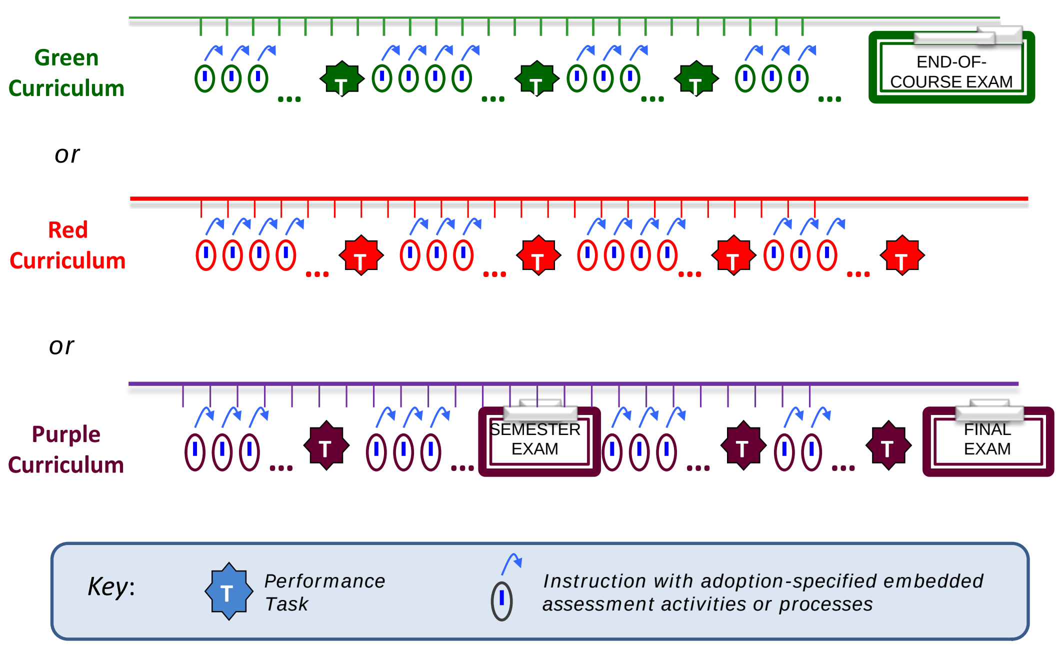


Figure 13. Option A: A curriculum-based model.

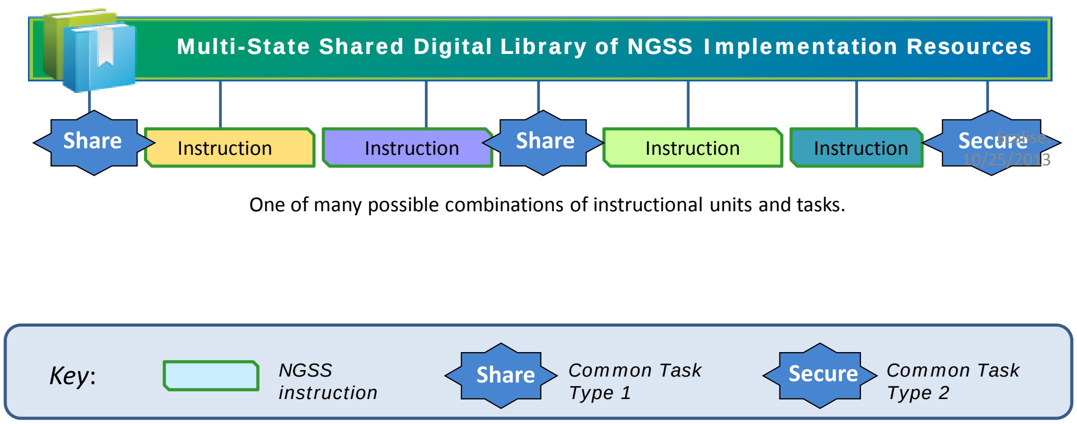


Figure 14. Option B: A common task model. NGSS = Next Generation Science Standards.

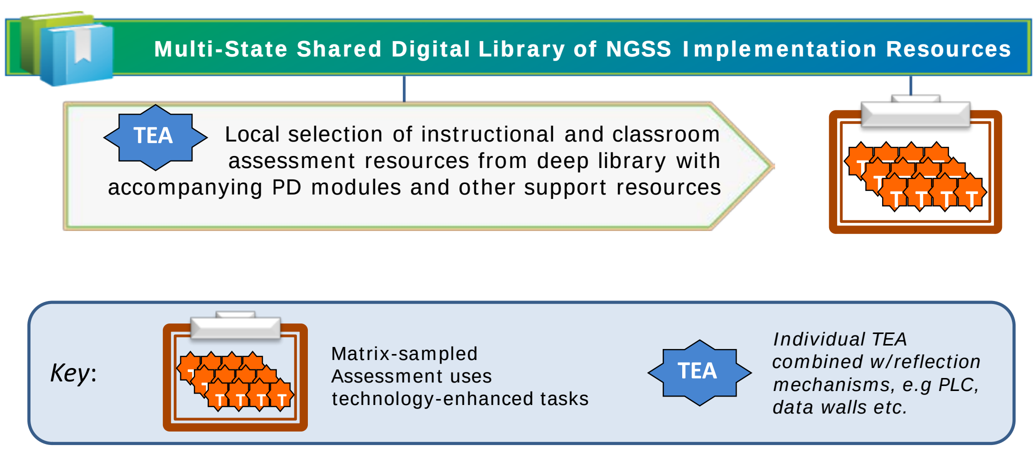


Figure 15. Option C: Traditional model—with inverted emphasis.
PD = professional development, PLC = professional learning community,
NGSS = Next Generation Science Standards, TEA = technology-enhanced assessment.

Scalise then provided a summary of the three system and evidence models and contrasted them to what might be a typical current statewide science assessment. Given the complexity of the constructs to be measured in the NGSS, Scalise concluded that it would not be possible to provide a valid and reliable measure of the NGSS through the use of, for example, a 40 minute multiple choice assessment. (see Figure 16).

	Example 1: Curriculum-Based Model	Example 2: Common Tasks Model	Example 3: Traditional But Inverted Model	Non-Example
Monitor	Accreditation Model	Common tasks, input survey, NAEP	Matrix sample TEA	40-min, 1x/year, multiple-choice test
Instruction	State/Locality Adoption Process	Shared extended common tasks	Classroom-based TEA	No or limited classroom materials
ProgramEval	Course Completion Data Profiles	Secure extended common tasks	Classroom-based TEA, PLC, data wall	Program eval based on Rows 1 & 2
Signify	Tasks exemplify NGSS	Tasks exemplify NGSS	Tasks exemplify NGSS	Questionable proxy (not NGSS-like tasks)

Figure 16. Three of many possible evidence models for science.
NAEP = National Assessment of Educational Progress, NGSS = Next Generation Science Standards, PLC = professional learning community, TEA = technology enhanced assessments.

Scalise concluded her presentation with encouragement to the state teams. She realized the challenges, but the three options afford opportunities that are understandable, manageable, and educationally sound.

Steps Toward Solutions

The session's final speakers provided state supervisors' perspectives. Peter McLaren, Rhode Island Department of Education, acknowledged the problems but quickly moved to recommending several steps toward a solution: (a) build awareness and understanding of both the Framework and NGSS, (b) align curricula to NGSS, (c) shift instruction, (d) attend to the issues associated with the multidimensionality of assessment, and (e) seek collaboration with the state's science education community. McLaren closed with the point that educators need quality exemplars and professional development, and that collaboration is critical. "We need to do this as a nation."

From Standards to Classrooms

Michael Muenks, Missouri Department of Elementary and Secondary Education, discussed the dimensions of purpose, policy, and practical (i.e., classroom) issues, pointing out that state assessments will drive local instruction and the numerous considerations for updated science assessments.

Summary of Part II

In summary, Part II of the symposium identified details of both the assessment challenges and opportunities presented by NGSS. Clearly, the challenge of NGSS involves a different way of thinking about K-12 science education.

This way of thinking involves understanding and assessing three dimensions: scientific and engineering practices, disciplinary core ideas of science, and crosscutting concepts of science.

Instruction designed to provide students opportunities to learn this new perspective was made real by videos and subsequent discussions. One theme that emerged was the need to take time, proceed with the best evidence available, and be patient.

A second theme was the evidence, through existence proofs of current assessments, that it is possible to design assessments for the NGSS. Indeed, large-scale assessments such as NAEP, PISA, and AP science demonstrate that the science education community need not start from scratch with NGSS assessments and that there are examples of computer-based interactive tasks that could be used as the basis for assessing the performance expectations in NGSS.

A third theme centered on formative assessments designed to improve the teaching and learning of NGSS. Here, it is essential to understand the short, medium, and long term cycles within the instructional sequence and use models, games, computer-based embedded assessments, and several different assessments tasks within an instructional sequence. Implementation of formative assessments with these characteristics will address challenges and provide positive and constructive opportunities to answer questions about formative assessments and NGSS.

Finally, there are challenges of task design and scoring, psychometrics, and logistics and practicalities. But, these can be reduced by, for example, building coherence into the NGSS assessment system, using evidence-centered design, making trade-offs where needed, and borrowing external information or using internal information from assessments.

There are at least three evidence models for science assessments. One is based on curriculum, a second involves common tasks, and a third is traditional, but the effort is inverted with major funding going to high quality NGSS technology-enhanced classroom assessment programs.

Part III

Moving Beyond the Challenges

Session 7

Panel Discussion on Moving Forward: The Policy and Practice Work Ahead

Nancy Doorey, Director of Programs at the K-12 Center at ETS, chaired this panel discussion. She began the discussion by reviewing themes from the NRC Framework and NGSS. Helen Quinn, Stanford University and Chair of the NRC committee that developed the Framework, continued the discussion by pointing out that the pieces (i.e., disciplinary core ideas, crosscutting concepts, and science and engineering practices) really are not new. They, however, are put together in a different way. The Framework recommends and the NGSS presents performance expectations that integrate the three dimensions or *pieces*.

Lisa Brody, a science teacher at Park View Middle School in Morton Grove, Illinois, pointed out that the changes in teaching practice emphasize the need for professional development and support from the community. David Evans, Executive Director, National Science Teachers Association (NSTA), supported Brody's points and gave examples of resources that NSTA is providing that will support teachers and build public support for the changes required by NGSS.

Richard Shavelson, Stanford University, observed "we have done this before, at least four times during my career." He made reference to *Tinkering Toward Utopia*, a 1997 book by David Tyack and Larry Cuban. The title expresses Shavelson's point. The four examples of science education reform referenced by Shavelson were the late 1950s and the Sputnik era of curriculum reform, the mid 1980s and *A Nation at Risk*, the 1990s and standards-based reform, and the contemporary Next Generation Science Standards. Shavelson identified a couple of lessons worth noting—build small tasks that exemplify the changes and build public understanding to counter questions and concerns.

Shirley Malcom, Head of the Directorate for Education and Human Resources Programs at the American Association for the Advancement of Science (AAAS), added her support for the themes of supporting teachers and making small advances. She cited one critical advance that must be addressed is equity and noted the equity issues are huge.

The point was made that, "Every classroom does not have a Lisa," and further that the great teachers, like Brody, are not distributed equitably. The likelihood of an excellent teacher in some districts is smaller than in other districts, and, within schools, the better teachers are sometimes more likely to be assigned to teach the better students.

Brody's response was, "I'm not different from other teachers," just fortunate to have had the benefit of high quality professional development and curriculum guidance. Other teachers can gain the skills and understanding needed to implement the NGSS with fidelity, she argued, if also given high quality professional development.

In the final few minutes of the panel, Doorey asked the panelists to briefly give their best advice on moving forward on the NGSS and assessments. Following are the panelists' recommendations.

- Helen Quinn advised—stay focused, maintain the course, and take the time needed to do it well.
- David Evans recommended—work with as many teachers as possible. Remember the real goal and value of the work is with teachers and students.
- Lisa Brody instructed—continue to communicate with and involve teachers. Reach out to teachers and open the doors to professional development.
- Shirley Malcom suggested—look to the nonfictional components of CCSS, as this may be a great opportunity for the science education community.
- Richard Shavelson counseled—continue providing leadership at the national, state, and local levels. Establish a strategic plan that includes assessment at these levels and addresses curriculum and teacher development.

Session 8

Next Steps in Working with States and Districts on Next Generation Science Standards (NGSS) and Assessments

Chris Minnich, Executive Director of the CCSSO, began with the reality of educational problems but provided encouragement, with recognition that we are changing standards, people are paying attention to our accomplishments, and the challenges are all within the educational community's capacity to address and resolve.

Stephen Pruitt, Senior Vice President at Achieve and leader of the team developing the NGSS, also provided some concluding comments. In the next steps of working with states and districts on Next Generation Science Standards and Assessments, educational leaders should: (a) emphasize the deep understanding of science described in NGSS; (b) recognize the importance of instruction and the organization of content; and (c) choose between what is easy and what is the right thing to do relative to assessing NGSS. He ended with the advice to look for opportunities, take time, and have the courage to be patient.

Moving Beyond the Challenges: A Concluding Discussion

The release of the NGSS in April 2013 set the stage for educational reforms at the national, state, and local levels. These reforms center on teachers' professional development, school programs, and assessments and accountability.

Recognizing the reform based on NGSS, the Center for K-12 Assessment & Performance Management at ETS joined with the CCSSO and the College Board to organize a two-day research symposium on the topic of next generation science assessments. The symposium was held on September, 24 and 25, 2013, in Washington, DC.

Approximately 200 leaders in science education and assessment attended the symposium. Two State Collaboratives on Assessment and Student Standards (SCASS)—TILSA (Technical Issues in Large-Scale Assessment) and the Science Assessment SCASS—represented approximately 100 attendees. In addition, participants included representatives of three groups critically involved in the development of the NGSS: the NGSS Writing Team and the National Research Council's committees on New Science Education Standards and Developing Assessment of Science Proficiency in K-12. A diverse group of science and measurement experts representing a range of higher education institutions and testing organizations made presentations and engaged in discussions of critical topics.

This report began with an overview of NGSS and continued with summaries of the presentations and discussions. This concluding discussion summarizes some perspectives that may help readers to move beyond the challenges.

Understanding the Influence of NGSS on the Educational System

The NGSS represent a significant departure from past approaches to science education. With policies representing a new vision of science education, one should reasonably ask—What changes are implied for states and districts adopting NGSS? How will NGSS affect curriculum, instruction, and assessments? And, in time, how will NGSS affect students' learning and achievement? It is worth noting that the US education system is a large, complex, and layered system. In that system, one recognizes the form, function, and decisions at national, state, local, and classroom levels of organization.

The National Research Council report *Investigating the Influence of Standards: A Framework for Research in Mathematics, Science, and Technology Education* (NRC, 2002) provides guidance in understanding the channels of influence within the education system and, in this discussion, specifically on assessments. Figure 17 describes three main routes or *channels* through which reform ideas based on national standards influence various levels of the educational system and eventually classroom teaching, student learning, and achievement. The primary channels through which the NGSS will influence the educational system are curriculum, teacher development, and assessment and accountability (see Figure 17). The latter, assessment and accountability, with assessment writ large as everything from formative assessment embedded within instruction to consequential assessments, was the focus of the K-12 Center conference and this report.

How has the system responded to the question, What are the introductions of nationally developed consequences for standards and student learning?

Contextual Forces

- Politicians and Policy Makers
- Public
- Business and Industry
- Professional Organizations

Channels of Influence Within the Education System

Curriculum
- State, district policy decisions
- Instructional materials development
- Text, materials selection

Teacher Development
- Initial preparation
- Certification
- Professional development

Assessment and Accountability
- Accountability systems
- Classroom assessment
- State, district assessment
- College entrance, placement practices

→ **Teachers and Teaching Practice** in classroom and school → **Student Learning**

Figure 17. The influence of standards on the educational system.

Many speakers either mentioned or gave curricular and instructional examples related to assessment. The perspective summarized in Figure 17 will help science education move beyond the immediate challenges of assessment and address other components of the educational system.

Translating NGSS to Assessment

Producing an accurate translation requires a thorough understanding of both the language and the culture of the original product (e.g., a play or poem), as well as the contemporary culture. In addition, one must understand the trade-offs, as well as the advantages, of a translation. So it is with translating NGSS to assessments. NGSS has a language and criteria that are important. Likewise, there are constraints and concepts of assessment that should be honored.

The many examples provided by the speakers give evidence that we have already moved beyond the challenges. The capacity to assess disciplinary core ideas and practices was clearly demonstrated. However, examples of assessing three dimensions were not clear. Thinking in terms of assessment units (e.g., PISA) is an excellent way to move beyond old perceptions of *a test*.

Use of evidence-centered design will be essential for those at the state and local districts who have to develop and implement assessments. The ECD approach will facilitate the process of appropriately synthesizing the various examples into initial assessments of NGSS.

To Conclude

As made clear by many speakers, translating NGSS to assessments presents numerous challenges, particularly for the design of consequential assessments if current constraints of testing time, cost, and uses of data are to be adhered to. We have to recognize and understand those challenges. As important as this may be, it is more important to move beyond the challenges and help those with the responsibility for developing the next generation science assessments.

References

Alonzo, A. (2013, September). *What can be learned from current large-scale assessments and what are the implications for assessing the NGSS?* Paper presented at the Invitational Research Symposium on Science Assessment, Washington, DC. Retrieved from http://www.k12center.org/rsc/pdf/alonzo.pdfAmerican Association for the Advancement of Science. (AAAS). (1993). *Benchmarks for science literacy.* New York, NY: Oxford University Press.

Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education, 85,* 536–553.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice, 5*(1), 7–74.

Buckley, B. (2012). Model based learning. *Encyclopedia of the Sciences of Learning.* , Springer Science+Business Media, LLC. doi 10.1007/978-1-4419-1428-6

Council of Chief State School Officers. (2008). *Attributes of effective formative assessment.* Washington, DC: Author.

DeBarger, A. H., Penuel, W. R., & Harris, C. J. (2013, September). *Designing NGSS assessments to evaluate the efficacy of curriculum interventions*. Paper presented at the Invitational Research Symposium on Science Assessment, Washington, DC. Retrieved from http://www.k12center.org/rsc/pdf/debarger-penuel-harris.pdfGobert, J., & Buckley, B. (2000). Introduction to model-based teaching and learning. *International Journal of Science Education, 22*(9), 891–894.

Gorin, J. S., & Mislevy, R. J. (2013, September). *Inherent measurement challenges in the Next Generation Science Standards for both formative and summative assessment.* Paper presented at the Invitational Research Symposium on Science Assessment, Washington, DC. Retrieved from http://www.k12center.org/rsc/pdf/gorin-mislevy.pdf

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77,* 81–112.

Herman, J. (2013, September). *Formative assessment for Next Generation Science Standards: A proposed model.* Paper presented at the Invitational Research Symposium on Science Assessment, Washington, DC. Retrieved from http://www.k12center.org/rsc/pdf/herman.pdf

Heritage, M. (2010). *Formative assessment: Making it happen in the classroom.* Thousand Oaks, CA: Corwin Press.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30,* 28–37.

Mislevy, R. J., Almond, R., & Lukas, J. (2003). *A brief introduction to evidence-centered design* (Research Report No. RR-03-16). Princeton, NJ: Educational Testing Service.

National Research Council. (1996). *National science education standards*. Washington, DC: The National Academies Press.

National Research Council. (2002). *Investigating the influence of standards: A framework for research in mathematics, science, and technology education.* Washington, DC: The National Academies Press.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas.* Washington, DC: The National Academies Press.

NGSS Lead States. (2013). *Next generation science standards: For states, by states.* Washington, DC: The National Academy Press.

Pellegrino, J., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know.* Washington, DC: National Academies Press.

Rutherford, F. J., & Algren, A. (1989). *Science for all Americans*. New York, NY: Oxford University Press.

Shavelson, R. J., Young, D. B, Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A. (2008). On the impact of curriculum-embedded formative assessment on learning. *Applied Measurement in Education, 21,* 295–314.

Tyack, D., & Cuban, L. (1997). *Tinkering toward utopia: A century of public school reform.* Cambridge, MA: Harvard University Press.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: Impact on student achievement. *Assessment in Education: Principles, Policy, and Practice, 11*(1), 49–65.

K-12 Center at ETS

Invitational Research Symposium on Science Assessment

Appendix A

Agenda

K-12 Center at ETS | CCSSO Council of Chief State School Officers | CollegeBoard

Invitational Research Symposium on
Science Assessment
Washington Hilton Hotel, Washington, DC September 24-25, 2013

Time	*Day 1: September 24, 2013*
7:30	REGISTRATION and Continental Breakfast
8:30	WELCOME: Pat Forgione, K-12 Center at ETS
8:45	1 **What Is the Vision of Science Proficiency and Science Instruction Underlying the NGSS, and What New Demands Do They Create?** Moderator: Jim Pellegrino, UIL-Chicago • Opening Comments by Moderator • The Ongoing Role of the K-12 Science Framework: Helen Quinn, Stanford Univ./NRC • The Vision of Science Instruction: Brian Reiser, Northwestern University, Joseph Krajcik, Michigan State University, and Lisa Brody, Middle School Teacher • Development Of the NGSS: Stephen Pruitt, Achieve/NGSS • What Does This Mean For States?: Juan-Carlos Aguilar, Georgia State Science Supervisor Table Discussion Followed by Group Discussion with Moderator
10:30	BREAK
10:45	2 **Measuring Across Multiple Dimensions: How Can We Measure the Ability to Engage in Science Practices in the Context of Rich Content?** Moderator: Richard Lehrer, Vanderbilt University • Opening Comments by Moderator • Update on the Work of the NRC Committee on Science Assessment: Mark Wilson, UC-Berkeley • Prototype Task To Measure Across Multiple Dimensions: Angela DeBarger, SRI, Christopher Harris, SRI, and William Penuel, UCO-Boulder Table Discussion Followed by Group Discussion with Moderator
12:15	LUNCH
1:10	3 **What Can Be Learned from Current Large-Scale Assessments and What are the Implications for Assessing the NGSS?** Moderator: Jonathan Osborne, Stanford University • Opening Comments by Moderator **Joint Presentation:** • What Can Be Learned from Current Large-Scale Assessment Programs to Inform Assessment of the Next Generation Science Standards? Alicia Alonzo, Michigan State University • Presentations Of Selected Items/Tasks by Developers Of Those Assessments: o NAEP: Peggy Carr, NCES o PISA: Janet Koster van Groos, ETS and Eric Steinhauer, ETS o Advanced Placement Physics: Karen Lionberger, College Board o New Zealand National Science Assessment: Jennifer Mackrell, NZQA Table Discussion Followed by Group Discussion with Moderator
3:10	BREAK
3:30	**4A Breakout Session:** 3:30pm-4:20pm *(With a Choice of Five Options Available)*
4:20	MOVEMENT TO NEXT BREAKOUT SESSION
4:40	**4B Breakout Session:** 4:40pm-5:30pm *(With a Choice of Four Options Available)*
5:30	**CCSSO/SCASS RECEPTION**: Open to all Symposium Attendees
	Reception provided by our Sponsors: 5:30pm-7:30pm

K-12 Center at ETS

Invitational Research Symposium on
Science Assessment

Day 2: September 25, 2013

Time	
7:00	Continental Breakfast
8:00	Welcome: Pat Forgione, K-12 Center at ETS
8:15	**5 How Can Formative Assessment Be Used to Support Improved Teaching and Learning of the NGSS?** Moderator: Joan Herman, CRESST/UCLA • Formative Assessment for Next Generation Science Standards: A Proposed Model. Joan Herman, CRESST/UCLA • Presentations of Selected Exemplars of Formative Science Assessments by the Developers: o Joe Krajcik, Michigan State University o Malcolm Bauer, ETS o Edys Quellmalz, WestEd o Nancy Butler Songer, University of Michigan Table Discussion Followed by Group Discussion with Moderator
10:00	BREAK
10:15	**6 NGSS Assessment System Designs: What Are the Challenges, Choices and Trade-Offs?** Moderator: Rodger Bybee, BSCS (retired) • Opening Comments by Moderator • Challenges In the Assessment of the NGSS: Joanna Gorin, ETS • System Design Options: Kathleen Scalise, University of Oregon • Making Strategic Choices To Support Improvement: A State Science Supervisor's Perspective: Peter McLaren, Rhode Island DOE • Making Strategic Choices To Support Improvement: A State Assessment Director's Perspective: Michael Muenks, Missouri DOE Table Discussion Followed by Group Discussion with Moderator
12:15	**LUNCH**
1:15	**7 Panel Discussion on Moving Forward: The Policy and Practice Work Ahead** Moderator: Nancy Doorey, K-12 Center at ETS • Opening Comments by Moderator • Panelist Discussion o Lisa Brody, Teacher, Park View School, Illinois o David Evans, National Science Teachers Association o Shirley Malcom, AAAS o Helen Quinn, Stanford University (Emeritus)/NRC o Richard Shavelson, Stanford University (Emeritus) Group Discussion with Moderator
2:30	**8 Closing Comments: Next Steps in Working with States and Districts on Next Generation Science Standards and Assessments** o Chris Minnich, CCSSO o Stephen Pruitt, Achieve/NGSS and the Science SCASS group at CCSSO
2:45	Adjournment: Nancy Doorey, K-12 Center at ETS
3:00	End

Appendix B

The Breakout Sessions

What Do We Know About Effective Professional Development for Science Educators, and What are the Implications for Statewide Transition to the NGSS? A Review of Large-Scale Reform Initiatives in Science

The success of the NGSS will depend to a great degree on the readiness of K-12 science educators to deliver high-quality, aligned instruction. Because these standards include important changes to reflect changes in the fields of science and engineering, states will need to ensure the availability of effective professional development resources and activities. Brian Reiser, Professor of Learning Sciences in the School of Education and Social Policy at Northwestern University, will present his newly released paper reviewing lessons learned from past effective professional development initiatives in K-12 science education and their implications for future efforts. Janet Coffey, a former middle school science teacher and faculty member in science education at University of Maryland, College Park's College of Education, now Program Officer for the Gordon and Betty Moore Foundation, will serve as Discussant.

- Presenter: Brian Reiser, Northwestern University
- Discussant: Janet Coffey, Gordon and Betty Moore Foundation
- Moderator: Donna Matovinovic, CTB/McGraw-Hill—Introduction, Q&A, and Discussion

Virtual Performance Assessment and Games: Potential as Learning and Assessment Tools

Advances in technology and measurement have the potential to play an important role in next-generation assessments. The Virtual Performance Assessment (VPA) Project (http://vpa.gse.harvard.edu) is studying the feasibility, practicality, and value of using immersive virtual performance assessments to measure science inquiry and problem solving. In this presentation, we will present research on how these assessments differ from conventional summative assessments and the consequential validity for using them in high-stakes settings. We found that in classroom settings this type of assessment is practical, affordable, and engaging for students. Also, insights from the design and scoring of the VPAs can guide the development of diagnostic assessments, formative for instruction, based on immersive authentic simulations. This type of assessment is potentially well suited to a variety of higher-order skills, including 21st century capabilities such as collaborative problem solving and leadership. Chris Dede and Jody Clarke-Midura of Harvard University will present their recent work, and

Kathleen Scalise of the University of Oregon will discuss the potential use of VPA for consequential purposes.

- Presenters: Chris Dede and Jody Clarke-Midura, Harvard University
- Presenter: Kathleen Scalise, University of Oregon
- Moderator: Tim Crockett, Measured Progress—Introduction, Q&A, and Discussion

Personalized Learning: The Promise of Ongoing Diagnosis and Adaptation

This session will feature presentations by Kristen DiCerbo of Pearson and André Rupp of ETS. DiCerbo will discuss the promise of adaptivity and personalization as ways to increase engagement, encourage practice, and ultimately improve learning. An overview of categories of methods for adapting to student proficiency levels will be discussed, from basic leveling to complex skill modeling. Examples from game design will be used to demonstrate the concepts discussed. Rupp will discuss opportunities and challenges for the automated identification, aggregation, and synthesis of evidence in the service of assessment needs that include creating diagnostic feedback and integrated score reports, as well as the identification of *unusual* learner behaviors and characteristics. He will discuss these issues specifically for performance-based assessment contexts that include data from end-of-state work products and workflow processes.

- Presenter: André Rupp, ETS
- Presenter: Kristen DiCerbo, Pearson
- Moderator: Doug Moore, Amplify—Introduction, Q&A, and Discussion

NGSS-aligned Formative Assessment Projects—Part I

Two of the presenters of formative science assessment projects from Session 5 will provide additional information about their projects and respond to questions. Joseph Krajcik, a former high school chemistry and physical science teacher and current Director of the CREATE for STEM Institute and faculty member in science education at Michigan State University, will discuss the Investigating and Questioning our World through Science and Technology (IQWST) project that is developing the next generation of middle school curricula designed to enable teachers with diverse knowledge and experiences to teach science effectively to students with a variety of backgrounds and strengths. Malcolm Bauer, a senior research scientist at ETS, will discuss the GlassLab project, which is exploring the potential for existing, commercially successful digital games to serve both as potent learning environments and real-time assessments of student learning.

- Presenter: Joe Krajcik, Michigan State University
- Presenter: Malcolm Bauer, ETS
- Moderator: Alexa Posny, Renaissance Learning—Introduction, Q&A, and Discussion

The New Zealand National Certificate of Educational Achievement: Melding Internal and External Assessments

Over the past decade, New Zealand has developed a new system of national student assessments. One unique aspect of this system is the incorporation of both external and internal (school-based) assessments. Those standards that can be validly and reliably assessed through conventional testing methods are externally assessed, but those that are more performance-based, such as scientific investigations, are internally assessed with moderation processes in place to ensure consistency. Jennifer Mackrell, a national assessment facilitator for the Secondary Examinations Qualifications Division of the New Zealand Qualifications Authority, will describe this system and the teacher training, scoring, and moderation of the internal assessments.

- Presenter: Jennifer Mackrell, New Zealand Qualifications Authority
- Discussant: Anita Bernhardt, Maine Department of Education
- Moderator: Marc Julian, Data Recognition Corporation—Introduction, Q&A, and Discussion

Measurement of Practices and Procedural Knowledge in Professional Licensure Exams

The use of simulations is quite new within high-stakes K-12 assessment, but it has been done for more than a decade within several professional licensure/certification examinations. Brian Clauser, Vice President for Measurement Consulting Services for the National Board of Medical Examiners and Editor of the *Journal of Educational Measurement*, will describe the simulations used within the examinations required to obtain a license to practice medicine within the United States. Chris Adams of FlightSafety International Inc. will share how virtual reality simulations are used in commercial pilot certification.

- Presenter: Brian Clauser, National Board of Medical Examiners
- Presenter: Chris Adams, FlightSafety International, Inc.
- Moderator: Nathan Wall, eMetric—Introduction, Q&A, and Discussion

Redesigned Advanced Placement Frameworks, Courses, and Lessons Learned to Date

In 2002, a committee of the NRC called for a revision of AP courses that would support focus on modern core competencies and knowledge, as well as in the sciences, allowing students to “experience problem solving, controversies, and the subtleties of scholarly investigation.” The College Board has since undertaken a major revision of the AP frameworks, courses, and assessments. This session will explore those revisions and the lessons learned to date.

- Presenters: Auditi Chakravarty, Tom Matts, Serena Magrogan, and LaTanya Sharpe, College Board
- Discussant: Jami-Jon Pearson, North Carolina Department of Education
- Moderator: Michael Bunch, Measurement Incorporated—Introduction, Q&A, and Discussion

The New Zealand System of National Certificates: Connecting Secondary Education to Future Careers

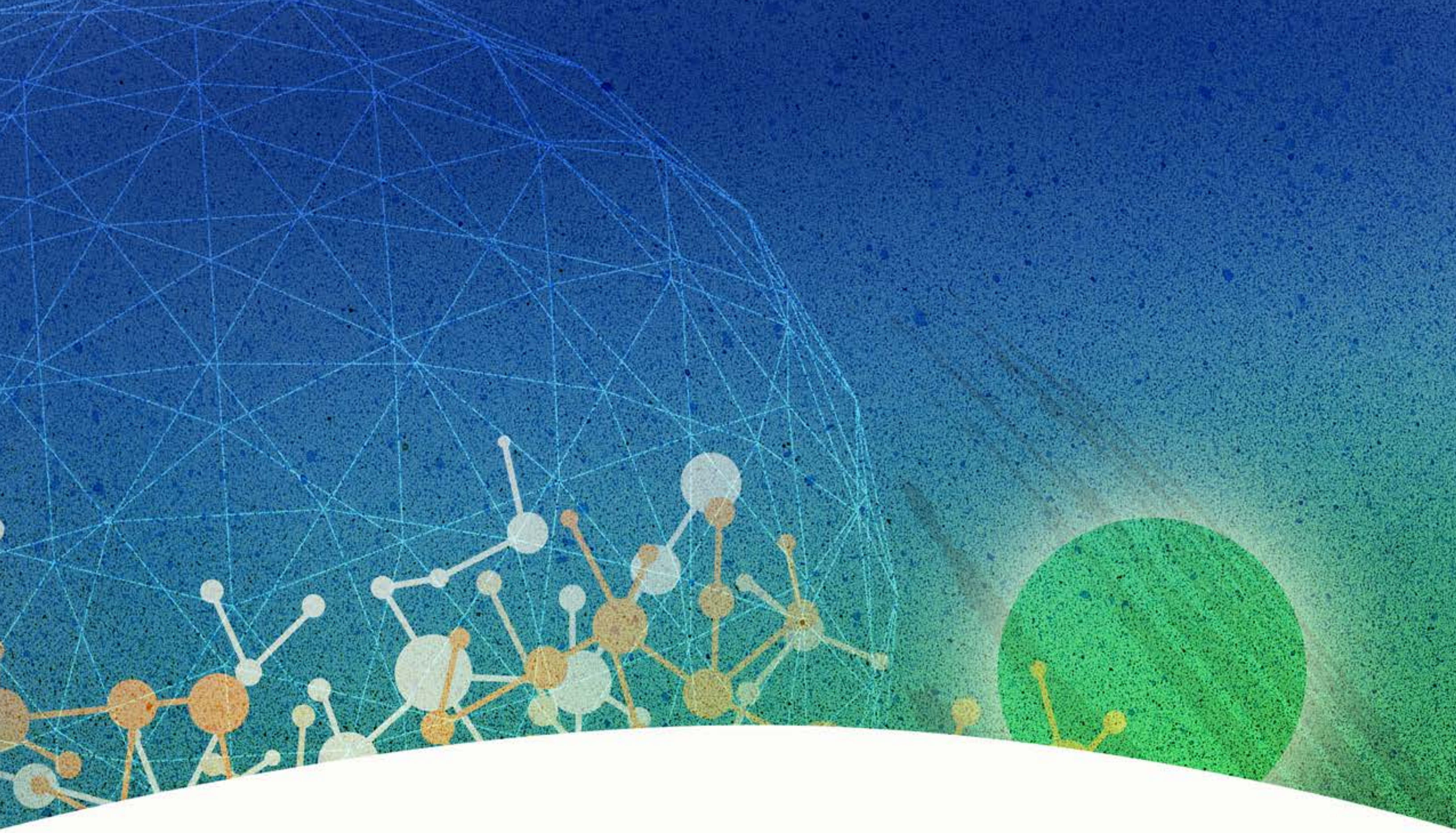
In order to increase student engagement and readiness for postsecondary study or work, New Zealand redesigned its secondary educational system and assessments in the early 2000s. High school students can accumulate credits toward more than 1,000 national certificates and diplomas that recognize skills and knowledge that meet nationally endorsed academic and industry standards. Students may take courses leading to national certificates and diplomas at high schools, polytechnics, or private training establishments. In addition, workplace-based training programs and apprenticeships also lead to national certificates and diplomas. Jennifer Mackrell, a national assessment facilitator for the Secondary Examinations Qualifications Division of the New Zealand Qualifications Authority, will describe this system.

- Presenter: Jennifer Mackrell, New Zealand Qualifications Authority
- Discussant: Pete Goldschmidt, New Mexico Department of Education
- Moderator: Phoebe Winter, Pacific Metrics—Introduction, Q&A and Discussion

NGSS-Aligned Formative Assessment Projects—Part II

Two of the presenters of formative science assessment projects from Session 5 will provide additional information about their projects and respond to questions. Edys Quellmalz, Director of Technology Enhanced Assessments and Learning Systems in the Science, Technology, Engineering, and Mathematics Program at WestEd, will discuss the new SimScientists module on human body systems and modeling. Nancy Butler Songer, Professor of Science Education and Technology and Director of the Center for Essential Science at the University of Michigan, will discuss her work on a curricular and learning progression focused on complex reasoning about ecology, biodiversity, and the impact of global changes on populations of animals and ecosystems.

- Presenters: Edys Quellmalz and Mark Loveland, WestEd
- Presenter: Nancy Butler Songer, University of Michigan
- Moderator: Deval Clearwater, HMH–Riverside Publishing—Introduction, Q&A, and Discussion



The Center for K-12 Assessment & Performance Management at ETS creates timely events where conversations regarding new assessment challenges can take place and publishes and disseminates the best thinking and research on the range of measurement issues facing national, state, and local decision makers.

K-12 Center at ETS

Invitational Research Symposium on
Science Assessment