Developing Assessments for the Next Generation Science Standards

Jim Pellegrino
Foci of Today’s Presentation

• Science Education: Past, Present, Future?
  • Defining Competence to Achieve Coherence
    – Unpacking the Components of Competence
• From NGSS Performance Expectations to Coherence in K-12 Science Education
• NRC Report Conclusions & Recommendations
• Moving Forward: Prospects & Issues
Standards & Frameworks as Guides for Reform in K-12 Science

1990s-2009

1/2010 - 7/2011

Issues in U.S. Science Education

- Multiple sets of standards – 50 states
- Lack of coherent instructional sequences – topics and modules -- mix and match
- Lots of “hands-on” but little “minds-on” inquiry activities in modules and kits
- Focus on declarative knowledge on tests
- Focus on “scientific method” absent content
- Poor performance on NAEP science
- Poor performance on PISA science
Foci of Today’s Presentation

• Science Educations: Past, Present, Future?
• Defining Competence to Achieve Coherence
  – Unpacking the Components of Competence
• From NGSS Performance Expectations to Coherence in K-12 Science Education
• NRC Report Conclusions & Recommendations
• Moving Forward: Prospects & Issues
New Definition of Competence

• The NRC Science Framework has proposed descriptions of student competence as being the intersection of knowledge involving:
  – important disciplinary practices
  – core disciplinary ideas,
  – and crosscutting concepts with
  – performance expectations representing the intersection of the three.

• It views competence as something that develops over time & increases in sophistication and power as the product of coherent curriculum & instruction
A Core Idea for K-12 Science Instruction is a Scientific Idea that:

• Has **broad importance** across multiple science or engineering disciplines or is a **key organizing concept** of a single discipline

• Provides a **key tool** for understanding or investigating more complex ideas and solving problems

• Relates to the **interests and life experiences of students** or can be connected to **societal or personal concerns** that require scientific or technical knowledge

• Is **teachable** and **learnable** over multiple grades at increasing levels of depth and sophistication
Evolution from Inquiry Standards to Scientific Practices

THE REAL WORLD
- Ask Questions
- Observe
- Experiment
- Measure

COLLECT DATA
TEST SOLUTIONS

Investigating

ARGUE CRITIQUE ANALYZE

THEORIES AND MODELS
- Imagine
- Reason
- Calculate
- Predict

FORMULATE HYPOTHESES
PROPOSE SOLUTIONS

Developing Explanations and Solutions

Social Interaction and Discourse

Emphasis on Knowledge Building

Inquiry

Inquiry
Scientific and Engineering Practices

1. Asking questions and defining problems
2. Developing and using models
3. Planning and carrying out investigations
4. Analyzing and interpreting data
5. Using mathematics and computational thinking
6. Developing explanations and designing solutions
7. Engaging in argument from evidence
8. Obtaining, evaluating, and communicating information
Crosscutting Concepts

- Some important themes pervade science, mathematics, and technology and appear over and over again, whether we are looking at an ancient civilization, the human body, or a comet. They are ideas that transcend disciplinary boundaries and prove fruitful in explanation, in theory, in observation, and in design.

  — American Association for the Advancement of Science

1. Patterns
2. Cause & Effect: Mechanism & Explanation
3. Scale, Proportion & Quantity
4. Systems and System Models
6. Structure and Function
7. Stability and Change
NRC Framework’s Goals for Teaching & Learning

• Coherent investigations of core ideas across multiple years of schooling
• More seamless blending of practices with core ideas
• Performance expectations that require reasoning with core disciplinary ideas
  - explain, justify, predict, model, describe, prove, solve, illustrate, argue, etc.
Foci of Today’s Presentation

• Science Education: Past, Present, Future?
• Defining Competence to Achieve Coherence
  – Unpacking the Components of Competence
• From NGSS Performance Expectations to Coherence in K-12 Science Education
• NRC Report Conclusions & Recommendations
• Moving Forward: Prospects & Issues
Two Major Features of the NGSS

• Built on the idea of Progressions in the Sophistication of Student Understanding - as previously articulated in the NRC Framework

• Include a new “Architecture” with a focus on Performance Expectations that draw from the intersections of disciplinary core ideas, science and engineering practices, and cross-cutting concepts
Students who demonstrate understanding can:

4-LS1-1. Construct an argument that plants and animals have internal and external structures that function to support survival, growth, behavior, and reproduction. [Clarification Statement: Examples of structures could include thorns, stems, roots, colored petals, heart, stomach, lung, brain, and skin.] [Assessment Boundary: Assessment is limited to macroscopic structures within plant and animal systems.]

4-LS1-2. Use a model to describe that animals receive different types of information through their senses, process the information in their brain, and respond to the information in different ways. [Clarification Statement: Emphasis is on systems of information transfer.] [Assessment Boundary: Assessment does not include the mechanisms by which the brain stores and recalls information or the mechanisms of how sensory receptors function.]

Connections to other DCIs in fourth grade: N/A

Articulation of DCIs across grade-levels:
1.LS1.A (4-LS1-1); 1.LS1.D (4-LS1-1); 3.LS3.B (4-LS1-1); MS.LS1.A (4-LS1-1); MS.LS1.B (4-LS1-2); MS.LS1.D (4-LS1-2)

Common Core State Standards Connections:
ELA/Literacy -
W.4.1 Write opinion pieces on topics or texts, supporting a point of view with reasons and information. (4-LS1-1)
SL.4.5 Add audio recordings and visual displays to presentations when appropriate to enhance the development of main ideas or themes. (4-LS1-2)
Mathematics -
4.G.A.3 Recognize a line of symmetry for a two-dimensional figure as a line across the figure such that the figure can be folded across the line into matching parts. Identify line-symmetric figures and draw lines of symmetry. (4-LS1-1)

* The performance expectations marked with an asterisk integrate traditional science content with engineering through a Practice or Disciplinary Core Idea.
Pluses & Minuses of Relying on Performance Expectations

+ Avoid vague cognitive verbs – “know” & “understand”
+ Stated as claims about students in terms of what they are supposed to be able to do to demonstrate their knowledge
+ Identify progressions as part of expectations
  - Don’t tell us how to get there – curriculum materials and instructional practices
  - Need to be “unpacked” in terms of the forms of evidence needed to support the student claim
NGSS as the Basis for Aligning C-I-A

Assessment

NRC Framework & NGSS

Curriculum

Instruction
Assessment Designed to Support Instruction

• To develop the skills and dispositions to use scientific and engineering practices needed to further their learning and to solve problems, students need to experience instruction in which they
  – use multiple practices in developing a particular core idea and
  – apply each practice in the context of multiple core ideas.

• Effective use of the practices will require that they be used in concert with one another, such as in supporting explanation with an argument or using mathematics to analyze data

• Assessments will be critical supports for this instruction.
Foci of Today’s Presentation

• Science Education: Past, Present, Future?
• Defining Competence to Achieve Coherence
  – Unpacking the Components of Competence
• From NGSS Performance Expectations to Coherence in K-12 Science Education
• NRC Report Conclusions & Recommendations
• Moving Forward: Prospects & Issues
Committee on the Assessment of K-12 Science Proficiency

Board on Testing and Assessment
and
Board on Science Education
National Academy of Sciences
Committee’s Charge

Identify strategies for developing assessments that validly measure student proficiency in science.

Review recent and current, ongoing work in science assessment and what additional research and development is required.

Make recommendations for state and national policymakers, research organizations, assessment developers, and study sponsors about steps needed to develop valid, reliable and fair assessments for the Framework’s vision of science education.
Committee Members

James W. Pellegrino, University of Illinois at Chicago (co-chair)
Mark R. Wilson, University of California, Berkeley (co-chair)
Peter McLaren, Rhode Island Department of Elementary and Secondary Education
Knut Neumann, Leibniz Institute for Science and Mathematics Education
Kathleen Scalise, University of Oregon
Richard Lehrer, Peabody College of Vanderbilt University
William Penuel, University of Colorado at Boulder
Brian Reiser, Northwestern University

Nancy Butler Songer, University of Michigan
Richard M. Amasino, University of Wisconsin, Madison (life sciences)
Helen R. Quinn, Stanford University (physics)
Roberta Tanner, Loveland High School, CO (engineering)
Edward Haertel, Stanford University
Joan Herman, CRESST, UCLA
Scott F. Marion, National Center for the Improvement of Education Assessment
Jerome M. Shaw, University of California, Santa Cruz
Catherine J. Welch, University of Iowa
Foundational Resources for Assessment Design & Use
1. **Assessment tasks should allow students to engage in science practices in the context of disciplinary core ideas and crosscutting concepts. This poses a significant design challenge.**
   - Multi-component tasks that make use of a variety of response formats will be best suited for this.
   - Selected-response questions, short and extended constructed response questions, and performance tasks can all be used, but should be carefully designed to ensure that they measure the intended construct and support the intended inference.

2. **Students will need multiple and varied assessment opportunities to demonstrate their proficiencies with the NGSS performance expectations.**
Main Messages (cont.)

3. A system of assessments will be required and should include classroom assessment, monitoring (large-scale) assessments, and indicators of opportunity to learn.

- Classroom assessment should be an integral part of instruction and should reinforce the type of science learning envisioned in the framework and NGSS.

- Monitoring (large-scale) assessments will need to include an on-demand component and a component based in the classroom (classroom-embedded) in order to fully cover the breadth and depth of the NGSS performance expectations.

- Indicators of opportunity to learn should document that students have the opportunity to learn science in the way called for in the framework and NGSS and that schools have appropriate resources.

- There are various ways to assemble a system – examples cases are presented
Main Messages (cont.)

4. Implementation should be gradual, systematic, and carefully prioritized, beginning with classroom assessment and moving to monitoring assessment.

5. Professional development and adequate support for teachers will be critical.
Multicomponent Tasks

• To adequately cover the three dimensions, assessment tasks will need to contain multiple components (e.g., a set of interrelated questions).

• Specific components may focus on individual practices, core ideas, or crosscutting concepts, but, together, the components need to support inferences about students’ three-dimensional science learning as described in a given performance expectation.
The picture below shows a place on the ocean floor where two plates are moving apart. At this plate boundary (shown at the dotted line), rock material is rising to the surface.

A. Draw on the picture to show what is happening in the mantle that causes the plates to move apart.

B. What is happening in the mantle that helps to explain why the two plates are moving apart?

C. Put an X on the places in the picture above where the oldest rock can be found in the crust.

D. Explain your answer.
Highlighting the Contrast

The major movement of the plates and description of plate boundaries of the Earth are...

A. Convergent
B. Divergent
C. Transform
D. All of the Above

A. Draw a model of volcano formation at a hot spot using arrows to show movement in the model. Be sure to label all parts of your model.
B. Use your model to explain what happens with the plate and what happens at the hot spot when a volcano forms.
C. Draw a model to show the side view (cross-section) of volcano formation near a plate boundary (at a subduction zone or divergent boundary). Be sure to label all parts of your model.
D. Use your model to explain what happens when a volcano forms near a plate boundary.
Characteristics of NGSS-Aligned Tasks

Include **multiple components** that reflect the connected use of different scientific practices in the context of interconnected disciplinary ideas and crosscutting concepts;

Address the **progressive nature of learning** by providing information about where students fall on a continuum between expected beginning and ending points in a given unit or grade; and

Include an **interpretive system** for evaluating a range of student products that are specific enough to be useful for helping teachers understand the range of student responses and provide tools for helping teachers decide on next steps in instruction.
The Assessment Coverage Challenge

• The NGSS describe specific goals for science learning in the form of *performance expectations*, statements about what students should know and be able to do at each grade level.

• Each performance expectation incorporates all three dimensions, and the NGSS emphasize the importance of the connections among scientific concepts.

• It will not be feasible to assess all of the performance expectations for a given grade level during a single assessment occasion.

• Students will need multiple—and varied—assessment opportunities to demonstrate their competence on the performance expectations for a given grade level.
Assessments in the Classroom

• Instruction that is aligned with the framework and NGSS will naturally provide many opportunities for teachers to observe and record evidence of students’ learning.

• Student activities that reflect such learning include
  – developing and refining models;
  – generating, discussing, and analyzing data;
  – engaging in both spoken and written explanations and argumentation;
  – reflecting on their own understanding.

• Such opportunities are the basis for the development of assessments of three-dimensional science learning.

• Report provides multiple examples of such assessments as they function in classroom teaching and learning.
Giving Precedence to Classroom Assessment
Example Task: *Biodiversity in the Schoolyard Zone*

- This example describes a cluster of three tasks that ask 5th grade students to determine which zone of their schoolyard contains the greatest biodiversity.

- The tasks require students to demonstrate knowledge of:
  - Disciplinary Core Idea -- biodiversity
  - Crosscutting Concept -- patterns
  - Practices – planning and carrying out investigations, analyzing and interpreting data, and constructing explanations.

- This is an example of formative assessment: Results from these tasks can help teachers spot strengths and weaknesses in students’ understanding and modify their instruction accordingly.
**Task 1:** Collect data on the number of animals (abundance) and the number of different species (richness) in schoolyard zones. The students are broken into three teams, and each team is assigned a zone in the schoolyard. The students are instructed to go outside and spend 40 minutes observing and recording all of the animals and signs of animals seen in their assigned zone. The students record their information, which is uploaded to a spreadsheet containing all the students’ combined data.

**Purpose:** Teachers can look at the data provided by individual groups or from the whole class to gauge how well students can perform the scientific practices of planning and carrying out investigations, and collecting and recording data.
Example Task: Biodiversity in the Schoolyard Zone

Task 2: Create bar graphs that illustrate patterns in data on abundance and richness from each of the schoolyard zones. Students are instructed to make two bar charts – one illustrating the abundance of species in the three zones, and another illustrating the richness of species in the zones – and to label the charts’ axes.

Purpose: This task allows the teacher to gauge students’ ability to construct and interpret graphs from data -- an important element of the scientific practice “analyzing and interpreting data.”
Task 3: Construct an explanation to support your answer to the question, “Which zone of the schoolyard has the greatest biodiversity?” Previously, students had learned that an area is considered biodiverse if it has both a high animal abundance and high species richness. In the instruction for this task, each student is prompted to make a claim, give his or her reasoning, and identify two pieces of evidence that support the claim.

Purpose: This task allows the teacher to see how well students understand the core idea of biodiversity and whether they can recognize data that reflects its hallmarks (high animal abundance and high species richness). It also reveals how well they can carry out the scientific practice of constructing explanations. This task could also be used as part of a “summative” end-of-unit assessment.
Assessments for Monitoring

- It is not feasible to cover the full breadth and depth of the NGSS performance expectations for a given grade level with a single external (large-scale) assessment.

- The types of assessment tasks that are needed take time to administer, and several will be required in order to adequately sample the set of performance expectations for a given grade level.

- Some practices, such as demonstrating proficiency in carrying out an investigation, will be difficult to assess using conventional formats of on-demand external assessments.
Assessments for Monitoring (cont.)

States will therefore need to rely on a combination of two types of external assessment strategies for monitoring purposes:

**On-Demand Assessments**
- Developed by the state
- Administered at a time mandated by the state

**Classroom-Embedded Assessments**
- Developed by the state or district,
- Administered at a time determined by the district/school that fits the instructional sequence in the classroom
Options for On-Demand Assessments

• Mixed item formats with written responses
  – Such as the AP Biology

• Mixed item formats with performance tasks
  – might involve both group and independent activities (NECAP example)
  – might involve some hands-on tasks, such as having students perform tasks at stations (NY example)

• Use matrix sampling, depending on the intended use and the need to report scores for individuals versus for groups.
Options for Classroom-Embedded Assessments

- **Types of assessments**
  - *Replacement units* (curriculum materials + assessments) developed outside of the classroom (by state or district)
  - *Item banks of tasks*, developed outside of the classroom
  - *Portfolio collections of work samples*, with tasks specified by state or district
Options for Classroom-Embedded Assessments

- Teachers administer them at a time that fits with the instructional sequence, possibly set by the school or district.

- Teachers receive training in how to administer them.

- Scoring can be done by teachers (trained to score them) or they can be sent to the district/state for scoring.

- Moderation and quality control procedures can be used to enhance the comparability of these assessments so they could support the desired inferences/comparisons needed for a monitoring purpose.
Issues Regarding Use of Performance Tasks

- Research will be needed to explore strategies for enhancing the comparability of results from performance tasks and portfolio assessments so that they yield results appropriate for the intended monitoring purpose.
- Appropriate use of such strategies will need to include acceptance of alternative concepts and varying degrees of comparability among assessments according to their usage.
- Research is needed on methods for statistically equating and/or linking scores and on methods for using moderation techniques. Such research should build on the existing literature base of prior and current efforts to enhance the comparability of scores for these types of assessments, including studies of approaches used in other countries.
Indicators of Opportunity to Learn

- Indicators would document variables such as:
  - time allocated to science teaching,
  - adoption of instructional materials that reflect the NGSS and framework’s goals,
  - classroom coverage of content and practice outlined in these documents.

- Such indicators would be a critical tool for monitoring the equity of students’ opportunities to learn.
A System Example

Components of Classroom Assessments

• State (or consortium of states) would develop collections of tasks aligned with the NGSS performance expectations for each grade or grade level.

• Teacher use the tasks in the classroom to support formative and summative assessment.

• Teachers would be trained to score these tasks
Components of the Monitoring Assessment

1. On-demand assessment -- mixed-item formats
   - Selected response, constructed response, performance tasks
   - Administered with combination of fixed form and matrix sampling
   - Fixed form would yield individual scores; matrix sample portion would yield school level scores.
   - Common test given as a fixed form would contain selected responses and constructed-response questions
   - Matrix sampled questions would include performance based tasks
2. Classroom-Embedded Assessment --replacement units designed by the state

- Given at a time determined by the district or LEA

- District or LEA would select from options for topics to be covered in the units

- Scored by the state, possibly using teachers
System Example (cont.)

**Indicators of Opportunity to Learn**

- State collects data to document that:
  - Teachers have access to professional development and quality curricular materials and administrative supports
  - Teachers are implementing instruction and assessment in ways that align with the framework and NGSS
  - All students have access to appropriate materials and resources.

- These indicators would be used for accountability purposes, along with other data.
Implementation: Bottom Up Approach

- The committee encourages a developmental path for assessment that is “bottom up” rather than “top down”: one that begins with the process of designing assessments for the classroom, perhaps integrated into instructional units, and moves toward assessments for monitoring.

- In designing and implementing their assessment systems, states will need to focus on professional development.

- States will need to include adequate time and resources for professional development so that teachers can be properly prepared and guided and so that curriculum and assessment developers can adapt their work to the vision of the framework and the NGSS.
Gradual, Prioritized Implementation

- The assessment system that the committee recommends differs markedly from current practice and will thus take time to implement, just as it will take time to adopt the instructional programs needed for students to learn science in the way envisioned in the framework and the NGSS.

- States should develop and implement new assessment systems gradually and establish carefully considered priorities. Those priorities should begin with what is both necessary and possible in the short term while also establishing long-term goals to implementation of a fully integrated and coherent system of curriculum, instruction, and assessment.
State and district leaders who commission assessment development should ensure that the plans address the changes called for by the framework and the NGSS.

They should build into their commissions adequate provision for the substantial amounts of time, effort and refinement that are needed to develop and implement the use of such assessments: multiple cycles of design-based research will be necessary.
Professional Development is Critical to Success

It is critically important that states include adequate time and material resources in their plans for professional development to properly prepare and guide teachers, curriculum and assessment developers, and others in adapting their work to the vision of the framework and the Next Generation Science Standards.
Some Challenges for Professional Development

- Practices may be unfamiliar to teachers
- Knowledge of crosscutting concepts and some core ideas may be incomplete for some teachers
- Thinking about learning progressions within and across grades
- Some teachers will need to make major changes in instruction & assessment approach
- Making connections across disciplines and to mathematics and ELA
Foci of Today’s Presentation

• Science Education: Past, Present, Future?
• Defining Competence to Achieve Coherence
  – Unpacking the Components of Competence
• From NGSS Performance Expectations to Coherence in K-12 Science Education
• NRC Report Conclusions & Recommendations
• Moving Forward: Prospects & Issues
Attention to Equity and Fairness

• The framework and the NGSS both stress that equity in science learning can only happen if all students have the opportunity to learn in the new ways called for and if science educators are trained to work with multiple dimensions of diversity.
  – A good assessment system can play a critical role in providing fair and accurate measures of the learning of all students and providing students with multiple ways of demonstrating their competency.

• Such an assessment system will include formats and presentation of tasks and scoring procedures that reflect multiple dimensions of diversity, including culture, language, ethnicity, gender, and disability.

• Individuals with expertise in diversity should be integral participants in developing state assessment systems.
Cost and Feasibility Considerations

- Developing the tasks that we advocate may be significantly more resource intensive than design and development of traditional assessment tasks (such as tests composed of multiple-choice items), particularly in the early phases.

However:

- Adopting new systems gradually and strategically, in phases, will be a key to managing costs.
- New and existing technologies offer possibilities for achieving assessment goals at costs lower than for other assessments including performance tasks.
- We recommend administering the monitoring assessments less frequently than is currently done in many states in other subjects.
- Consider state consortia or other types of collaborations.
Cost and Feasibility Considerations (cont.)

- An important advantage of the approach recommended is that many assessment-related activities—such as task development and scoring moderation sessions in which teachers collaborate—will have benefits beyond their assessment function.

- Much of what we recommend for classroom assessment will be integral to curriculum planning and professional development and thus it is both a shared cost and a shared resource with instruction.

- Although the combination of classroom-based and monitoring assessments we propose may take longer to administer in the classroom, it will also be a benefit in terms of usefulness for instruction.
States should support the use of existing and emerging technologies in designing and implementing a science assessment system that meets the goals of the framework and the Next Generation Science Standards.

New technologies hold particular promise for supporting the assessment of three-dimensional science learning, and for streamlining the processes of assessment administration, scoring, and reporting.
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.
Will We Have Assessments Worth Teaching With & To?

- Desires of the policy community often conflict with the capacities of the R&D & practice communities
  - Need for better coordination and communication
    - USDoe, States, IES & NSF, R&D Community, Teachers, Administrators, & Professional Education Groups

- **Standards are the beginning not the end** – not a substitute for the thinking and research needed to define **progressions of learning** that can serve as a basis for the integration of **curriculum, instruction and assessment**.
Assessment Should not be the “Tail Wagging the Science Education Dog”
The perfect is the enemy of the good.

Voltaire
Questions & Comments