Measuring What Matters: Opportunities & Challenges in the Assessment of Science Proficiency

James W. Pellegrino

UIC LEARNING SCIENCES RESEARCH INSTITUTE
Measuring What Matters: Proficiency in Science

• About Assessment & Assessment Design

• Next Generation Science Learning & Next Generation Science Assessment

• Grand Challenges in Science Assessment
Measuring What Matters: Proficiency in Science

• About Assessment & Assessment Design
Knowing what Students Know: The Science and Design of Educational Assessment
Assessment as Part of a Larger Coordinated System

Assessment

Curriculum

Instruction
Contexts and Purposes: Distinctions Making a Difference

- **Contexts:**
  - small scale: individual classrooms
  - intermediate scale: schools, districts
  - large scale: states, nations, international

- **Purposes:**
  - assist learning (formative)
  - measure individual achievement (summative)
  - evaluate programs (accountability)

- **Problem:** One size does not fit all
  - Educators at different levels need different information at different grain sizes and time scales
  - Differing priorities, constraints, & design tradeoffs
Assessment as a Process of Reasoning from Evidence

- **cognition**
  - model of how students represent knowledge & develop competence in the domain

- **observation**
  - tasks or situations that allow one to observe students’ performance

- **interpretation**
  - method for making sense of the data

*Must be coordinated!*
Scientific Foundations of Educational Assessment

• Advances in the Sciences of Thinking and Learning -- the cognition vertex
  – informs us about what observations are important and sensible to make

• Contributions of Measurement and Statistical Modeling -- the interpretation vertex
  – Informs us about how to make sense of the observations we have made
How People Learn

Brain, Mind, Experience, and School

How Students Learn

HISTORY, MATHEMATICS, AND SCIENCE IN THE CLASSROOM
Why Models of Development of Domain Knowledge are Critical

• Tell us what are the important aspects of knowledge that we should be assessing.
• Give us strong clues as to how such knowledge can be assessed
• Can lead to assessments that yield more instructionally useful information
  – diagnostic & prescriptive
• Can guide the development of systems of assessments
  – work across contexts & time
A common problem is that the assessments we develop for use in education often lack validity relative to the constructs of interest and the desired interpretive uses.

We need to make our assumptions and methods of assessment design more **EXPLICIT** -- rather than believe in miracles.

“I think you should be more explicit here in Step Two.”
Assessment design spaces vary tremendously & involve multiple dimensions

- Type of knowledge and skill and levels of sophistication
- Time period over which knowledge is acquired
- Intended use and users of the information
- Availability of detailed theories & data
- Distance from instruction and assessment purpose

Need a principled set of processes that can help structure going from theory, data and/or speculation to an operational assessment
Unpacking the domain

Conceptual Assessment Framework

Assessment Implementation

Assessment Delivery

- What is important about this domain?
- What work and situations are central in this domain?
- What knowledge & skills are central to this domain?

• How do we represent key aspects of the domain in terms of an “assessment argument.”
• Assessment design structures: ECD -- domain model (claims), evidence, and task models

• How do we choose and present tasks
• How do we gather and analyze responses?

• How do students and tasks actually interact?
• How do we report examinee performance?
Measuring What Matters: Proficiency in Science

• About Assessment & Assessment Design

• Next Generation Science Learning & Next Generation Science Assessment
A bit of the historical context for this discussion
The Vision of Science Learning

Knowing how to use and apply what you know...

empowers you – in your own learning about the world and your participation in it.

Goal is for every student, from the earliest grades onward, to have coherent and sequenced instruction that provides opportunities to do the “walk and talk” of science and engineering.
What is Really New in the NRC Framework & NGSS?

1. Focus on explaining phenomena or designing solutions to problems

2. Three–Dimensional learning
   - Organized around disciplinary core ideas
   - Use of crosscutting concepts
   - Central role of scientific and engineering practices

3. Standards expressed as performance statements that integrate the 3 dimensions

4. Coherence: building and applying ideas over time and across disciplines

5. Focus on all learners
What is Three-Dimensional Learning?

Three-dimensional learning shifts the focus of the science classroom...

...to where students use disciplinary core ideas, crosscutting concepts with scientific practices to explore, examine, and explain how and why phenomena occur and to design solutions to problems.
How the New Standards are Different

Standards expressed as **performance expectations**:

- Combine practices, core ideas, and crosscutting concepts into a single statement of *what is to be assessed*

- Requires students to demonstrate *knowledge-in-use*

- Performance Expectations are not instructional strategies or objectives for a lesson – *they describe achievement, not instruction*

- Intended to describe the end-goals of instruction – *the student performance at the conclusion of instruction*
**An NGSS performance expectation**

**MS-PS1 Matter and its Interactions**

Students who demonstrate understanding can:

**MS-PS1-2.** Analyze and interpret data on the properties of substances before and after the substances interact to determine if a chemical reaction has occurred. [Clarification Statement: Examples of reactions could include burning sugar or steel wool, fat reacting with sodium hydroxide, and mixing zinc with hydrogen chloride.][Assessment boundary: Assessment is limited to analysis of the following properties: density, melting point, boiling point, solubility, flammability, and odor.]

The performance expectation above was developed using the following elements from the NRC document *A Framework for K-12 Science Education:*

### Science and Engineering Practices

- **Analyzing and Interpreting Data**
  - Analyzing data in 6–8 builds on K–5 and progresses to extending quantitative analysis to investigations, distinguishing between correlation and causation, and basic statistical techniques of data and error analysis.
  - Analyze and interpret data to determine similarities and differences in findings.

### Disciplinary Core Ideas

#### PS1.A: Structure and Properties of Matter
- Each pure substance has characteristic physical and chemical properties (for any bulk quantity under given conditions) that can be used to identify it.

#### PS1.B: Chemical Reactions
- Substances react chemically in characteristic ways. In a chemical process, the atoms that make up the original substances are regrouped into different molecules, and these new substances have different properties from those of the reactants.

### Crosscutting Concepts

- **Patterns**
  - Macroscopic patterns are related to the nature of microscopic and atomic-level structure.
The challenge:

How can we create assessments that integrate the three dimensions of the NGSS and help teachers assess student’s progress toward achieving performance expectations?
NGSA Project Aims

A. Articulate a principled design approach for constructing classroom-based assessments that align to NGSS

B. Use the approach to develop and test technology-based assessment tasks and rubrics (middle school physical & life science)

C. Engage in a co-design process with science teachers to develop guidelines and strategies for classroom use
Overview of NGSA Design Approach

Task Design Process Schematic

1. Identify Target Performance Expectations
   a. Unpack Disciplinary Core Ideas
   b. Unpack Practices
   c. Unpack Crosscutting Concepts

2. Domain Analysis
   - Create Integrated Dimension Map

3. Domain Modeling
   - Articulate Learning Performances

4. Specify Task Design Patterns
   - Determine KSAs & Evidence Statement
   - Determine Task Design Features
   - Apply Fairness/Equity Framework

5. Develop Tasks and Rubrics
   - Technology Environment Affordances
MS-PS1-4. Develop a model that predicts and describes changes in particle motion, temperature, and state of a pure substance when thermal energy is added or removed.
## Unpacking disciplinary core ideas

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Example (Energy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elaboration of DCI</td>
<td>All particles of matter have kinetic energy because they are in motion. Thermal energy transferred to a substance causes particles to move faster.</td>
</tr>
<tr>
<td>Proficiency boundaries</td>
<td>Student do not need to understand “heat” as thermal energy transfer.</td>
</tr>
<tr>
<td>Prior knowledge</td>
<td>Grade 5: Matter is made of particles too small to be seen.</td>
</tr>
<tr>
<td>Student challenges</td>
<td>Belief that only things that are warm or hot have thermal energy.</td>
</tr>
</tbody>
</table>
## Unpacking science practices

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Example (Develop models)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essential knowledge and skills</td>
<td><strong>Specify</strong> model elements/relationships and <strong>describe</strong> why they are needed</td>
</tr>
<tr>
<td>Evidence of a high level of performance</td>
<td><strong>Specifies only</strong> the <strong>appropriate</strong> and <strong>necessary</strong> elements/relationships in the model</td>
</tr>
<tr>
<td>Relationships to other practices</td>
<td><strong>Models illustrate mechanisms, informing the construction of scientific explanations</strong></td>
</tr>
<tr>
<td>Prerequisite knowledge and skill</td>
<td>Knowledge that models help <strong>explain</strong> or <strong>predict</strong> phenomena</td>
</tr>
</tbody>
</table>
Overview of NGSA Design Approach

Task Design Process Schematic

1. Identify Target Performance Expectations
   - a. Unpack Disciplinary Core Ideas
   - b. Unpack Practices
   - c. Unpack Crosscutting Concepts

2. Domain Analysis
   - 2a. Unpack Disciplinary Core Ideas
   - 2b. Unpack Practices
   - 2c. Unpack Crosscutting Concepts

3. Create Integrated Dimension Map

4. Articulate Learning Performances

5. Domain Modeling
   - 5a. Determine KSAs & Evidence Statement
   - 5b. Determine Task Design Features
   - 5c. Apply Fairness/Equity Framework

6. Specify Task Design Patterns
   - 6a. Develop Tasks and Rubrics
   - 6b. Technology Environment Affordances
MS-PS1-4. Develop a model that predicts and describes changes in particle motion, temperature, and state of a pure substance when thermal energy is added or removed.

DCI Concept Map

- Energy transfer
  - Directly proportional

- Particle motion change
  - Can cause change in
  - States
    - Differ in their
      - Particle spacing & change in location
      - Temperature change
    - Measures average
MS-PS1-4. Develop a model that predicts and describes changes in particle motion, temperature, and state of a pure substance when thermal energy is added or removed.

DCI Concept Map

SEPs & CCCs added

Energy transfer
- Directly proportional
- SP: Models
- CC: C&E, SPQ, E&M

Particle motion change
- Can cause change in
  - SP: Models
  - CC: C&E, Patterns
- Measures average
  - SP: Models; Explanation
  - CC: C&E

States
- Differ in their
  - SP: Models
  - CC: Patterns

Particle spacing & change in location

Temperature change
How do we **Assess** *toward* the PEs?

**Assess *toward* Performance Expectations**
Overview of NGSA Design Approach

Task Design Process Schematic

1. Identify Target Performance Expectations
   - a. Unpack Disciplinary Core Ideas
   - b. Unpack Practices
   - c. Unpack Crosscutting Concepts

2. Domain Analysis
   - Create Integrated Dimension Map

3. Domain Modeling
   - Articulate Learning Performances

5. Specify Task Design Patterns
   - Determine KSAs & Evidence Statement
   - Determine Task Design Features
   - Apply Fairness/Equity Framework

6. Develop Tasks and Rubrics
   - Technology Environment Affordances
Qualities of “Good” Learning Performances

Each **Learning Performance** separately:

- Blends disciplinary core ideas, scientific practices and crosscutting concepts
- Helps to identify an important opportunity that teachers should attend to and assess *before* the end of a unit
- Is assessable in a 5-10 minute task

Collectively the set of all **learning performances**:

- Identify “what it takes” to make progress toward meeting NGSS performance expectations
Example: Learning Performances building towards a Specific PE

MS-PS1-4. Develop a model that predicts and describes changes in particle motion, temperature, and state of a pure substance when thermal energy is added or removed.

- **LP E-01:** Students evaluate a model that uses a particle view of matter to explain how states of matter are similar to and/or different from each other.

- **LP E-02:** Students develop a model that explains how particle motion changes when thermal energy is transferred to or from a substance without changing state.

- **LP E-03:** Students develop a model to explain the change in the state of a substance caused by transferring thermal energy to or from a sample.
Overview of NGSA Design Approach

Task Design Process Schematic

1. Identify Target Performance Expectations
   - a. Unpack Disciplinary Core Ideas
   - b. Unpack Practices
   - c. Unpack Crosscutting Concepts

2. Domain Analysis
   - Create Integrated Dimension Map

3. Domain Modeling
   - Articulate Learning Performances

4. Specify Task Design Patterns
   - Determine KSAs & Evidence Statement
   - Determine Task Design Features
   - Apply Fairness/Equity Framework

5. Develop Tasks and Rubrics
   - Technology Environment Affordances
Assessment as an Argument from Evidence: 3 Integrated Questions

- What do we want students to know and be able to do? (Described by our learning performances)
- What kinds of evidence will students need to provide to demonstrate proficiency?
- What kinds of tasks / task features will elicit the desired evidence?

When we have logical and coherent answers to these three questions, we have an assessment argument.
Overview of NGSA Design Approach

Task Design Process Schematic

1. Identify Target Performance Expectations
   - a. Unpack Disciplinary Core Ideas
   - b. Unpack Practices
   - c. Unpack Crosscutting Concepts

2. Domain Analysis
   - Create Integrated Dimension Map

3. Domain Modeling
   - Articulate Learning Performances

5. Specify Task Design Patterns
   - Determine KSAs & Evidence Statement
   - Determine Task Design Features
   - Apply Fairness/Equity Framework

6. Develop Tasks and Rubrics
   - Technology Environment Affordances
Tasks are built to align to individual LPs

• Claims in the Learning Performances, and their associated evidence statements, are used to identify task characteristics

• Assessment tasks are designed using characteristic and variable task features

• Exemplar responses written for each task, checked against the evidence statement

• One LP will have multiple tasks that could be designed – can vary systematically relative to level of challenge
Shawn had 3 dishes of water at room temperature. She cooled one dish, causing thermal energy to transfer from that dish to the surroundings. She kept the middle dish at room temperature. She transferred thermal energy into the third dish by heating it. Then, Shawn dropped a red-coated chocolate candy into each dish. Watch what happened using the video.
Nami wondered if mixing an acid with an alcohol would cause a chemical reaction. She did the following experiment:

<table>
<thead>
<tr>
<th>Step 1</th>
<th>Step 2</th>
<th>Step 3</th>
<th>Step 4</th>
<th>Step 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Acid" /> <img src="image2" alt="Alcohol" /></td>
<td><img src="image3" alt="Mixed" /></td>
<td><img src="image4" alt="Heated" /></td>
<td><img src="image5" alt="Layer 1" /> <img src="image6" alt="Layer 2" /></td>
<td><img src="image7" alt="Layer 1" /> <img src="image8" alt="Layer 2" /></td>
</tr>
</tbody>
</table>

- Measured and tested properties of acid and alcohol at room temperature. Recorded data in Table 1.
- Mixed the acid and alcohol in a test tube.
- Heated the test tube with the mixture.
- After heating, observed Layer 1 and Layer 2 form.
- Measured and tested properties of substance in Layer 1 and Layer 2 at room temperature. Recorded data in Table 1.

She measured the boiling point and mass, and calculated the density of the substance, then recorded the data in Table 1.

### Question #1

Based on Nami’s results in Table 1, write a scientific explanation about whether this experiment had a chemical reaction or not. In your scientific explanation make sure that you:

1. Write a claim stating whether the acid and alcohol chemically reacted.
2. Include evidence to support your claim.
3. Give reason(s) why the evidence you included supports your claim.

### Table 1. Data for liquids before and after the experiment

<table>
<thead>
<tr>
<th>Substance</th>
<th>Density</th>
<th>Boiling Point</th>
<th>Mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acid</td>
<td>1.2 g/cm³</td>
<td>100 °C</td>
<td>6.9 g</td>
</tr>
<tr>
<td>Alcohol</td>
<td>0.80 g/cm³</td>
<td>98 °C</td>
<td>9.0 g</td>
</tr>
<tr>
<td>Layer 1</td>
<td>0.91 g/cm³</td>
<td>82 °C</td>
<td>13.2 g</td>
</tr>
<tr>
<td>Layer 2</td>
<td>1.0 g/cm³</td>
<td>100 °C</td>
<td>2.7 g</td>
</tr>
</tbody>
</table>
For a class project, Jaden’s science teacher asked him to develop a model to show how energy flows through a natural system that involves a consumer. Jaden chose to use the koala as the consumer in his model. Koalas live in eucalyptus trees and eat mainly eucalyptus leaves. Jaden’s model is shown to the right.

**Question #1**

Describe 2 parts of Jaden’s model that show you how energy flows through the system.

Type answer here

- The Sun transfers energy directly to the koala. This helps the koala break down food.
- Koala eats eucalyptus leaves for food. Digesting food results in transferring energy around Koala’s body to be used for generating heat.
- Energy from the breaking down of food is taken to muscle.
- While resting, energy from food eaten earlier is used to power important reactions in the body.
- Koala must use energy for motion.

**Question #2**

What feedback would you give to Jaden to help him improve his model? Take a snapshot of the model and circle 2 parts that need improvement.

Then use the text box to describe improvements you would make to the parts of the model you circled.
How can science educators effectively support the integrated 3-dimensional learning called for by the NGSS?

A big challenge facing teachers who are shifting instruction to meet the vision of the Framework for K-12 Science Education and the Next Generation Science Standards (NGSS) is how to support students' progress toward achieving the new standards.

The Next Generation Science Assessment (NGSA) group is a multi-institutional collaborative that is applying the evidence-centered design approach to create classroom-ready assessments for teachers to use formatively to gain insights into their students' progress on achieving the NGSS performance expectations.

We are a high-caliber interdisciplinary team with expertise in:

- science disciplinary knowledge and practice,
- science teaching and learning,
- classroom-based assessment,
- technology-enhanced instruction and assessment, and
- K-12 professional development.
Measuring What Matters: Proficiency in Science

- About Assessment & Assessment Design
- Next Generation Science Learning & Next Generation Science Assessment
- Grand Challenges in Science Assessment
Proficiency in science is being defined through performance expectations that intertwine science practices, cross-cutting concepts, and core content knowledge. These descriptions of what it means to know and do science pose challenges for assessment design and use, whether at the classroom instructional level or the system level for monitoring the progress of science education. There are systematic ways to approach assessment development that can address design challenges, as well as examples of the application of such principles in science assessment. This Review considers challenges and opportunities that exist for design and use of assessments that can support science teaching and learning consistent with a contemporary view of what it means to be proficient in science.
Where Do We Stand in Meeting these Challenges?

• We have increasing awareness of what the development of competence means and the implications for designing coherent science education.

• We have examples of thinking through in detail the juxtaposition of *disciplinary practices* and *core content knowledge* to guide the design of curriculum, instruction, and assessment.

  – *Designing Assessments for Middle School Physical & Life Science*
  – *Redesign of the College Board’s Advanced Placement Science Courses & Assessments*
Assessment Should not be the “Tail Wagging the Science Education Dog”
Thank You!!!