Measuring What Matters: Defining & Assessing Science Proficiency

Jim Pellegrino
Chinese Curse?

• There is a Chinese curse which says 'May he live in interesting times.' Like it or not we live in interesting times. They are times of danger and uncertainty; but they are also more open to the creative energy of men than any other time in history.

Components of Vision and Change Across K-16+

- NRC’s *Science Education Framework & Next Generation Science Standards*
- College Board Redesign of *AP Curriculum Frameworks and Exams for Biology, Chemistry & Physics*
- AAMC’s *Competency Framework & Redesign of the Medical College Admission Test (2015)*
A FRAMEWORK FOR K-12 SCIENCE EDUCATION
Practices, Crosscutting Concepts, and Core Ideas
NATIONAL RESEARCH COUNCIL OF THE NATIONAL ACADEMIES

NEXT GENERATION SCIENCE STANDARDS

PRACTICES
CONTENT
CROSSCUTTING
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
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.
### SCIENCE EDUCATION WILL INVOLVE LESS:

| Rote memorization of facts and terminology |
| Learning of ideas disconnected from questions about phenomena |
| Teachers providing information to the whole class |
| Teachers posing questions with only one right answer |

### SCIENCE EDUCATION WILL INVOLVE MORE:

| Facts and terminology learned as needed while developing explanations and designing solutions supported by evidence-based arguments and reasoning. |
| Systems thinking and modeling to explain phenomena and to give a context for the ideas to be learned |
| Students conducting investigations, solving problems, and engaging in discussions with teachers’ guidance |
| Students discussing open-ended questions that focus on the strength of the evidence used to generate claims |

Assessment as Part of a Larger Coordinated System

Diagram:

- Assessment
- Curriculum
- Instruction

Arrows indicate the interconnection between assessment, curriculum, and instruction.
Assessment is a Process of Reasoning from Evidence

• cognition
  – Theories, models & data on how students represent knowledge & develop competence in the domain

• observations
  – tasks or situations that allow one to observe students’ performance

• interpretation
  – method for making sense of the data

*Must be coordinated!*
<table>
<thead>
<tr>
<th>Grand Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Where Do We Stand in Meeting these Challenges?

• We have increasing awareness of what the development of competence should mean and the possible 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

  – **AP Science Redesign Project (NSF)**
  – **Designing Assessments for Middle School Physical & Life Science (NSF & Moore Fdn)**
Why an AP Science Redesign?

• A 2002 NRC Report identified ways to improve advanced study of math and science in the U.S. The Report’s recommendations are applicable to all AP course subjects:
  – Emphasize deep understanding rather than comprehensive coverage -- avoid “mile wide & inch deep” syndrome
  – Reflect current understanding of how students learn in a discipline
  – Reflect current research directions within the disciplines
  – Emphasize the development of inquiry and reasoning skills
Conceptual Approach Built Upon Work of Others:


Collaboration with Jeanne Pemberton, Mark Reckase, Meryl Bertenthal, John Eggebrecht, Kristen Huff, Cindy Hamen, Marcia Wilbur et al.

Supported by NSF and the College Board
# Components & Timeline

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Process Design:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Based on how people learn, identified shortcomings, and discipline expertise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Establish Foundations:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disciplinary experts identify essential concepts and reasoning skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model of Knowing &amp; Learning:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Framework for Curriculum and Assessment are claims and evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Assessment Design:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tasks derived from claims and evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exam score descriptions from claims and evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Professional Development:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provide AP teachers with the curriculum resources &amp; training needed to teach the redesigned courses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Curriculum Design:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formative assessments and interpretive framework Course Description</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Involving AP and Professional Community:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review approach with and solicit feedback from instructors (secondary and post-secondary) and scientific communities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>OPERATIONS:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deliver the redesigned course &amp; exam to students, and incorporate into ongoing operational processes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Application of ECD: Three Critical Design Phases

- Domain Analysis
  - Content
  - Skills

- Domain Model
  - Claims
  - Evidence
  - ALDs

- Assessment Framework
  - Task models
  - Form assembly specifications

Increasing specificity
1. Use representations and models to communicate scientific phenomena and solve scientific problems.

2. Use mathematics appropriately.

3. Engage in scientific questioning to extend thinking or to guide investigations within the context of the AP course.

4. Plan and implement data collection strategies in relation to a particular scientific question.

5. Perform data analysis and evaluation of evidence.


7. Connect and relate knowledge across various scales, concepts, and representations in and across domains.
Most scientists regarded the new streamlined peer-review process as ‘quite an improvement.’
The unifying concepts or Big Ideas increase coherence both within and across disciplines. A total of Four Big Ideas:

<table>
<thead>
<tr>
<th>BIG IDEA</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The process of evolution drives the diversity and unity of life.</td>
</tr>
<tr>
<td>2</td>
<td>Biological systems utilize energy and molecular building blocks to grow, reproduce, and maintain homeostasis.</td>
</tr>
<tr>
<td>3</td>
<td>Living systems retrieve, transmit, and respond to information essential to life processes.</td>
</tr>
<tr>
<td>4</td>
<td>Biological systems interact, and these interactions possess complex properties.</td>
</tr>
</tbody>
</table>
Biology: “Enduring Understandings”

For each Big Idea, there are enduring understandings which incorporate core concepts that students should retain. Total of 17 enduring understandings across the four Big Ideas.

**BIG IDEA 1**

The process of evolution drives the diversity and unity of life.

- **Enduring Understanding 1.A:** Change in the genetic makeup of a population over time is evolution
- **Enduring Understanding 1.B:** Organisms are linked by lines of descent from common ancestry
- **Enduring Understanding 1.C:** Life continues to evolve within a changing environment
- **Enduring Understanding 1.D:** The origin of living systems is explained by natural processes
1. Use representations and models to communicate scientific phenomena and solve scientific problems.

2. Use mathematics appropriately.

3. Engage in scientific questioning to extend thinking or to guide investigations within the context of the AP course.

4. Plan and implement data collection strategies in relation to a particular scientific question.

5. Perform data analysis and evaluation of evidence.


7. Connect and relate knowledge across various scales, concepts, and representations in and across domains.
Level 1: work with scientific explanations & theories

Level 2:
1. justify claims with evidence
2. construct explanations of phenomena based on evidence produced through scientific practices
3. articulate the reasons that scientific explanations and theories are refined or replaced
4. make claims and predictions about natural phenomena based on scientific theories & models.
5. evaluate alternative scientific explanations
Level 1: work with scientific explanations and theories
Level 2: justify claims with evidence
Level 3: Description of what observable data are needed to support the claim that a student can achieve the Level 2
Level 3: *Examples of Evidence Found in Students’ Work*

- Robustness of evidence (from investigations, theories, or models) mustered in support of claim
- Appropriateness of reasoning behind selection and exclusion of evidence
- Appropriateness of model incorporated
- Consideration of data from multiple sources (e.g., investigations, scientific observations, the findings of others, historic reconstruction, and/or archived data historical experiments)
- Differentiation between a claim and the evidence that supports it
- Differentiation between evidence and explanation
- Inclusion and reasonableness of a statement of prediction or existence of a phenomena
Three Critical Design Phases

- **Domain Analysis**
  - Content
  - Skills

- **Domain Model**
  - Claims
  - Evidence
  - ALDs

- **Assessment Framework**
  - Task models
  - Form assembly specifications

Increasing specificity
### AP Integrating the Content and Science Practices

<table>
<thead>
<tr>
<th>Content</th>
<th>Science Practice 5.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essential Knowledge 1.B.2</td>
<td>The student connects phenomena and models across spatial and temporal scales</td>
</tr>
<tr>
<td>Phylogenetic trees and cladograms are graphical representations (models) of evolutionary history that can be tested</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Science + Practice + Learning Objective (1.B.2 &amp; 5.3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning Objective</td>
</tr>
<tr>
<td>The student is able to evaluate evidence provided by a data set in conjunction with a phylogenetic tree or a simple cladogram to determine evolutionary history and speciation</td>
</tr>
</tbody>
</table>
Connecting the Domain Model to Curriculum, Instruction, & Assessment

Domain Model

- Claims and supporting evidence
- Descriptions of Achievement levels

Assessment Framework

Summative Assessment

Curriculum Framework

Instructional Resources
Structure of the AP Biology Curriculum Framework

- 4 Big Ideas
- Enduring Understandings
- Essential Knowledge
- Science Practices: Science Inquiry & Reasoning
- Learning Objectives
Connecting the Domain Model to Curriculum, Instruction, & Assessment

- Domain Model
  - Claims and supporting evidence
  - Descriptions of Achievement levels

- Assessment Framework

- Summative Assessment

- Curriculum Framework

- Instructional Resources
The New AP Biology Exam

No test items focus on low levels of cognitive demand -- declarative knowledge/recall

For each exam item, students either produce the evidence (CR) or engage with the evidence (SR/MC)

► explain
► justify
► predict
► evaluate
► describe
► analyze
► pose scientific questions
► construct explanations
► construct models
► represent graphically
► solve problems
► select and apply mathematical routines
What’s The Impact Of Curriculum Changes On New AP Biology Exam?

*Because of use of Big Ideas*….in 2008, 12% of questions had something to do with evolution
In 2013 exam, 35% of questions had something to do with evolution

*Because of emphasis on science practice and mathematical skills*…new types of questions are being asked, e.g., grid-ins

*Because of use of evidence*…the number of Multiple Choice questions was reduced from 100 questions on 2012 exam to 63 on the 2013 exam.
Lessons Learned from the AP Redesign Project

• **No Pain -- No Gain!!!** -- this is hard work

• **Backwards Design** and **Evidence Centered Design** are challenging to execute & sustain
  – Requires multidisciplinary teams
  – Requires sustained effort and negotiation
  – Requires time, money & patience

• **Value-added** -- Validity is “designed in” from the start as opposed to “grafted on”
  – Elements of a validity argument are contained in the process and the products
A FRAMEWORK FOR K-12 SCIENCE EDUCATION
Practices, Crosscutting Concepts, and Core Ideas
NATIONAL RESEARCH COUNCIL OF THE NATIONAL ACADEMIES

NEXT GENERATION SCIENCE STANDARDS

PRactices

CONTENT

CROSScutting
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.]
Pluses & Minuses of Relying on NGSS 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
Committee on the Assessment of K-12 Science Proficiency

Board on Testing and Assessment
and
Board on Science Education

National Academy of Sciences
NRC Assessment Report: Give Precedence to Classroom Assessment
Next Generation Science Assessment

Developing Assessments Aligned to the NGSS: Middle School Science

NSF

Gordon and Betty Moore Foundation

UIC Learning Sciences Research Institute

SRI Education

The Concord Consortium

CREATE for STEM Institute
Projects’ Overall Goals

(1) Use a systematic design approach to guide the development of tasks aligned with the performance expectations in the NGSS,

(2) Develop and test technology-based assessment items and rubrics related to these performance expectations, and

(3) Develop guidelines and materials for teachers to use these assessments in the classroom for diagnostic and formative purposes.
Our Design Approach

Three distinct phases:

- unpacking
- assessment
- creating tasks and rubrics
Our Design Approach

Phase Two: Assessment Argument

1. Identify a Cluster of PEs
   - 2a. Unpack Science Practices
   - 2b. Unpack Disciplinary Core Ideas
   - 2c. Unpack Crosscutting Concepts

3. Integrated Concept Maps
   - 4a. Learning Performances (LPs)
     - 4b. Evidence Statements for Each LP
     - 4c. Task Design Features

5. Tasks, rubrics, and resources for teachers

Instructional Setting

Task Authoring Environment and Delivery Requirements
Our Assessment Argument Structure

- What claims do we want to be able to make about what students know and can do?
- 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.
Learning Performances

Why use Learning Performances?
- Specifies “knowledge in use”
- We conceptualize understanding as embedded in practice and not as memorizing static facts

What are Learning performances?
- Defined using cognitive terms for what it means for learners to “understand” a particular idea
- Define how the knowledge is used in reasoning about scientific questions and phenomena
- Combine practices and content ideas at a smaller grain size than the PE
An Example Item

Steven found four different bottles filled with unknown pure liquids. He measured the properties of each liquid. The measurements are displayed in the data table below. Steven wonders if any of the liquids are the same substance.

<table>
<thead>
<tr>
<th>Liquid</th>
<th>Density</th>
<th>Color</th>
<th>Volume</th>
<th>Boiling Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0 g/cm³</td>
<td>Clear</td>
<td>6.1 cm³</td>
<td>100 °C</td>
</tr>
<tr>
<td>2</td>
<td>0.89 g/cm³</td>
<td>Clear</td>
<td>6.1 cm³</td>
<td>211 °C</td>
</tr>
<tr>
<td>3</td>
<td>0.92 g/cm³</td>
<td>Clear</td>
<td>10.2 cm³</td>
<td>298 °C</td>
</tr>
<tr>
<td>4</td>
<td>0.89 g/cm³</td>
<td>Clear</td>
<td>10.2 cm³</td>
<td>211 °C</td>
</tr>
</tbody>
</table>

Use the data in the table to:
1) Write a claim stating whether any of the liquids are the same substance.
2) Provide at least two pieces of evidence to support your claim.
3) Provide reason(s) that justify why the evidence supports your claim.

Variable Task Features
- Number of properties included as data/evidence – 2 (density and boiling point)
- State of matter of substances – all liquids
- Inclusion of irrelevant data – yes
- Level of scaffolding to develop claim, evidence, and reasoning – yes
How can science educators effectively support the integrated 3-dimensional learning called for by the Next Generation Science Standards?

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.

Our current work:
The Next Generation Science Assessment team is currently developing technology-enhanced assessment tasks, rubrics, and accompanying instructional resources in physical science and life science for middle school classrooms. Our ongoing work aims to create exemplary tasks that can be used during instruction to garner good evidence that students are building proficiency with the NGSS performance expectations. We currently have a robust set of tasks available in physical science and expect to have our first draft of life science items ready to showcase in late spring 2016. Our tasks can be viewed through the NGSA portal.

Contact us!
What’s Left to Do? – A LOT!!!

• We need to translate the K-12 standards into effective models, methods and materials for curriculum, instruction, and assessment.
  – Need to unpack & clarify performance expectations
  – Need precise claims & evidence statements
  – Need task models & templates

• We need to use what we know already to evaluate and improve the assessments that are part of current practice, e.g., classroom assessments, large-scale exams, etc.
Will We Have Assessments Worth Teaching With & To?

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

- Frameworks & 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.
Changes in STEM in Higher Education

SCIENCE EDUCATION

Challenge faculty to transform STEM learning

Focus on core ideas, crosscutting concepts, and scientific practices


Models for higher education in science, technology, engineering, and mathematics (STEM) are under pressure around the world. Although most STEM faculty and practicing scientists have learned successfully in a traditional format, they are the exception, not the norm, in their success. Education should support a diverse population of students in a world where moting deep learning and is well aligned with other international initiatives. These strategies were developed for K-12 (primary and secondary education), but we believe the approach is valid for the first 2 years of college.

CORE IDEAS, CROSSCUTTING CONCEPTS. Disciplinary experts have a great deal of knowledge—organized and contextualized around important concepts (5). Students should develop knowledge around these “disciplinary core ideas” rather than try to assemble understanding from many disparate ideas and activities. Core ideas should be advanced over time through carefully developed progressions of learning activities national-level initiatives and the research literature, we believe that core ideas must be negotiated locally by faculty in each discipline in order to build ownership and buy-in.

For example, core ideas that emerged from cross-disciplinary discussions at our institution, Michigan State University (MSU), include “evolution” for biology, “structure and properties” for chemistry, and “interactions cause changes in motion” for physics. Focusing on core ideas within each discipline allows reduction of the amount of material that many agree has become overwhelming (the “mile-wide, inch-deep” problem). Faculty agreement on what is centrally important moves the conversation from what to eliminate to what

now, what claims can they support?
– What’s missing? What should be changed?
Some Benefits of Adopting an ECD Approach

- Two ways to make use of an ECD approach regarding assessment activities
  - **Forward Direction** – design new sets of assessment tasks/situations aligned with the goals of curriculum and instruction
  - **Backward Direction** – reverse engineer existing assessments to determine what claims can be supported by the specific forms of evidence available from the tasks
Assessment Should not be the “Tail Wagging the Science Education Dog”
Thank You!!!