engage^ny

Our Students. Their Moment.

New Science Assessments Measuring the NYS P-12 Science Learning Standards

Overview of Presentation

- Assessment Design Process
- Informing Instruction with Materials Resulting from Assessment Design
- Timelines for New Assessments and (Tentatively) Releasing Test Development Documents
- Opportunities for Partnership With NYSED to Prepare the Field for New Science Assessments

ASSESSMENT DESIGN PROCESS

Note: This section will provide context for each of the other topics.

What are test specifications?

- Test specifications are documented processes and design criteria that are followed when developing test items and assembling test forms

Why do we need specifications?

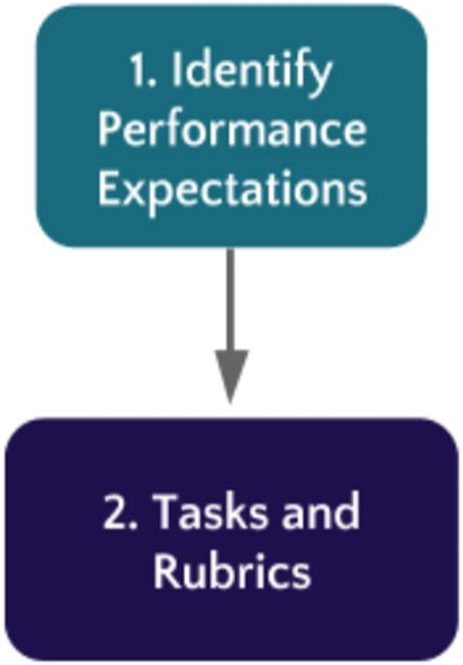
- To ensure:
  - Fair and consistent measurement of the content (fairness & reliability)
  - That the scores generated by tests consistently inform the questions they were designed to answer (validity)

How do we get specifications?

- Principled Assessment Design
  1. Identify the key ‘things’ a student should be able to do at the end of instruction
  2. Identify evidence that indicates a student can actually do those things
  3. Unpack specific knowledge, skills, and concepts
  4. Parse knowledge and skills across the range of performance
  5. Create task shells that allow students to show knowledge/skills
  6. Select knowledge/skills/concepts and fit with task shells (i.e., write questions)
  7. Build tests using specified numbers of type of task
  8. Administer tests

- Why this works
  - Ensures all aspects of assessment are connected and, thus, results will inform initial questions/claims
  - Allows for consistent development and administration of tests that are comparable
  - Focuses on conceptual and applied student understanding

Principled Assessment Design



The typical approach to writing questions involves:

- **Implicit** design decisions
- **Inconsistent** elicitation of knowledge and skills
- **Unexplainable variation** in contexts, difficulty, evidence elicited from students

- A principled assessment design approach connects the learning goals, evidence of those goals, and the tasks that produce this evidence.
  - Explicitly discussed in the NRC's report on developing assessments aligned to the NGSS (Pellegrino et al., 2014).

Claims

- Identify the key ‘things’ a student should be able to do at the end of instruction

- What do we want to be able to say about a student based on how this student performs on this assessment?

Evidence

- Identify evidence that indicates a student can actually do those things

- What does a student need to do to show us that he or she is meeting the goals we have outlined?

Domain Analysis

- “Domain analysis” is the fancy term; this could also be referred to as “unpacking the Learning Standards.”

- “Unpacking” allows us to identify the essential components of the standard, to understand what the standard really means, and to pinpoint the subject-area knowledge that students need to demonstrate proficiency in the standard

Domain Analysis

- When we “unpack” the Learning Standards, we unpack the skills that a student must demonstrate and the subject-area knowledge that they must acquire and show.
- The skills (and some subject-area knowledge) are unpacked in the Performance Level Descriptions (PLDS).
- Subject-area knowledge typically must be unpacked beyond what is in the PLDs and is handled differently based on the course.
  - e.g., science has three dimensions so we have to unpack each of them (see Harris, et al., 2016).

PLDs

- Performance Level Descriptions (PLDs) parse knowledge and skills across the range of performance
- What are the specific knowledge and skills, differentiated by performance level, that we are looking for students to demonstrate?
- A numeric score has no meaning unless we design the test in such a way that it can be tied to specific performances in terms of knowledge and skills.

Task Models/Design

- Create task shells that allow students to provide/produce that evidence

- What are the best types of tasks for students to demonstrate the (differentiated) knowledge and skills?

Note: this is where we start to think about overall form design

Examples from Intermediate Test

- We will look at some of the specifications produced for the new Intermediate Science Test (Grade 8) in science as examples.

- These are in draft form and still being worked on by NYS educators.

Example Claim

In a general sense, claims for the Intermediate Science Test explicate what students are able to do in science at the end of grade 8.

- Claims merge concept and skill to support less emphasis on declarative factual recall and more emphasis on developing skills as a vehicle to learn and apply concepts (research shows that this leads to longer retention of concepts).
- Not all combinations of concept and skill will be appropriate given the time and format constraints of the exam, the intended purpose, audience, and rigor (i.e., some Performance Expectations (PEs) will not be able to be tested on the exam).

Example Claim

Drafted Claim for Earth and Space Sciences:

- A student can apply scientific practices, principles and technologies related to the cyclic patterns and scale properties of objects in the solar system and the role of gravity in the motions of objects within space systems, the evidence from geoscience processes and plate tectonics at varying scales to explain the history of Earth, the flow of energy that drives the cycling of Earth's materials resulting in an uneven distribution of resources, the causes for the change in weather and climate patterns, and the impact humans have on Earth's systems and the mitigation of the effects of natural hazards on humans.

*Note: this claim covers all Earth and Space Sciences (ESS) topics for grades 6-8; 15 total PEs

Example Evidence

Evidence identifies what a student needs to do/say/produce to support acquisition of the claim

- Operationalize claim
- Lessen the ambiguity for teachers in understanding what is meant by a claim
- Define the specific language choices in the claim (What is meant by explain vs. describe?)

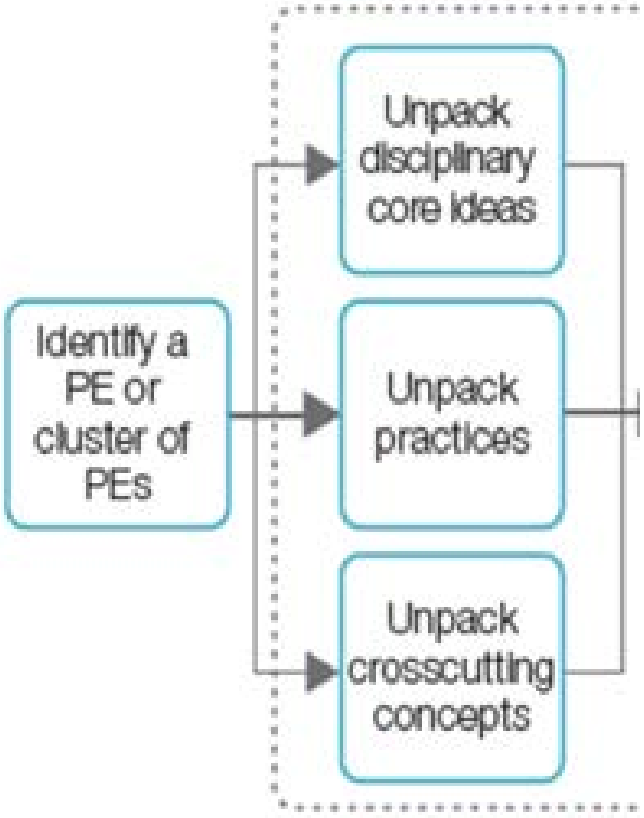
Example Evidence

Excerpt from Claim: "…the evidence from geoscience processes and plate tectonics at varying scales to explain the history of Earth…"

Evidence: *A student demonstrates understanding of the "History of Earth" through application, evaluation, analysis, and/or synthesis using science and engineering practices, core ideas, and crosscutting concepts related to:*

- Scientific explanations using geologic evidence to organize the 4.6-billion-year-old history of Earth [MS-ESS1-4]
- Scientific explanations based on evidence for how Earth's surface has changed at varying temporal and spatial scales [MS-ESS2-2]
- Data using geologic evidence to provide support for past plate motions [MS-ESS2-3]

Example Domain Analysis



- What complex of knowledge, skills, or other attributes should be assessed?
- Focus on relevant content to be assessed
- Includes a thorough unpacking of the three-dimensions (Harris, et al., 2016)

Example of Domain Analysis

Disciplinary Core Idea (DCI)

Unpacking Aspects of Disciplinary Core Idea HS-ESS1.C related to Earth's History PE:ESS1-4

Aspect of DCI	**ESS1.C: The History of Planet Earth:** The geologic time scale interpreted from rock strata provides a way to organize Earth's history. Analyses of rock strata and the fossil record provide only relative dates, not an absolute scale
Elaborating the meaning of key sub-ideas	• Older rock strata and fossils are typically found at the bottom of a column with the younger and more recent rocks, fossils and events occurring near the surface (unless overturning has occurred). • Rock strata can be correlated using fossils, rock types, and/or layers of debris from catastrophic events such as volcanic eruptions and asteroid impacts.

Other domain analysis elements for the DCI include:
- Defining expectations for understanding (within the target grade band)
- Assessment Boundaries
- Prerequisite Knowledge
- Challenges for Students (Preconceptions/Misconceptions)
- Relevant Phenomena

Example of Domain Analysis

Cross Cutting Concept (CCC)

Unpacking Crosscutting Concepts of Scale, Proportion, and Quantity related to PE: MS-ESS1-4

Key aspect of the concept	**Scale, Proportion, and Quantity:** Time, space, and energy phenomena can be observed at various scales using models to study systems that are too large or too small. • Develop scale models of geologic time to represent Earth's 4.6-billion-year-old history.
Intersections with practices (SEPs)	• **Analyzing and Interpreting Data** recorded from observations of landforms and rock strata. • **Developing and Using Models** such as cross sections to be consistent with real world examples of rock layers and fossils showing changes over time. • **Engaging in an Argument from Evidence** to support a scale timeline using evidence from rock strata and fossil records.

Other domain analysis elements for the CCC include:
- Evidence required to demonstrate application
- Prerequisite Knowledge

Example of Domain Analysis

Science and Engineering Practices (SEP)

Unpacking the Science Practice of Constructing Explanations and Designing Solutions related to PE: MS-ESS1-4

Key aspect of the practice	**Constructing Explanations and Designing Solutions -** Construct a scientific explanation based on valid and reliable evidence obtained from sources (including the students' own experiments) and the assumption that theories and laws that describe the natural world operate today as they did in the past and will continue to do so in the future. • Construct scientific explanations for phenomena related to Earth's history using relative dating, rock strata, and fossils that provide evidence of Earth's 4.6 billion year age and changes to Earth's surface over time.
Intersections with practices (SEPs)	• **Developing and Using Models** such as cross sections to be consistent with real world examples of rock layers and fossils showing changes over time.

Other domain analysis elements for the SEP include:
- Evidence required to demonstrate practice
- Prerequisite Knowledge

Example Performance Level Description (PLD)

PLDs are detailed descriptions of subject-specific and grade-specific knowledge and skills for the full range of each performance level.

- What does it mean, in terms of specific knowledge and skills, to get a specific performance level (i.e., 1, 2, 3, or 4 for Elementary and Intermediate Science Tests and 1, 2, 3, 4, or 5 for HS Regents Exams)?
- What are the claims you can make about students at each level of performance?
- Always define performance level in terms of what students can do, not what they cannot.
- A numeric score has no meaning unless we design the test in such a way that it can be tied to specific performances in terms of knowledge and skills.

G 3-8 Performance Levels (Draft)

NYS Level 4

- (…**exceed** grade-level expectations of learning standards)

NYS Level 3

- (…**meet** grade-level expectations of learning standards)

NYS Level 2

- (…**partially meet** grade-level expectations of learning standards)

NYS Level 1

- (…demonstrate knowledge and skills below Level 2)

HS Performance Levels (Draft)

NYS Level 5

(…**meet** grade-level expectations of learning standards **with distinction**.)

NYS Level 4

(…**meet** grade-level expectations of learning standards…likely prepared to succeed in the next level of coursework)

NYS Level 3

(…**minimally meet** grade-level expectations of learning standards…meet the content area requirements for a Regents diploma but may need additional support to succeed in the next level of coursework.)

NYS Level 2

(…**partially meet** grade-level expectations of learning standards.)

NYS Level 1

(…demonstrate knowledge and skills below Level 2.)

Example PLD

DCI	Level 4	Level 3	Level 2	Level 1
ESS1.C: The History of Planet Earth (MS-ESS1-4)	Construct a scientific explanation based on evidence from multiple sources for how the geologic time scale is used to organize Earth's 4.6 billion year history, and determine patterns of relative age for rock strata, fossils, and past geologic events.	Construct a scientific explanation based on evidence from rock strata for how the geologic time scale is used to organize Earth's 4.6 billion year history.	Based on a scientific explanation, determine the relative ages of the rock strata or the fossils within the strata.	Based on a scientific explanation, identify the youngest or oldest layer in rock strata.

Example Task Model

Task models identify the structures or shells for collecting evidence for the claims

- Task model variables will include things like item type, content, material given to examinees (stimuli, sources), things that drive difficulty up and down…etc.
- Many items can be written from one task model and there will be many models for a given assessment
  - Helps support form comparability because items between forms are developed from the same task models

Example Task Model

- Students are given a stimulus graph and asked to extract relevant information to answer a question.

Guidelines for New Test Development

We Do Talk About:

- The standards
- The practices
- The instructional shifts
- A hypothetical NYS student
- What skills students need to succeed in science

We Try Not To Talk About:

- The old Exam
- The old Core Curriculum
- Item types (e.g., MC, CR, etc.) until later in development process

Performance Tests

- The current Elementary- and Intermediate-level Science Tests each include a performance test that students complete during class time, prior to the written test.
- We are hearing consistently from science educators that, based on the content of the new learning standards, performance tests should still be a part of the new tests.
- We have begun exploring the development of performance tasks for the new Elementary and Intermediate Science Tests and getting recommendations from NYS science educators on how to proceed.
  - We are mindful of the impact that changes to the performance tests will have on schools (e.g., purchasing new supplies) and will take that into account.
- Conversations about Regents Exams (i.e., performance tests or required lab activities) have not yet begun.

Computer-based Testing

- If funding is available, the new Elementary and Intermediate Science written tests will be available via computer.
- Computer delivery would use the same system as the Grades 3-8 English Language Arts and Mathematics Tests to maximize student and teacher familiarity.
- Computer delivery creates opportunities for more innovative scientific exploration, such as computer simulations, further down the road.

engage ny

Our Students. Their Moment.

INFORMING INSTRUCTION WITH MATERIALS RESULTING FROM ASSESSMENT DESIGN

Informing Instruction

- Hopefully, much of what you saw in the these slides jumped out as being useful to informing instruction.
  - Claims & Evidence (syllabus, general info)
  - Domain Analysis (deep dive into learning standards)
  - PLDs (many, many things!)
  - Task Models (creating classroom assessment tasks)

- #1 recommendations are: PLDs and the documents and work related to the domain analysis
  - Other documents may also prove useful depending on the goal you're trying to achieve.

Informing Instruction

- Everything shown in this presentation can be adapted for use in smaller settings than state assessments: district, school or classroom level.
  - *Formative – even short activities*
  - *Summative – allows for stronger interpretation of results*
  - *Dynamic – PLDs let us make adjustments on-the-fly*

- The same materials can be used, following the same process.
  - *You can also create your own materials. Just follow this or another principled design process to ensure that you are anchored to a desired goal (claim).*

- The important thing is the focus on the interpretations that you hope to make with the results.
  - *If you want to assess how a specific skill has developed following instruction, develop tasks that measure the skill and provides information about what students are able to demonstrate in terms of the skill.*

Communicating Test Results

- PLDs also play a crucial role in communicating with students, teachers and parents about test results.
- Scale scores fall within a particular performance level (which has an overarching policy definition)
- Each performance level then has the list of specific knowledge and skills that are typical of students in that level (i.e., the PLDs).
- So we can take a student's test result and connect it to the specific knowledge and skills that they likely showed proficiency in.
  - Because the PLDs are on a continuum, we can also see what the student mastered in the past and where they can go in the future.

As A Reminder…

- NYS educators and other science stakeholders (e.g., steering committee) were engaged to divide the HS standards into course maps for four HS courses.
  - Course maps are posted on SED website.

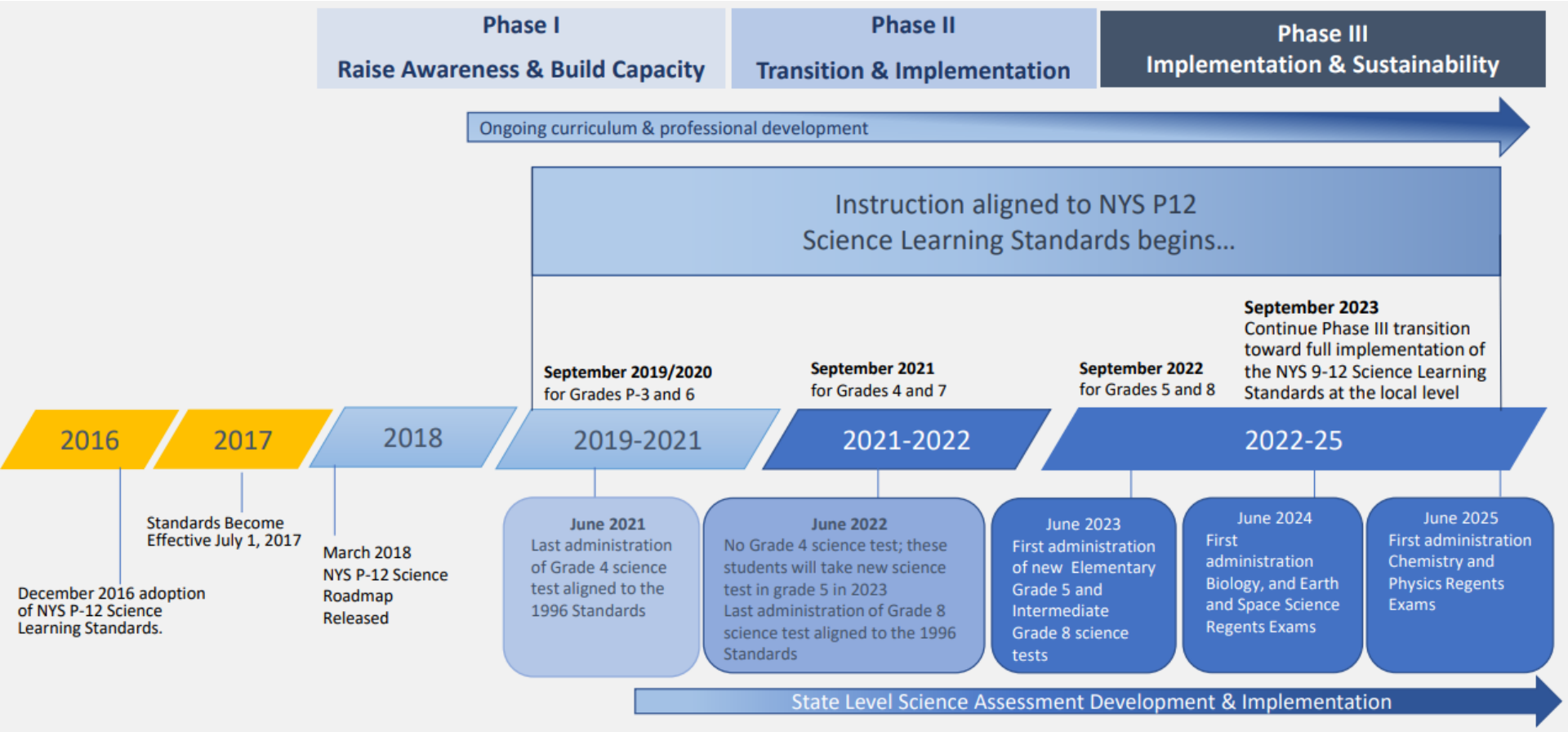
- The Office of Curriculum & Instruction has a website devoted to Science Standards Implementation Resources to support the "Raise Awareness & Build Capacity" phase of the Strategic Plan.

engage ny

Our Students. Their Moment.

TIMELINES FOR NEW ASSESSMENTS AND (TENTATIVELY) RELEASING TEST DEVELOPMENT DOCUMENTS

Timeline (updated April 2020)



New Assessments

- May/June 2023
  - Elementary Science Test
  - Intermediate Science Test
- June 2024
  - Regents Exam in Earth & Space Sciences
  - Regents Exam in Biology
- June 2025
  - Regents Exam in Chemistry
  - Regents Exam in Physics

Test Development Documents Release Philosophy

- Communicate (raise awareness)
- Plenty of PD (build capacity)
- Do no harm (avoid confusion)

Projected Released of Test Development Documents

- Elementary & Intermediate Science
  - PLDs: August 2020
  - Claims, Evidence and Additional Materials to follow in 2020-21 School Year

We expect to follow similar plans for the release of Regents Exam development materials, relative to their first administration dates.

OPPORTUNITIES FOR PARTNERSHIP WITH NYSED TO PREPARE THE FIELD FOR NEW SCIENCE ASSESSMENTS

Partnership Opportunities

1. Share today's messages
2. Get involved in test development
3. PD with the development materials
4. Other ideas?

Thank You!

- Questions about the NYS Science Learning Standards and timelines related to instruction can be directed to the Office of Curriculum and Instruction at: emscurric@nysed.gov

- For questions about test development and assessment or development document timelines, please contact the Office of State Assessment at: emscassessinfo@nysed.gov

- Information about opportunities to participate in test development can be found at: http://www.p12.nysed.gov/assessment/teacher/home.html

References

- Harris, C. J., Krajcik, J. S., Pellegrino, J. W., & McElhaney, K.W. (2016). *Constructing assessment tasks that blend disciplinary core Ideas, crosscutting concepts, and science practices for classroom formative applications.* Menlo Park, CA: SRI International.

- Pellegrino, J. W., Wilson, M., Koenig, J., & Beatty, A. (Eds.). (2014). *Developing assessments for the Next Generation Science Standards*. Washington, DC: National Academies Press.