Chapter 9
Assessment

2016 Science Framework
for California Public Schools
Kindergarten through Grade 12

Chapter Nine
Assessment

Introduction: Assessment as Science

Assessment is like science, and three-dimensional science learning should be assessed by applying the same three dimensions as the learning itself. To assess our students, we plan and conduct investigations about student learning and then analyze and interpret data to develop models of what students are thinking. These models allow us to predict the effect of additional teaching addressing the patterns we notice in student understanding and misunderstanding. Assessment allows us to improve our teaching practice over time, spiraling upward. Because of this strong link between assessment and instruction, this chapter is targeted to teachers and focuses on classroom assessment. It does not provide recommendations for district or state testing.

Purpose of Assessment

Assessment has two fundamental purposes: summative and formative. The key difference between these two purposes of assessment is *how* the information assessments provide is used: Either to guide and advance learning (usually while instruction is under way) or to obtain evidence of what students have learned, often for use beyond the classroom (National Research Council [NRC] 2014). For example, assessment for summative purposes helps determine whether students have attained a certain level of competency or proficiency after a more or less extended period of teaching and learning, typically after several weeks, at the end of a semester, or annually (American Educational Research Association et al. 2014). Inferences made from the results of these assessments can be used for accountability purposes, for making decisions about student placement, certification, curriculum, and programs, and for assigning grades. By contrast, formative assessment provides information about student learning day-by-day, week-by-week in order to guide next steps in teaching and learning and secure progress toward short-term goals. It is assessment that is tied to immediate learning goals and may involve both formal tasks as well as activities conducted as part of a lesson, such as classroom dialogue and observation. Often in formative assessment, instructional activities and assessment activities may be intertwined or even indistinguishable. For example, evidence of learning may be obtained from a classroom discussion or a group activity in which students explore and respond to each other's ideas and learn as they go through this process (NRC 2014). Formative assessment should assist students in guiding their own learning by evaluating and revising their own thinking or work; and foster students' sense of autonomy and responsibility for their own learning (Andrade and Cizek 2010, cited in NRC, 2014).

An important rule of thumb in educational assessment is that one-size-does-not-fit-all. In other words, assessment that serves one purpose may not appropriately serve another. As Hamilton and Stecher (2002) note, "requiring tests to serve multiple purposes sometimes results in the reduction of utility of the test for any one of these purposes" (Hamilton and Stecher 2002, 135). The *purpose for which* learners are being assessed should determine the choice of assessment instruments and their use.

Assessment Cycles

One way to think about assessment for different purposes is to conceptualize assessment as operating in different time frames or cycles: long, medium and short (Wiliam 2006). Each cycle provides information at varying levels of detail and inferences drawn from the assessment results are used to address specific questions about student learning and inform a range of decisions and actions.

Long cycle: Annual assessments, for example, are long-cycle assessments. They cover a year's worth of learning and, by their nature, provide a large grain size of information about student achievement relative to the standards.

Some of the questions that results from these assessments can help teachers answer are:

- What have my students learned? Have they met the standards assessed?
- What are the overall strengths and weaknesses in my class's learning?
- What are the strengths and weaknesses in individual's and groups' learning?
- What are the strengths and weaknesses in my/our curriculum and my instruction?
- Have the improvement strategies I/we put in place worked?

Medium Cycle: Interim/benchmark assessments are medium-cycle and address intermediate goals on the way to meeting end-of-year, or end-of-course goals. Typically administered quarterly or every six weeks, they cover a shorter period of instruction than long-cycle assessments and, consequently, provide more detail about student learning, although not enough to guide day-to-day teaching and learning. Results from interim assessments provide periodic snapshots of student learning throughout the year. These snapshots assist teachers to monitor how student learning is progressing and to determine who is on track to meet the standards and who is not. Medium cycle assessments can help teachers address these questions:

- What have my students learned so far?
- Who has and who hasn't met intermediate goals?
- Who is and who is not on track to meet end-of-year or end-of-course goals?
- What are the strengths and weaknesses in individual's/groups' learning?
- Who are the students most in need? What do they need?
- What are the strengths and weaknesses in curriculum and instruction?
- What improvements do I need to make in my teaching?

Assessments that teachers develop, or that are included in the curricular materials and are administered at the end of a unit of study, are also medium cycle. These can serve a summative purpose to evaluate student achievement with respect to the goals of the unit. If such assessments are given to students before the end of the unit when there is still time to take some instructional action before moving on to the next unit, then they can serve a formative purpose. Some questions that these assessments can help teachers answer are:

- Have my students met the goals of the unit?
- Are there some students who need additional help to meet the goals of the unit?
- What help do they need?
- What improvements do I need to make in my teaching next time I teach this unit?

Short-cycle: This cycle of assessment occurs when evidence of learning is gathered day-by-day from a variety of sources during ongoing instruction for the purpose of moving learning forward to meet short-term goals (i.e., lesson goals).

Short-cycle assessment provides the most detailed information for teachers to adjust their instruction or plan subsequent instruction, and for students to reflect on their learning and adjust their learning tactics as needed. Short-cycle assessment should help teachers answer these questions:

- Where are my students in relation to learning goals for this lesson?
- What is the gap[1] between students' current learning and the goal?
- What false preconceptions are evident?
- What individual difficulties are my students having?
- What are the next immediate steps in learning for my students?
- What do I need to do to improve my teaching?
- What feedback do I need to provide in order to help students move their learning forward?

Teachers are not the only assessors in short-cycle formative assessment. Students also need to be involved because ultimately it is the learner who has to take action to move learning forward. Short-cycle assessment should help students answer the following:

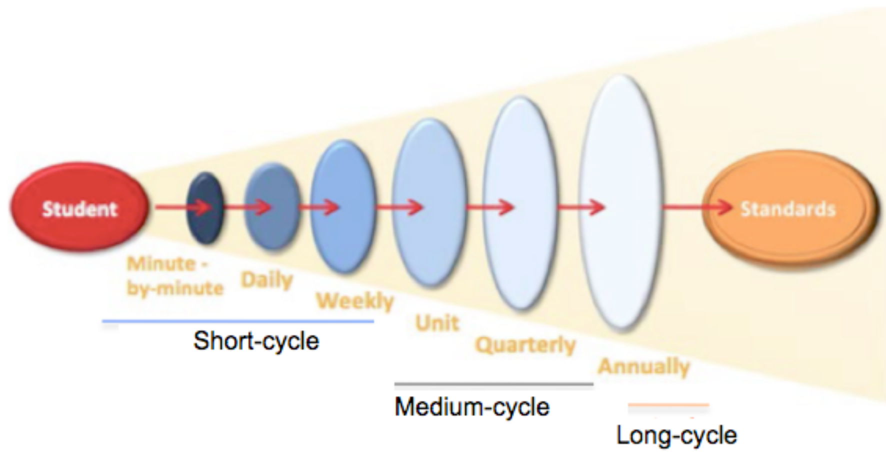
- Where is my learning now in relation to the learning goals for this lesson?
- Am I on track to meet the learning goals?
- What difficulties am I experiencing in my learning?
- What can I do about these difficulties?
- What are the strengths in my work? Where do I need to improve?
- What are my immediate next steps to move my learning forward?

Figure 9.1 shows a coherent assessment system with assessments of different time frames and of different grain sizes for different decision-making purposes. Importantly, assessments within each time frame gather evidence of learning toward the same set of goals so as to push teaching and learning in a common direction (Herman 2010).

---

[1] The gap refers to the distance between where the students' learning currently stands at particular points in the lesson (a lesson can be several periods) and the intended learning goal for the lesson. The purpose of short-cycle formative assessment is to close this gap so that all students meet the goal (cf. Sadler 1989).

Figure 9.1. A Coherent Assessment System



Source: Adapted from Herman and Heritage 2007.

Plan for Statewide Science Assessments

Because the Next Generation Science Standards for California Public Schools, Grades Kindergarten Through Grade Twelve (CA NGSS*)* are multifaceted, California faces a great challenge to implement a statewide assessment system that is comprehensive but not a burden on classroom time or other resources.

As required by the federal Department of Education, California students will take three statewide CA NGSS assessments during their K–12 education (table 9.1). In California, the California Department of Education (CDE) and State Board of Education (SBE) have made these decisions: each test event will take less than 2.5 hours (including instructions) and will be delivered entirely on a computer. The state test will include no hands-on performance tasks but will include performance assessment items on at least two of the three dimensions in Next Generation Science Standards (NGSS), including the practices, which can be completed on a computer.

Table 9.1. All Students Take Three Statewide CA NGSS Assessments

Grade	Material covered
Five	K–5 PEs
Eight	All middle school PEs (grades 6–8)
Once during Ten, Eleven, or Twelve.	All high school PEs (all students tested on all domains: Life Science, Physical Science, Earth & Space Science, Engineering, Technology & Applications of Science)

California's new NGSS-aligned state science assessment will, for the first time, include science performance expectations (PEs) taken from all grades in a span, not just the grade in which the test takes place. The SBE's rationale for this design is to promote science instruction across all grades, not just the grade in which the test is administered. The process for developing the new state summative assessments will begin with a pilot, followed by a census field test, and then operational administration currently scheduled for spring 2019.

In May 2016, the SBE took action to add student test scores from the state's science test, when available, to the state's accountability reporting for possible assistance or intervention, as well as to the federal Department of Education. In California's new integrated accountability model, the SBE expects student test scores on science, once available, to also be reported in district Local Control and Accountability Plans (LCAPs) under Priority 4, Student Outcomes.

A complete description of California's plan for an innovative, hybrid model, computer adaptive state summative science assessment design is available from the SBE (State Board of Education 2016), but a few details are relevant for designing instruction, preparing complementary classroom assessment as part of the overall assessment system, and interpreting the results of the assessments. Table 9.2 describes those key features and part of the rationale or motivation for each.

Table 9.2. Key Features of the Statewide CA NGSS Assessments

Test Feature	**Rationale or Motivation**
Test Features that May Influence Instruction and Curriculum Design	
Tests cover the PEs of a grade span (K–5, 6–8, or 9–12) rather than a single grade level or course. • Grade five assessment, consisting of grade five PEs and matrix sampling of PEs from kindergarten through grade four; • Grade eight assessment, consisting of middle school (grades six through eight) PEs; • Grade ten, eleven, or twelve assessment, consisting of high school PEs	The CA NGSS progressively build up understanding from grade to grade. Since knowledge is cumulative, the test provides incentives for schools to teach science every year and provide all students equal access to all standards.
Portions of the test will involve "doing science" through innovative item types or performance tasks presented on the computer.	The CA NGSS learning occurs when students engage in science and engineering practices.
Every test item will assess the integration of at least two dimensions at a time.	The CA NGSS are three dimensional.
Test Features that May Affect Interpretation of Test Results	
Students will be assessed on different PEs even when they take the test at the same time in the same room.	Test designers use statistical sampling techniques such that schools will be able to identify strengths and weaknesses in their overall program without having to increase testing time.
Two types of scores will be reported: individual student scores and group scores.	Each test includes PEs from multiple grades and understanding of the science and engineering practices (SEPs) and crosscutting concepts (CCCs) builds progressively over many grades thus encouraging science instruction in all grades. The addition of a group score allows for the inclusion of a more broad array of content making it a more powerful tool in identifying program strengths and weaknesses.

The remainder of this chapter focuses on how teachers and curriculum developers can emphasize these same features in their everyday classroom assessment system of the CA NGSS.

Assessing Three Dimensional Learning

Three dimensions of science learning are combined to form each standard: the core ideas of the disciplines of life science, physical sciences, earth and space sciences, and engineering and technology; the practices through which scientists and engineers do their work; and the key crosscutting concepts that link the science disciplines. Three-dimensional science learning refers to the integration of these dimensions. According to the report, *Developing Assessment for the Next Generation Science Standards* (NRC 2014) NGSS aligned assessments that address three-dimensional learning should be designed to:

1. Examine students' performance of science and engineering practices in the context of disciplinary core ideas and crosscutting concepts;
2. Contain multiple components (e.g., a set of interrelated questions). It may be useful to focus on individual practices, core ideas, or crosscutting concepts in the various components of an assessment task, but, together, the components need to support inferences about students' three-dimensional science learning as described in a given performance expectation;
3. Accurately locate students along a sequence of progressively more complex understanding of a core idea and successively more sophisticated applications of practices and crosscutting concepts;
4. Include an interpretive system for evaluating a range of student responses that are specific enough to be useful for helping teachers understand the range of student learning.

Measuring the three-dimensional learning described in the CA NGSS will require assessments that are significantly different from those in current use. For example, as shown in figure 9.2, items that assess disciplinary ideas alone are inadequate for assessing three-dimensional learning.

Figure 9.2. Example of Single Item vs Multi-component Task

Single item to assess one-dimensional learning:	Multi-component task to assess three-dimensional learning:
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	Subtask 1. Draw a model of a volcano forming at a hot spot using arrows to show movement in your model. Be sure to label all parts of your model. Subtask 2. Use your model to explain what happens with the plate and what happens at the hot spot that would result in the formation of a volcano. Subtask 3. 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 of the parts of your model. Subtask 4. Use your model to explain what happens at a plate boundary that causes a volcano for form.

Source: NRC 2014.

Classroom Assessment

The CA NGSS place an emphasis on classroom assessment, an integral part of instruction. Classroom assessment should include both formative and summative assessment: Formative assessment to guide instructional decision making and support students' own agency in learning while the learning is occurring; and summative tasks to make judgments about student learning (e.g., assign student grades) after a period of learning. Through carefully planned classroom assessment teachers can monitor student understanding of disciplinary core ideas, how they are reasoning and engaging in science and engineering practices, and the degree to which they are making connections through crosscutting ideas. Instructional practice that is aligned to the CA NGSS will include activities for teachers to gather evidence of three-dimensional learning, such as "when students develop and refine models, generate, discuss and analyze data, engage in both spoken and written explanations and argumentation, and reflect on their own understanding of the core idea and the subtopic at hand" (NRC 2014). As part of the CA NGSS performance expectations, teachers should also be aware of the assessment boundaries (identified in red following a PE) which clarifies the scope and detail appropriate to that grade level.

Conceptual Approaches to Designing Three-Dimensional Assessment

The CA NGSS were constructed with Evidence-Centered Design in mind (also see NRC 2014). Evidence-Centered Design treats assessment design and development much like the construction of an **argument [SEP-7]** in the CA NGSS. The objective is to make a claim about what students know by gathering evidence from what students say, do, make, or write to support the claim. In order to gather this evidence, teachers must invite students to engage in carefully designed tasks. Any claim that our students understand targeted disciplinary core ideas (DCIs), SEPs, and CCCs must be inferred from relevant, observable evidence. The PEs from the CA NGSS outline the tasks students can demonstrably accomplish when they attain the desired level of understanding.

PEs are quite broadly stated and need to be instantiated in specific classroom tasks that educators construct and engage students in. Three particularly useful resources supplement the PEs and help teachers design or evaluate assessments:

- *NGSS progressions*. What do students need to understand about ***cause and effect [CCC-2]*** at the high school level that they didn't already know in middle school? How much do students need to understand about Earth systems (ESS2.A) in middle school versus elementary school? Since the *CA NGSS* were designed to deliberately spiral upward, these distinctions (and many more like them) are important in designing grade-appropriate assessments. The progressions describe what students should understand and know at the end of each grade span for every sub-item in all three dimensions of *NGSS*. Simple tables of the progressions appear in Appendices E, F, and G of the original *NGSS* standards and are collected in one place in Appendix 3 of this *Framework*.

- *Evidence Statements***.** While a PE may take up a single line, the Evidence Statements released to supplement the *NGSS* expand on every single PE by describing the evidence that teachers would need to collect to ensure that students have met the PE. The Evidence Statements identify the underlying knowledge required for each DCI included in the PE, the key elements of the SEP that teachers should look for, and how the CCCs can be used to deepen understanding in this PE. Evidence statements are available on the Achieve website at http://www.nextgenscience.org/evidence-statements
- (Achieve 2015).
- *Assessment Boundaries and Clarification Statements*. These brief statements appear in red beneath each PE in the standards. They present a very abbreviated version of what the previous two resources describe. Assessment Boundaries usually place the PE in the context along the progression of complexity and the Clarification Statements highlight some of the details that are expanded upon in the evidence statements.

Both the progressions and evidence statements are hard to describe in a sentence or two, but they are extremely valuable as teachers design instruction and assessment. The *Framework* writers used them as a constant reference. Readers that are not already familiar with them should consider stopping and viewing them before continuing on.

Performance Tasks

CA NGSS instruction is centered around phenomena and *NGSS* assessment should be as well. Such authentic assessment require that students apply their full three dimensional 'toolset' to new phenomena or new problems. The goal of three-dimensional assessment is therefore not to test *what* students know, but to see how successfully they can *use and apply* what they know. One way to accomplish this form of assessment is through classroom-embedded performance tasks. As students conduct science and engineering within the classroom, they record their work in ways indicated by the performance task and this record provides the basis for assessment. The tasks may involve hands-on work, investigation using simulations, or analysis of data produced by others.

Performance tasks that assess the CA NGSS:

- Present students novel phenomena or problems.
- Assess a single PE or a bundle of related PEs.
- Include multiple tasks that may focus on at least two NGSS dimensions.
- Can be formative or summative.
- Can be hands-on, computer-based, or a hybrid of the two.
- Provide instruction and context so that students understand the nature of new phenomena before being assessed about them.
- May include intermediate instruction between tasks.
- Can be teacher-developed as part of formative assessment, embedded within a curriculum package, or developed and distributed by the state or districts as self-contained scenarios.

There are many models for how performance tasks can be delivered in a classroom. These tasks can be developed by teachers as part of their regular instruction and formative assessment, or they can be fully contained scenarios provided by districts or the state to be administered by teachers at the correct time within the flow of a course. Technology can enhance the delivery of performance tasks, especially when they will be centrally scored. Tasks can also be hybrid where students perform part of an investigation using hands-on materials in their classroom and part of the investigation using computer simulations or computer-based assessment prompts.

Teachers may need to deliver instruction as part of the assessment in order to introduce the specific scenario being investigated, which is one way in which instruction and assessment begin to merge in the CA NGSS. Even once students understand the phenomena, there may need to be instruction embedded between different tasks in the multi-part performance tasks. For example, a performance expectation might require that students **develop a model [SEP-6]** of a system and then use it to write an **explanation [SEP-6]** describing a specific cause and effect relationship in the system. These practices are interrelated, but what if a student is unable to develop a viable model during the assessment? An assessment would likely include multiple tasks that each focus on one of the two practices. The second task may not show a clear picture of the student's ability to construct explanations unless there is an intermediate stage of instruction between the two tasks to make sure that students have a viable model before continuing on. Within a computer-based assessment, the instruction can be done through software tutorials. Because the tasks are presented sequentially, educators still gain insight into where individual students are along the continuum of skill for performing individual SEPs and applying individual DCIs and CCCs.

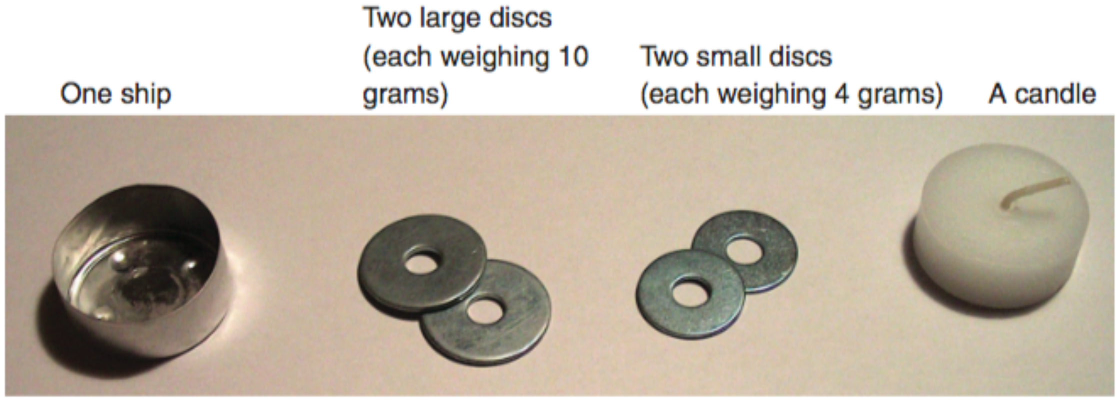
Example Performance Task 1: Primary Grades Hands-on Investigation

NRC (2014) present a performance task for students in the primary grades based on a hands-on investigation. The description that follows is an abbreviated version of what appears in that document. While this task is research-based, it is was written before the *CA NGSS* and employs DCIs that are not introduced in the primary grade span within the *CA NGSS* and therefore is not a 'classroom-ready' *CA NGSS* assessment. Despite this shortcoming, it is included in this *Framework* as an example of using a hand-on performance task with young children to assess three dimensional learning.

Students receive a set of materials shown in

Figure 3. In the task, students investigate floating and sinking, but the task assumes no prior knowledge about why objects float (or do not float). Instead, the task uses this novel phenomenon to probe students' use of SEPs and broader understanding of CCCs. Out of the six prompts, several SEPs and one CCC are assessed multiple times. Two of the prompts focus on a single SEP (with CCCs), but students must apply multiple SEPs for the majority of the tasks.

Figure 9.3. Materials Provided for Performance Task 1



Source: NRC 2014.

Prompt for Question 1

Your ship can be loaded in different ways. We will try out one way. In a few minutes, you will place the small disc as cargo in the ship. You will put the disc on the inside edge of the ship, not in the center. What will happen when you put the ship in the water? In the space below, draw a picture of what you think will happen. On the lines below, write an explanation of what you think will happen.

Scoring Rubric for Question 1

3 Points: Drawing/answer that reflects the following ideas: The ship is floating but is tilted to one side. The placement of the disc on the inside edge of the ship caused the ship to float unevenly.

2 Points: Drawing/answer that reflects the following concept: The ship is floating but is tilted to one side. There is no explanation for why it tilts.

1 Point: Drawing/answer that indicates that the ship floats, but there is no recognition that the off-center placement of the weight causes the ship to float unevenly.

0 Points: Drawing/answer that indicates that the ship sinks—or other answers/drawings.

Commentary

This prompt helps set the stage for the rest of the task and has less assessment value than some of the later questions. Since classroom performance tasks are opportunities for both teaching and assessing learning, sometimes prompts may be inserted for learning value rather than for assessment purposes. This prompt forces the student to make a prediction and establish their preconceptions, an important aspect of conceptual change theory.

SEPs. Students must apply a mental model of floating objects to make a prediction. Mental models, however, cannot be assessed (because they are inside students' heads) and so this particular item does not do an effective job of assessing the modeling [SEP-2].

DCIs. The task requires background physical science knowledge about buoyancy and balance, though these ideas do not correspond directly with any of the primary grade DCIs in *CA NGSS*.

CCCs. Level 1 on the rubric scale is for responses that fail to recognize the ***cause and effect relationship [CCC-1]*** between the boat being off center and the placement of the weight.

Prompt for Question 2

Place the disc in the ship as was demonstrated for question 1 and then place the ship onto the water. Observe what happens. In the space below, draw a picture of what happened. On the lines below, write an explanation of what happened. Try to include as details in your drawing and explanation that you think might help explain why the ship behaves the way it does.

Scoring Rubric for Question 2

2 Points: The drawing contains the following elements: the water surface, the ship floating tilted in the water, the lowest point of the ship is the side containing the disc. The written explanation indicates that the ship floats but is tilted.

1 Point: The drawing contains some points of the correct solution (e.g., it may contain two elements, such as the water surface and tilted ship, but part of the explanation is missing).

0 Points: Other

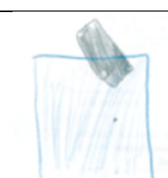
Commentary

SEPs. The rubric requires that students identify all the key elements in their pictures (Figure 9.4), which is essentially deciding what sort of data to collect. This decision is part of **planning an investigation [SEP-3]**. Students write a brief **explanation [SEP-6]** of their prediction and **communicate [SEP-8]** using a drawing. This prompt elicits these practices at the level expected in the primary grade span, but this example should not be used as an exemplar of assessing these practices at a higher level. An **explanation [SEP-6]** for a higher grade level requires students to connect the phenomena to scientific principles, rather than just this prompt's evidence-based account of what happened. **Communication [SEP-8]** at a higher grade level requires intent to communicate to a specific audience, rather than this example's drawing that simply illustrates scientific ideas.

DCIs. This prompt does not require knowledge of DCIs.

CCCs. This prompt does not require understanding the CCCs.

Figure 9.4. Example Responses for Question 2

2 points	1 point	1 point	0 points
"The disc makes the ship heavy on one side."	"The ship floats but tilts and water comes in."	"It turns over."	"It constantly moves to the edge."
	No image drawn.		

Source: NRC 2014

Prompt for Question 3

What else would you like to know about the ship and what happens when it is loaded with the discs? Write your question below.

Scoring Rubric for Question 3

3 Points: Questions or hypotheses similar to "Does the ship sink when I load it evenly with all four discs?"

2 Points: Questions or hypotheses similar to "What happens if I load the ship with two large discs?"

1 Point: No real question/question not related to material/problem recognizable

0 Points: Other questions (e.g., How far does it splash when I throw the discs into the water?) or statements (e.g., Put the disc into) the ship.

Commentary

SEPs. Students **generate their own questions [SEP-1]**.

DCIs. This rubric does not measure knowledge of DCIs.

CCCs. The rubric score gives high priority to questions that probe ***stability and change [CCC-7]***, though the prompt does not specifically cue students to view the problem through this lens. This rubric may miss 'outside the box' thinking if students ask really insightful questions that are not related to sinking.

Prompt for Question 4

Research your question. Perform an experiment to find the answer to your question. Draw and write down what you have found out.

Scoring Rubric for Question 4

2 Points: Answer fulfills the following criteria: 1) Tight relation to question: Design provides answer to the posed question/problem; and 2) the observations (drawing and text together) are detailed (e.g., The ship tilted to the left, the load fell off and sank quickly).

1 Point: Answer fulfills the following criteria: 1) Somewhat connected to the question: Design is at least directed toward the posed question/problem; 2) the observations (drawing and text together) are understandable but incomplete or not detailed (e.g., The ship tilted).

0 Points: Other answers

Commentary

SEPs. This prompt is an authentic and brief opportunity to **plan and carry out a simple investigation [SEP-3]**.

DCIs. This prompt does not require knowledge of DCIs.

CCCs. This prompt does not require understanding the CCCs.

Prompt for Question 5

Consider what you could learn from the experiments you have just done. Mark “Learned” if the statement indicates something you could find out from these experiments. Mark “Not Learned” if it is something you could not learn from these experiments.

Learned	Not Learned	
X		When discs are placed at the edge of a ship, it can turn over and sink.
	X	Ships need a motor.
X		The heavier a ship is, the deeper it sinks into the water.
X		A ship made from metal can be loaded with iron and still float.
	X	Round ships float better than long ships.

(Correct answers are marked above).

Commentary

SEPs. Each of these statements is a claim, and students must decide if the investigation provided **evidence to support that claim [SEP-7]**.

DCIs. This prompt does not require knowledge of DCIs.

CCCs. This prompt also assesses **cause and effect relationships [CCC-2]**, as students should only claim to have learned about the items where both the cause and the effect were observed. The items learned can be related to DCIs about forces and weight.

Example Performance Task 2: Secondary Scenario-based Assessment

Oakland Unified School District (OUSD), an early NGSS implementer, has developed NGSS performance tasks where students apply different SEPs to answer a single big question over multiple days. In the seventh-grade task, students learn about and engage in an entirely new situation based around a fictional scenario storyline:

Student Storyline

In order to prepare for the Mars One Mission, a company called Biodome has decided to send a team of scientists and doctors to live under a dome on Earth. You are an environmental scientist working for the Biodome Company to help analyze any data that the scientists collect. A catastrophe has occurred and death is imminent. Your task is to find out what is wrong based on data collected from the monitoring devices before it's too late.

The first day of the performance task, the teacher introduces the task and students read a one-page summary that provides context and background about the conditions on Mars and explains how the Biodome operates as a closed system to provide a livable habitat. Students then learn about the real-life Biosphere project on Earth. They apply their **mental model [SEP-2]** of the ***cycling of energy and matter [CCC-6]*** in photosynthesis and respiration (LS2.1C) from previous instruction to construct **explanations [SEP-6]** that form the basis for assessing MS-LS1-6 (“Construct a scientific explanation based on evidence for the role of photosynthesis in the cycling of matter and flow of energy into and out of organisms.”).

Prompt for Questions 1 and 2

Task Problem

About 20 years ago, a project under the name Biosphere 2 began a two-year experimental study in a closed environment, but something went terribly wrong. Learning from the scenario below will help make the current Biodome project a more successful one. You will need to employ your expertise of matter and energy involved in chemical reactions, especially in photosynthesis and respiration, to explain what happened.

SCENARIO: Data from the environment in the Biosphere 2 project showed that the percentage of sunlight that was transmitted through the glass ceiling was *20 percent less* than what was expected.

Answer the following questions:

1. Explain how this decrease in sunlight affected the *plants' ability to grow*.
2. Explain how this decrease in sunlight leads the *people in Biosphere 2 to struggle with not having enough food to survive*.

Scoring Rubric for Questions 1 and 2

Expert **3**	**Proficient** **2**	**Emergent** **1**	**Novice** **0**
-Includes all the elements of the Proficient level AND -Details included like: • plants performing respiration to use stored energy for growth • specific structures that allow matter to enter and exit the organism or that perform the reactions	-Explanations demonstrate how energy is needed as an input to convert matter in photosynthesis to products needed for growth -Explanations include how a change affecting the products of photosynthesis affects the reactants of respiration and in turn the energy output -All ideas are scientifically accurate	-Explanations explains connection to either photosynthesis OR respiration -Some ideas may not be scientifically accurate.	-Explanations are either unclear or are largely scientifically inaccurate

Commentary:

SEPs. While the prompt calls for an **explanation [SEP-6]**, the rubric does not specifically measure the qualities of the explanation itself. Additional subscales could be added.

DCIs. This rubric is almost entirely focused on DCIs related to photosynthesis and respiration (LS1.C).

CCCs. Level 2 of the rubric does invoke the ***flow of energy and cycling of matter [CCC-5]*** while level 3 students also include ***structure/function relationships [CCC-6]***. Even though these CCCs are mentioned, the rubric scale itself does not assess varying levels of understanding of these CCCs. If the intent of the rubric is to assess the depth of understanding of the CCC, it would need a separate subscale that determined if students were achieving the middle school level of mastery according to Appendix 3 (i.e., energy may take different forms and matter is conserved because atoms are conserved).

On the second day of the performance task, students learn the details of the 'crisis' in the fictional Biodome scenario. The scenario includes specific data about the levels of oxygen. Students examine these data to track down the source of the problem in the Biodome. Their work forms the basis of the assessment of MS-LS2-1 ("Analyze and interpret data to provide evidence for the effects of resource availability on organisms and populations of organisms in an ecosystem.").

Prompt for Questions 3 and 4

Task Problem

Imagine Biodome has been up and running for one year. This Biodome project improved the design of the glass structure to allow more sunlight to come in. However, you just received the latest report from the doctors that they are concerned that the Biodomians are complaining about having very little energy and seem very unhappy. The scientists have reported that the plants and crops in the Biodome's ecosystem are starting to die. You have 24 hours to figure out what is going wrong in Biodome's ecosystem before an emergency is declared and the project is terminated.

Over the next two days, you will eventually figure out:

• *What is causing the plants' slowed growth, and*

• *Why the scientists and doctors in the Biodome feel like they have less and less energy.*

Answer the following questions:

3. *Data Analysis*. For each of the columns in the data table below, write a sentence to describe the trend of the data for each factor (temperature, light intensity, CO_2 level, O_2 level, water taken up by roots, photosynthesis rate)
4. *Graphing and Interpretation***.** On the graph paper provided, create two graphs from the data. Each graph should have a title and labeled axes.
    - The first graph must show the photosynthetic rate over time
    - The second graph must show how *the factor you believe is causing the problem* changes over time (plot just one factor: Temperature, Light Intensity, CO_2 level, O_2 level, Water taken up by roots)
    - Under each graph, explain:
        (i) the story of the two sets of data and how they are connected.
        (ii) the importance of any relevant breakpoints in the data.

Weekly Average Environmental data recorded at 12:00 p.m. (noon)

Week	Temp (°C)	Light Intensity (%)	CO_2 (% of air)	Fraction of Water (H_2O) taken up by roots	O_2 (% of air)	Photosynthesis Rate (O_2 production)
1	25	100	0.030	1.0	21	100
2	24	100	0.030	1.0	20	100
3	25	100	0.028	1.0	19.5	90
4	24	100	0.026	0.9	19	80
5	25	100	0.025	0.9	18.5	80

6	25	100	0.022	0.8	18	70
7	24	100	0.018	0.5	17	50
8	24	100	0.015	0.3	16.5	30
9	25	100	0.014	0.3	16	30

Week 1 - Entry from a Biodomian's notebook: Everything seems to be functioning properly here in the Biodome. We started to plant our own crops such as hyacinth beans and sweet potatoes. I have measured the Hyacinth Beans to be 1.8m high. I am very excited to see how they continue to grow.

Week 3 - Entry from a Biodomian's notebook: I am starting to really get sick of all the sweet potatoes we are eating here. The hyacinth beans seem to be having trouble adjusting to the environment here as they are now 1.5m tall and some of the leaves are beginning to turn brown. I am noticing that the scientists are complaining that it seems like it is getting harder to breathe and stay entertained.

Week 6 - Entry from a Biodomian's notebook: We are getting really worried about the crops here because the hyacinth beans have wilted and are now only 1.2m tall. We also found dead insects and worms in the soil. Our doctors have reported that everyone has complained about low energy levels.

Week 9 - Entry from a Biodomian's notebook: I am starting to feel extremely exhausted. I woke up in the middle of the night feeling like I could not breathe. Hopefully the doctor can figure out what is happening. I went to check on the crops earlier this week and only half of the hyacinth beans are still alive and only 1m tall. The birds in the Biodome haven't been making much noise recently.

Scoring Rubric for Questions 3 and 4

Question	**Expert 3**	**Proficient 2**	**Emergent 1**	**Novice 0**
3. Data Analysis	-Includes all the items in *proficient* level -Interprets relationships	-Identifies correctly all trends in the data -Supports trends	-Identifies some trends in the data correctly -Does not	-Description of trends are unclear or largely incorrect

	between multiple factors -Identifies the optimal range for each factor for photosynthesis	with specific numeric data (numbers) -Describes all factors fully and correctly	use specific numbers to prove patterns	
4. Graphing and Interpretation	-Includes all the items in *proficient* level -Could include details like: * Determines a best fit line * Indicates a slope or mathematical representation for any relationships	-Plots photosynthesis rate accurately - Plots the correct factor accurately -Title and both axes labeled properly on both graphs -Describes the story of how the factor and photosynthesis rate are connected -Explains the relevance of a breakpoint in the data	- Plotting has some errors - Might be missing a title or labels -Description of connection or breakpoint is inaccurate	-Graph is hard to read or many elements missing -Description of connection or breakpoint is unclear or absent

Commentary:

SEPs. The rubrics for these two prompts separate out two independent subskills within SEP-7. **Analyzing data [SEP-7]** involves reading the table in question 3 and can reasonably be assessed one-dimensionally. **Interpreting data [SEP-7]**, however, requires a direct link to the other two dimensions.

DCIs. Level 3 of the rubric for question 4 has students relate photosynthesis (LS1.C, LS2.B) to other factors. Students draw on their understanding of relationships between parts of ecosystems (LS2.A) as part of their reasoning about this relationship.

CCCs. The breakpoint in the data mentioned in Level 3 of the rubric is an example of ***stability and change [CCC-7].***

During the final day of the performance task, students make a claim about the cause of the problem in the Biodome. They support their claim with evidence from the previous day and reasoning based on their understanding of ***cause and effect relationships [CCC-2]*** and ecosystem functioning and dynamics (LS2.C). This argument forms the basis of a three-dimensional assessment of MS-LS2-4 ("Construct an argument supported by empirical evidence that changes to physical or biological components of an ecosystem affect populations.").

Prompt for Question 5
After examining the data, make a clear claim as to which factor is causing the plants to die and the Biodomians' loss of energy. Be sure to support this with evidence from the reading and data resources provided. Make sure to include each of the following in your explanation:

- reasoning that includes the role of *photosynthesis* in this problem
- reasoning that includes the role of *cellular respiration* in this problem
- an argument against another factor being the cause of the problem

Scoring Rubric for Question 5

Criteria	Expert 3	Proficient 2	Emergent 1	Novice 0
Argument Claim: Cause and Effect	-Claim for factor causing the problem is clearly stated and connects to the chemical reactions driving the change to system	-Claim for factor causing the problem is clearly stated and best fits the data	-Claim for factor causing the problem seems possible and is clearly stated -Multiple factors may be given	-Claim for factor is unclear or absent
Argument	-Includes all	-Provides	-Provides	All evidence

Evidence: Supporting Claim	the items in *proficient* level -Organizes evidence to leave the audience with your strongest piece of evidence	*appropriate and sufficient* evidence from the data and reading resources -Includes analysis that compares factor data to photosynthetic rate at different points with specific quantitative data to support claim -Identifies source.	*appropriate evidence, but needs more* to support the claim -Source may or may not be identified	is in-appropriate and/or *DOES NOT* support the claim -Source may or may not be identified
Argument Reasoning: Photosynthesis/ Respiration Connection	-*Accurately* explains why the evidence supports the claim. -Includes all items from *proficient* level -Possible details included like: • Explains how energy is stored in the bonds of certain molecules and released during chemical reactions • Explains how	-Explains why the evidence supports the claim *with minor corrections needed*. -Describes how each piece of evidence is connected to photosynthesis and/or respiration -Demonstrates how plants and animals are interconnected through the products and reactants of the reactions	-Explains why the evidence supports the claim. -Reasoning demonstrates connections to photosynthesis and respiration, but many ideas are inaccurate.	-Explanation of connections between evidence and claim are *unclear with major inaccuracies*

	molecules can be rearranged in the body to perform different functions	-Explains how changes in matter with the reactions relates to energy and the use of energy		
Argument Rebuttal	Rebuttal *CONVINCINGLY* disproves another claim	Rebuttal addresses another claim, but *does not disprove it*	Rebuttal actually proves the alternative claim and *weakens the overall argument*	Rebuttal is unclear or absent

Commentary:
One could argue that Question 5 is not perfectly aligned to MS-LS2-4 because the focus in the PE should be the effects on 'populations', which implies shifts in the number of individuals or characteristics. The biosphere crisis in this scenario affects individual 'organisms' within a population and students have minimal data about the populations overall. The potential misalignment illustrates the challenge of developing authentic performance tasks with coherent storylines that also fit into the narrow specifications of the CA NGSS PEs. Despite this shortcoming, the prompt represents a culmination that requires integration of all three CA NGSS dimensions.

SEPs. This rubric measures the subcomponents of an effective **argument [SEP-7],** a claim, evidence, reasoning, and addressing a counter-claim.

DCIs. The 'Reasoning' criteria in the rubric focuses on how matter and energy are related in organisms (LS1.C) and ecosystems (LS2.C). The highest level rubric also shows students drawing connections to LS1.A (Structure and function) and PS1.B (Chemical reactions)

CCCs. The highest level of the 'Claim' criteria includes a specific causal mechanism not mentioned in the lower levels. This distinction reflects the fact that the middle school understanding of ***cause and effect [CCC-2]*** highlights the difference between correlation and causation (Appendix 3).

Strategies for Three Dimensional Assessment

The previous section illustrated examples of rich, multi-component assessments. These assessments included a series of simpler sub-tasks that may assess only two dimensions at a time. The sections below provides ideas, insights, and strategies that teachers can use to design some of these subtasks. The snapshots below pull out individual SEPs to give simple pictures of an otherwise overwhelming world of three dimensional assessment. The examples are organized by SEP because assessment design does require that students “do” something in order to demonstrate their learning, but assessment of DCIs and CCCs is embedded within each example. As teachers integrate strategies like these into their teaching, they can eventually be able to construct fully integrated performance tasks of their own that simultaneously assess multiple practices, or evaluate assessment tasks written by others to ensure that they include rigorous assessment of all three dimensions.

Asking Questions and Defining Problems

While questions stem from natural curiosity, the CA NGSS is trying to cultivate students' ability to ask productive scientific questions by the end of the K–12 progression. Questions are often the entry point into scientific processes that spur innovations and discoveries, so assessment of this SEP might focus on evaluating whether or not questions are scientifically productive. The form of the assessment varies based on the grade level.

Assessment Snapshot 9.1: Distinguishing Between Helpful and Unhelpful Questions in Primary and Elementary School
Mrs. J's first grade class has just completed the snapshot "Matching Environment and Needs." She then tells students that she has a mystery animal and they will need to ask questions to figure out what the animal is. After having students write their own question in their science notebook, Mrs. J provides students a list of questions and asks them to categorize each as either helpful or unhelpful. Mrs. J notices that many of the students identify the question, "Does the animal drink water?" as 'helpful.' She leads a class discussion about the question and reminds students that all living things need water to survive. In explaining their answers, Mrs. J realizes that students have a misconception that fish 'drink' the water in order breathe, so this question is helpful for deciding if the animal is a fish. She draws a chart on the board comparing the words 'drink' and 'breathe' and has students help her describe the differences. Mrs. J has students return to their initial questions and revise them in order to make them 'more helpful.' They then get to ask them and discover what the mystery animal is (Inspired by Jirout and Klahr 2011). **Commentary**: *SEPs*. At the primary level, students **ask questions [SEP-1]** to "find more information about the natural…world" (Appendix 3). By the time students enter elementary grades, they should be asking questions that require investigation and not just gathering information, but this task is on target for primary grades. *DCIs*. The task requires that students connect animal parts with their functions (LS1.A) and that all animals need food to grow (LS1.C). *CCCs*. This task does not assess understanding of CCCs.

Assessment Snapshot 9.2: Asking Questions about Cause and Effect for Middle and High School

The CA NGSS emphasizes the *student's* ability to ask questions, so formative assessment of this practice involves providing opportunities for students to ask questions and then evaluate them. Dr. D has students ask questions at the end of each class period about what they want to know next.

In this example, students just spent the class analyzing graphs of earth's temperature during the 20th century (MS-ESS3-5). Once all students have submitted their questions to an online tool using their smartphones, Dr. D has them use the tool to vote on which would be the most 'productive' questions to pursue during the next class period. Dr. D asks students to evaluate specific questions by asking, "Would answering this question help us determine a ***cause and effect relationship [CCC-2]***?" The questions are displayed anonymously and because Dr. D uses this strategy regularly in his class, he has established a climate that the voting process is not a popularity contest; it is a learning process and the whole class benefits from having a range of questions to compare. After class, Dr. D individually reviews the questions and quickly assigns the questions to a rubric scale (

Figure 5), noting which criteria his students have mastered and which they have not. He wants to share his results with his professional learning community that meets after school to see how they compare to other classrooms. Perhaps his colleagues have had more success and can offer tips about how he can help focus the student questions.

Commentary:

In this case, the questions themselves are formative assessments of individual students and the voting process provides feedback about the overall class level of understanding of the elements of effective questions. See the rubric in

Figure 5 for how this task assesses SEPs, CCCs, and DCIs.

Figure 9.5. Rubric for Asking Questions About a Cause and Effect Relationship in Global Climate

	3	2	1
Science and Engineering Practice **[SEP-1] Asking Questions**	Question draws on specific evidence in the graphs and could be answered through application of other SEPs (i.e., is scientifically testable).	Question draws on specific evidence or is scientifically testable, but not both.	Question may express curiosity, but does not build on evidence presented and is not specific enough to be testable.
Disciplinary Core Idea **ESS2.A Earth Materials and Systems**	Question invokes energy and mass transfer between Earth's systems or the flow of energy into/out of the system.	Question invokes interactions between Earth's systems.	Question does not build on existing DCI knowledge or invokes DCI material that is not relevant to the phenomena.
Crosscutting Concept **[CCC-2] Cause and Effect**	Question asks about a specific cause-and-effect mechanisms or acknowledges the possibility of multiple contributing causes.	Question inquires about the existence of cause and effect relationships but is not specific.	Question does not probe cause-and-effect mechanisms.

This rubric could be revised for other phenomena primarily by modifying the DCI subscale.

Examples from a task interpreting graphs of average global temperature in the 20th century (MS-ESS3-5): "Is the temperature warming?" (Rubric score 2, 1, 1 on SEP, DCI, and CCC, respectively)

Commentary

This question could be measured and investigated, but this question ignores the data presented in the task that already answer this question. While answering this question might inspire other questions about cause and effect, this question by itself does not probe any cause and effect relationships.

"Why does the temperature go up and down so much?" (Rubric score 2, 1, 2)

Commentary

This question is based on observations, but is not specific enough to investigate. The word 'why' probes cause and effect.

"Could the temperature increase be caused by the sun getting brighter?" (Rubric score 3, 3, 3)

Commentary

This question correctly interprets a warming trend on the graph, draws on DCIs that relate climate to energy from the sun, is specifically testable if data about the Sun's brightness was available, and inquires about a specific cause and effect relationship.

(Adapted from d'Alessio 2014)

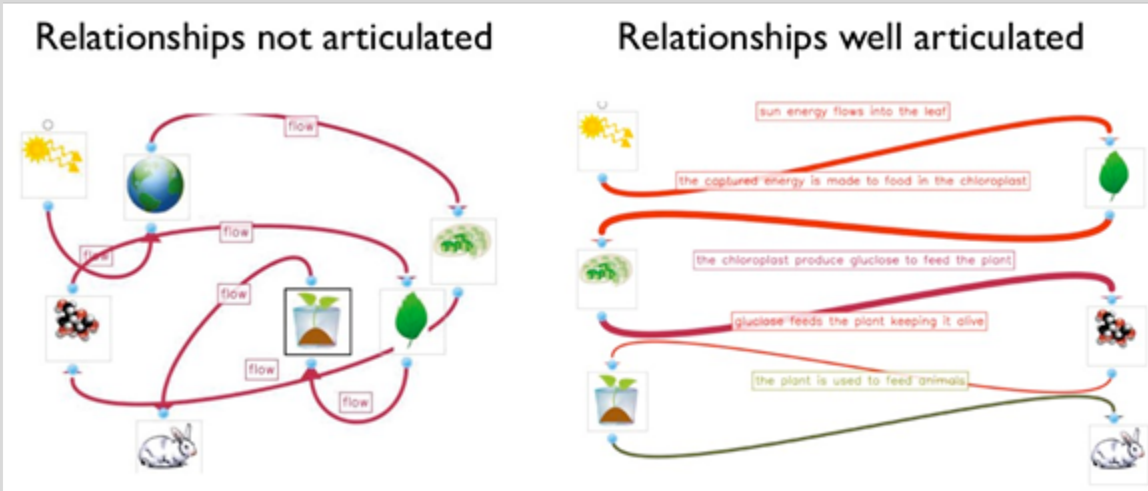
This rubric scale focuses on a task of **asking questions [SEP-1]**, but is also an indicator of the understanding of DCIs, and understanding of ***cause and effect relationships [CCC-2]***.

Developing and Using Models

In the early grades, models are typically more tangible representations such as physical models or pictorial models/diagrams. By high school, these models can be more abstract conceptual models represented by concept maps, mathematical models, or even computer codes. In almost all cases, these are models of ***systems [CCC-4]***. The NGSS Evidence Statements (Achieve 2015) define three key elements that are a part of every model: components, relationships, and connections. Systems have components that interact with one another (these interactions are called 'Relationships' in the NGSS Evidence Statements). Models can be applied to understanding phenomena and predicting the behavior of the overall system (these applications are called 'connections' in the NGSS Evidence Statements). One way to assess whether or not students have developed models of systems is to provide mediums for them to illustrate the mental models that are inside their heads. These mediums can be materials to make physical models or abstract representations such as pictorial models.

Assessment Snapshot 9.3: System Models in Middle and High School
Ms. P assigns her middle school students a task to draw a **model [SEP-2]** that illustrates the ***flow of energy [CCC-5]*** in an ecosystem (MS-LS2-3). Ms. P used to have students draw their models on a piece of paper, but she found that students really did not understand what a model was or how to represent it. She decided to use a computer tool to help scaffold the process, in this case the free MySystem tool (part of WISE, http://wise.berkeley.edu Students select different illustrations of objects that will act as components in the ***system [CCC-4]*** and drag them onto the workspace. Then, they make connections between the objects to represent interactions between the components. The tool requires that students describe these relationships with labels. Ms. P is able to distinguish between different levels of understanding by just glancing at the system diagrams (figure 9.6). Ms. P also finds that the labels of the relationships provide her particular direct insight into student mastery of DCIs. For example, a student that has built up a strong knowledge of DCIs labels a relationship "the captured energy is made to food in the chloroplast" while another says simply "flow.

Figure 9.6. Example Student Models of Energy Flow in an Ecosystem



Source: WISE 2015

Ms. P is trying to decide which rubric to use to score the models and is deciding between a simple holistic rubric (figure 9.7) and a criterion-based rubric (**Error! Reference source not found.**8). Neither rubric makes a distinction between the SEP and the DCIs or CCCs being assessed since successful completion of the item requires combined application of the three. While she likes the simplicity of the holistic rubric, she is worried that she will be inconsistent in its application.

Figure 9.7: Holistic Knowledge Integration System

6	**Systemic:** Students have a systemic understanding of science concepts.
5	**Complex:** Students understand how more than two science concepts interact in a given context.
4	**Basic:** Students understand how two scientific concepts interact in a given context.
3	**Partial:** Students recognize potential connections between concepts but cannot elaborate the nature of the connections specific to a given context.
2	**Isolated:** Students have relevant ideas but do not connect them in a given context.
1	**Irrelevant:** Students have irrelevant ideas in a given context.

Source: Technology Enhanced Learning in Science 2011.

She opts for the criterion-based rubric because it provides her students more specific feedback about where they can improve. Because it is more detailed, she decides to spend time introducing the rubric to her class and having them learn to score their peers' system models. While she finds that they are not able to reliably score one another (they have a hard time judging accuracy), she does feel that the exercise helps them focus on the key elements of a successful model. She has the students revise their models after their peer scoring and many make critical improvements.

Figure 9.8: Sample Criterion-based Rubric for System Models

	3	**2**	**1**
Components	All essential components of the system are included. The model does not include irrelevant components.	Major components of the situation are present, but smaller details are missing. --OR Extra components are included that are not appropriate to explain the phenomenon.	Omits one or more major components.
Relationships (arrows)	All components that interact are connected.	Some essential relationships are missing. -- OR Some components are incorrectly connected.	Major flaws exist in the way the components are connected in the diagram.
Relationships (labels)	Relationships are labeled with a clear description of the physical process that connects them.	Some of the labels are unclear or inaccurate.	Some labels are vague or missing.

Source: Table by M. d'Alessio

Commentary:

SEPs. In this task, students **develop a model [SEP-2]**. This prompt does not ask students to use or apply their model, which is a separate component of SEP-2 that would need to be assessed with another prompt.

DCIs. The rubric records DCI understanding as students use scientifically accurate components, relationships, and labels related to the cycling of matter and energy transfer in ecosystems (LS2.B).

CCCs. Students describe the interactions between components in a ***system [CCC-4]***. When looking more closely at the description of CCC-4 in Appendix 3, this task really probes systems at the elementary level. In middle school, students are expected to extend their understanding of systems to include systems made of interacting subsystems. This prompt could be extended to ask students to depict what goes on inside each of the organisms in the same diagram as the overall ecosystem.

In elementary grades, models might be simpler but should still emphasize the relationships between components. Figure 9.9 shows two student responses to the prompt, “Draw a model of a volcano formation at a hot spot using arrow to show movement in your model. Be sure to label all of the parts of your model.” Both models include labels of the components, but neither one effectively illustrates how the components relate to one another.

Figure 9.9. Example Student Models at the Elementary Level

***Commentary**:*

SEPs. Students **develop a model [SEP-2]** that illustrates the relationship between objects. In the example diagrams, all the relationships are spatial (an important aspect of a model). The prompt directs students to use arrows to show movement in the model. Assuming that students noticed this instruction, the absence of motion arrows in the examples likely indicates that students do not understand the cycling of Earth materials and how it relates to this context.

DCIs. This task goes beyond the elementary grade understanding of ESS2.B because elementary students are primarily expected to recognize patterns in volcanoes. They don’t link volcanoes to plate motions and the cycling of matter until middle school (ESS2.A).

CCCs. Students describe the interactions between components in a ***system [CCC-4]***.

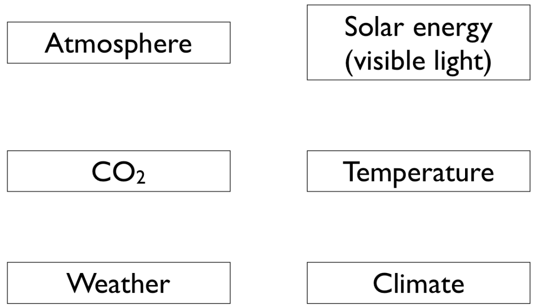
Source: NRC 2014.

At the high school level, students still struggle identifying interactions between components.

Figure 9.10 shows how an abstract system model can be used as a quick formative assessment to build this way of thinking.

Figure 9.10. Quick Formative Assessment of Systems in High School

Below are six different components of a simplified system. Draw arrows showing which components are related and add detailed labels of the relationships.



Commentary:

SEPs. Students **develop a model [SEP-2]** that illustrates the relationship between objects. At the high school level, students should be able to identify the components themselves, but this task is designed as a 'quick formative assessment' where the focus is on the relationships between components.

DCIs. This task asks students to articulate core ideas about Earth's energy budget and driving forces of weather and climate (ESS2.D).

CCCs. Students describe the interactions between components in a ***system [CCC-4]***.

Prompts with four to six components make easy warm up exercises and can be done individually or collaboratively.

Students must not only develop models, but they must use them to explain how a phenomena happen. In the NGSS Evidence Statements (Achieve 2015), PEs with SEP-2 include a 'Connections' section that articulates possible applications of the model. Teachers can use these 'Connections' to construct assessment items. For example, the evidence statement of HS-ESS2-4 says that students should be able to use their models describe the "net effect of all the competing factors in changing the climate." After developing the model in figure 9.10, teachers could prompt students, "In the first two decades of the 21st century, the amount of solar output went down slightly while the amount of CO_2

in the atmosphere went up dramatically. Use your model to explain what will happen to the planet's temperature."

Language is one avenue for formatively assessing student models because they must make their thinking public. A teacher might ask a student, "Can you explain your model to me?", turning an internal mental model into a conceptual model. This everyday usage of the word 'explain' is not the same as the NGSS practice of **constructing an explanation [SEP-6]**. Perhaps the teacher could use the phrase, "describe your model to me" to avoid such confusion. In this case, the description is a verbal representation of the model. Such verbal representations complement pictorial models when students present a diagram to the class and describe what it shows. The 'Connections' section in the evidence statements for PEs with SEP-2 often give guidance for what students should be able to describe. For example, in MS-LS2-3, students should be able to use their model to describe what happens "when organisms consume other organisms" and indicates that student responses should describe how "there is a transfer of energy and a cycling of atoms that were originally captured from the nonliving parts of the ecosystem by producers." After students develop the model in Figure 9.6, teachers could prompt students to, "use your model to describe what happens to atoms when an animal eats another organism."

Planning and Carrying Out Investigations

Investigations come in many different formats, so performance tasks related to investigations can be hands on or conducted entirely on computers. Technology-enhanced investigations can be contrived 'virtual labs,' realistic computer simulations, or investigations using digital data such as satellite imagery.

The important components of this SEP are that students start from an open scientific question and end with realistic data. While this process needs to be scaffolded to help move students along a developmental progression, by the end of grade twelve, students should be able to:

1. Start with an open ended scientific question and convert it into a scientifically testable question,
2. decide how to test it (considering appropriate scientific practices of repeatability and consistency),
3. decide what specific data need to be collected, and then
4. actually carry out the investigation.

Along the way, there are a number of formative assessment strategies that can provide practice and feedback for students at the key skill of planning.

Because carrying out investigations is so time consuming, formative assessment is especially important for planning investigations that are likely to succeed (though there is certainly a balance between letting students learn from their mistakes and helping them learn to avoid the mistakes). Specific strategies for formative assessment focus in on specific pieces of the planning process. To help students articulate the purpose of an investigation, they can select from a list of possible purpose statements, discussing their choice with peers (this strategy works even better if students can anonymously submit their own statements and then have the students select the best exemplars from their class). Students must identify the specific evidence that addresses the purpose of the investigation. They can decide which quantities can be measured and the appropriate tools to determine those ***quantities [CCC-3].*** A scaffolded approach could have students prepare blank data tables and graphs, or select the correct tables and graphs from options presented by the teacher. Students can predict the appropriate ***scale [CCC-3]*** for graph axes (before they even collect the data). They can begin to consider how they will **analyze and interpret the data [SEP-5]**. To plan procedures, students could write them up or sketch a storyboard. To make the task less open ended, students can be given a list of procedures in a mixed up order, identify intentional errors in a procedure provided to them, or write a brief justification for each step of a complete procedure presented to them. With each of these tasks, teachers can monitor progress and provide feedback.

Assessment Snapshot 9.4: Experimental Design for High School

Dr. S and Ms. H want to see if transitioning their high school science courses to NGSS-style student-driven investigations helps their students understand experimental design better. They recruit all the teachers in their department to administer a short one-page assessment to their students at the beginning and again at the end of the year about planning experiments (Figure 9.11). Some of their teachers are transitioning to NGSS already while some are using more traditional teaching techniques with recipe-style labs. At the end of the year, they blind score all the tests (using Table 3). Students that designed their own experiments throughout the year showed a much better ability to investigate a question about a health claim in the media. The two teachers share their results at a department meeting after school to encourage their colleagues and decide to read more about the developmental progression of experimental design and common preconceptions (Dasgupta, Anderson, and Pelaez 2014).

Source

Inspired by Sirum and Humburg 2011.

Commentary:

Effective rubrics for summative assessment in NGSS place development along a continuum of understanding. The binary checklist in Table 9.3 is not a good example of this, but it does serve as a good formative assessment of what specific subideas students consistently fail to remember or understand. Dr. S and Ms. H could identify specific aspects of experimental design that students consistently fail to include and then add or revise their lab activities to ensure that students learn these ideas.

SEPs. This rubric assesses the ability to **plan investigations [SEP-3].**

DCIs. This prompt does not require knowledge of DCIs.

CCCs. This snapshot emphasizes one of the Nature of Science CCCs, "Scientific Knowledge Assumes an Order and Consistency in Natural Systems." To measure understanding of this CCC along a continuum, a different rubric would be needed than the scoring checklist of table 9.3.

Figure 9.11. Experimental Design Ability Test

Pre-test prompt: Advertisements for an herbal product, ginseng, claim that it promotes endurance. To determine if the claim is fraudulent and prior to accepting this claim, what type of evidence would you like to see? Provide details of an investigative design.

Post-test prompt: The claim has been made that women may be able to achieve significant improvements in memory by taking iron supplements. To determine if the claim is fraudulent and prior to accepting this claim, what type of evidence would you like to see? Provide details of an investigative design.

Source: Sirum and Humburg 2011

Table 9.3. Experimental Design Ability Test Scoring Checklist

+1	Recognition that an experiment can be done to test the claim (vs. simply reading the product label).
+1	Identification of what variable is manipulated (independent variable is ginseng vs. something else).
+1	Identification of what variable is measured (dependent variable is endurance vs. something else).
+1	Description of how dependent variable is measured (e.g., how far subjects run will be measure of endurance).
+1	Realization that there is one other variable that must be held constant (vs. no mention).
+1	Understanding of the placebo effect (subjects do not know if they were given ginseng or a sugar pill).
+1	Realization that there are many variables that must be held constant (vs. only one or no mention).
+1	Understanding that the larger the sample size or # of subjects, the better the data.
+1	Understanding that the experiment needs to be repeated.
+1	Awareness that one can never prove a hypothesis, that one can never be 100 percent sure, that there might be another experiment that could be done that would disprove the hypothesis, that there are possible sources of error, that there are limits to generalizing the conclusions (credit for any of these).
/10	**Total**

Source: Sirum and Humburg 2011

Not all investigations are considered 'experiments' where parameters are varied or held constant and compared against controls. Large advances in science have come from purely observational investigations (including the mapping of the human genome, the discovery of planets around distant stars, and the recording of seismic waves that probe Earth's interior). An overemphasis on experimental design is not developmentally appropriate for the early grades when it may be more valuable to stress these curiosity-driven 'exploriments'. Teachers can even assess student attitudes towards science to see how well they are advancing the CCC that ***science is a human endeavor [CCC-NoS]*** using the Draw a Scientist test (Chambers 1983) or other validated survey.

Analyzing and Interpreting Data

Data are at the core of science. Analyzing and interpreting data can therefore be assessed alongside almost all the other SEPs. Students can use data to **explain [SEP-6]** *what* happened, to support an **argument [SEP-7]** about *why* it happened, and to predict what *will* happen (when combined with **models [SEP-2]** or **mathematical thinking [SEP-4]**). Students can **communicate [SEP-8]** using representations of data when data can be interpreted clearly (as in infographics), and **ask questions [SEP-1]** when they cannot.

Grammarians remind us that the word 'data' is plural, reflecting the fact that data are a collection of individual cases. To a scientist, each case has little meaning unless it is compared to the data as a whole. Seeing data as both its whole and its parts is a skill that students acquire over time. They learn to recognize ***trends and patterns [CCC-1]*** as well as individual cases that deviate from those patterns. Expert scientists have developed an internal 'library' of common data patterns (bell curves, exponential growth, linear trends, sine curves, etc.) that are each mentally linked to a set of tools for interpretation and physical processes that might ***cause [CCC-2]*** the pattern. Assessment allows teachers to determine where students are along the progression from a novice that only sees individual cases to an expert that fluidly sees the parts and the whole together.

Many of the skills for analyzing data at the early elementary level focus on helping students learn to record their observations, looking for ***patterns [CCC-1]*** in the observations, comparing observations and predictions. As students progress through the grades, they are able to deal with these same three skills in increasing complexity.

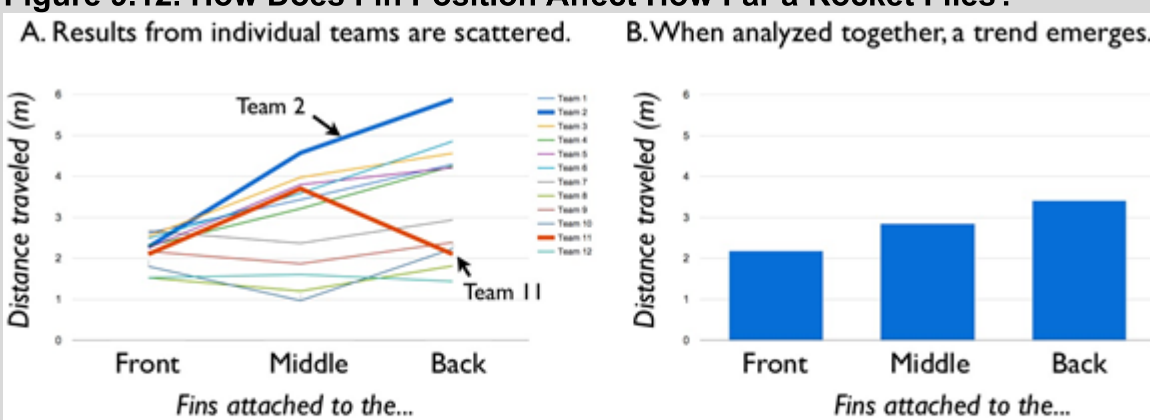
Data collected by students in the real world are messy. Imprecise measurement tools and impatient students often generate data that are too noisy to recognize the critical trends and patterns. Scientists need to collect enough data so that random errors cancel out, but classroom time for investigation is often limited. Technology can help solve some of these problems by providing ways for classes to quickly combine the data from multiple student groups and instantly display the results from all groups side-by-side. When students see their data in comparison to others, it prompts them to **ask questions [SEP-1]** about why results might differ from one another (d'Alessio and Lundquist 2013). Experts do this automatically, comparing new data to internal representations of how the data 'should' look, but students still benefit from external comparisons. When pooled together, patterns become clearer (Vandergon et al. 2016).

Assessment Snapshot 9.5: Analyzing Data for Upper Elementary
Mrs. L gives her fifth graders a design challenge to build small paper rockets launched by blowing into a straw. Their goal is to modify the rocket so that it travels as far as possible, which requires testing and iteration. Everyone receives a template for the same rocket body and same shape fins because researchers have found that using a common prototype as a starting point can lead to bolder innovations in classroom design projects (Sadler, Coyle, and Schwartz 2000). Before students begin their free design, Mrs. L presents a fictional dialog between students that highlights some of the decisions they will have to make about how the ***structure will enable the rocket's function [CCC-6].*** Amara: "The fins should go in the middle so it glides like an airplane." Brian: "No! They should go in the back like feathers on an arrow." Carrie: "Wings? Feathers? This is a rocket, not a bird! They should go in the

front so that they can help guide the rocket forward."

She asks students to **plan an investigation [SEP-3]** to figure out which student's idea works best (see figure 9.12). All teams in the class agree to systematically test the same rocket body with wings attached in three different positions. Mrs. L sets up an online form for them to submit their results. She projects a graph on the screen that will automatically display the results. It begins blank and Mrs. L asks students to sketch in their science notebooks what the graph would look like if Amara is correct, and then has them add the other two students. Students then perform their trials with the paper rockets and the graph updates with their data (**Error! Reference source not found.**A). Once all trials are complete, Mrs. L asks students if they can answer the original question posed by the Amara, Brian, and Carrie. A student from Team 2 uses the systematic progression in her team's data to agree with Brian, but a student from Team 11 says that her team found that Amara's suggestion worked best. Mrs. L is glad to see students using their data to support their arguments, but each student only uses data from his or her own team and does not examine the data as a whole (a common developmental stage). Students won't be required to calculate mean values until sixth grade (6.SP.5c), but students can relate to the 'middle' or average of a set of data. She asks students to come to the board to draw where they think the average is for each fin location in figure 9.12A. She invites classmates to call out 'higher' or 'lower' to get across that this method of determining averages is somewhat subjective. She informs the class that there is a simple way to **calculate [SEP-5]** the average, and that she set up the computer spreadsheet to do this automatically. She projects 9.12B and has students **compare their own visual estimate to the calculation [SEP-4]**. She asks teams to discuss what might have ***caused [CCC-2]*** their individual rockets to differ from the average. One student notices a ***pattern [CCC-1]*** that the results with the fins in the front are all pretty similar, but some rockets went a lot farther when the fins were in the back while others did not. The students want to know why but Mrs. L says, "I am impressed by your observations, but I don't really know the answer for sure." Mrs. L discussed the ideas of repeatability and variability and then asked students to revisit the possible causes of the differences. At the end of the activity, Mrs. L asks students to write an **argument [SEP-7]** using the sentence frames: "When I build my rocket, the best place to put the fins is _____ because ________. This position is better than the others because _____." She also asks students to sketch a graph of the data that supports their argument. A large number of students sketch something similar to Figure 9.12A and claim that fins should go in the middle or front, continuing to cite only their team's individual experience. Mrs. L. decides to find another activity for next week that further emphasizes the idea that combining large amounts of data can create a clearer picture.

Figure 9.12. How Does Fin Position Affect How Far a Rocket Flies?



Students submitted their results using an online form. During data collection, graph A projected on the screen. After student discussion of the variation between each trial, the teacher projected graph B that illustrates a clear trend. Graphs by M. d'Alessio

Commentary:

SEPs. The class discussion of the two graphs and the evidence students choose to include in their argument are Mrs. L's formative assessment of students' ability to **analyze data [SEP-4]**. In particular, the argument allows her to assess how well her students "use data to evaluate and refine design solutions" (appendix 3). She is trying to move them toward the ability to "Consider limitations of data analysis (e.g., measurement error), and/or seek to improve precision and accuracy of data with better technological tools and methods (e.g., multiple trials),” which is a middle school level of **data analysis [SEP-4]** (appendix 3). The example does not provide a rubric, but Mrs. L uses trends in the student arguments to add a new lesson that re-teaches the key idea that students missed about measurement error.

DCIs. In ETS1.B (Developing Possible Solutions), students should understand that tests are often designed to identify failure points or difficulties, which suggest the elements of the design that need to be improved. This task addresses ETS1.B but does not offer any assessment of it.

CCCs. Students “use graphs and charts to identify ***patterns [CCC-1]*** in data” (a middle school level understanding from appendix 3).

Resources

NASA Jet Propulsion Laboratory n.d.

Using Mathematics and Computational Thinking

Different aspects of mathematics and computational thinking pair with other SEPs and should therefore be assessed in tandem with those practices. For example, statistical thinking is important for **analyzing and interpreting data [SEP-4]**. Understanding measurement and units is a critical part of **planning and carrying out investigations [SEP-3**]. Understanding the application of computer simulations is part of **developing and using models [SEP-2]**.

Assessment Snapshot 9.6: Mathematical Thinking for Early Elementary

Mr. A’s kindergarten class is conducting an investigation when they realize that they need to use **mathematical thinking [SEP-5].** Mr. A’s class receives a package of silkworm eggs and is amazed how they all hatch on almost the same day! One student asks how quickly they will grow and another wonders how big they will get. The students decide that they would like to track the ***growth [CCC-7]*** of their silkworms and measure them daily. Mr. A wants the students to come up with a way to answer the question, “How ***big [CCC-3]*** are they today?” through a visual display of their measurement data. The students need to find a way to summarize all their measurements using a graphical display. Mr. A was guided by research about the different developmental levels in understanding how to display data (table 9.4).

Table 9.4. Developmental Levels of the Ability to Display Data

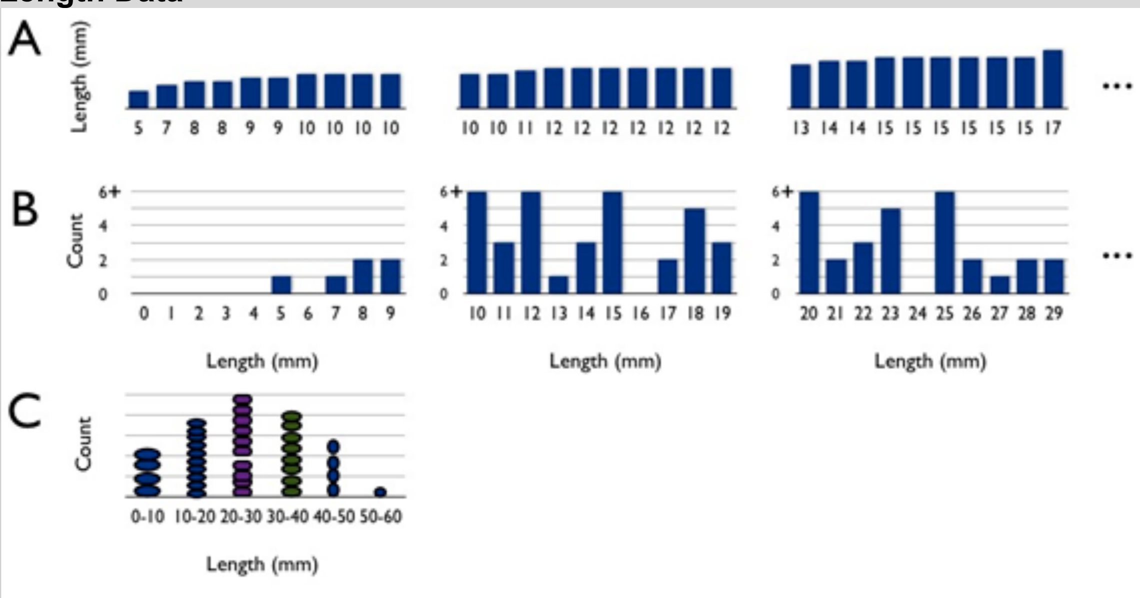
Level	Descriptor
6	Create and use data representations to notice trends, patterns, and be able to recognize outliers.
5	Create and use data representations that recognize scale as well as trends or patterns in data.
4	Represent data using groups of similar values and apply consistent scale to the groups.
3	Represent data using groups of similar values (though groups are inconsistent).
2	Identify the quantity of interest, but only consider each case as an individual without grouping data together
1	Group data in ways that don't relate to the problem of interest.

Source: Adapted from NRC 2014

One group orders each of the 261 measurements by magnitude, making a bar for each worm. The display uses a full 5 feet of wall space! (**Error! Reference source not found.**A; level 2 on **Error! Reference source not found.**4). Another group makes a bar graph with a bin size of just 1 mm per bin, which leads to 50 different bars (figure 9.13B; level 4 on **Error! Reference source not found.**4). Also, this group's vertical axis only extends to six worms at the top of the paper, so bars with more than six worms and got cut off. A third group creates a more traditional bar graph with measurements placed into bins. Rather than using bars, the group uses circles stacked one on top of the other. Unfortunately, different students draw the circles for each bin and they are not the same size and therefore not comparable (**Error! Reference source not found.**C; level 3 on **Error! Reference source not found.**4).

Mr. A leads a discussion about which representations are most useful for understanding silkworm growth. Mr. A recognizes that each representation is at a different developmental level and uses that understanding to highlight different concepts with different students (grouping versus consistent grouping, for example). As students **examine the graphs [SEP-5]** with better understanding of what they represent, they notice a ***pattern [CCC-1]*** that there are more 'medium sized' silkworms and fewer short or long ones (level 5 on table 9.4), which allows Mr. A to introduce the concept of variability. Students begin to ask questions about why some silkworms are growing so much faster than others. Mr. A's targeted guidance about how to represent data helped elevate the scientific discussion.

Figure 9.13. Facsimiles of Student-Created Representations of Silkworm Length Data



Groups A and B continue off to the right with additional pages.

Source: Adapted from Lehrer 2011.

Commentary:

SEPs. The emphasis of the rubric is on the ability to count and recognize similar values, examples of using **mathematical thinking [SEP-5]** at the primary level.

DCIs. While the activity supports the DCIs that plants and animals have unique and diverse lifecycles (LS1.B) and that individuals can vary in traits (LS3.B), the task does not assess student understanding of these DCIs.

CCCs. Students cannot complete this task without attention to ***scale and quantity [CCC-3]***, including the use of standard units to measure length. The rubric in table 9.4 emphasizes student ability to recognize ***patterns [CCC-1]*** as they create their data representations.

Resource:

Based on NRC 2014

Constructing Explanations

Students use evidence and reasoning based on DCIs to explain phenomena. Explanations are closely coupled with **models [SEP-2]**, and have some commonalities with scientific **arguments [SEP-7].** When students construct an explanation, they are often reporting about a conceptual model – phenomenon being explained can be thought of as an overall system property, and the interactions between components are part of the reasoning. As such, one strategy for formatively assessing explanations is to ask students to apply their conceptual models and report the results. Many of these questions can be presented as multiple choice items that call for high order conceptual thinking, often with distractors that probe for specific preconceptions. In a classroom, students can use colored index cards, personal white boards, clickers, or smartphone based apps to simultaneously report their thinking. After they report their initial answer, students discuss questions with small groups of peers and revote, if necessary. The technology students use to submit their choices is unimportant (Lasry 2008), but the peer discussion is very significant (Mazur 2009; McConnell et al. 2006). Students describe their thinking during these "assessment conversations" with one another and later with the teacher (Duschl and Gitomer 1997). These conversations often straddle the border between **argument [SEP-7]** and **explanation [SEP-6]** because students must defend their positions to peers and the teacher. In order to promote both these practices, questions must be higher order conceptual questions that require discussion of conceptual models, not simple recall. The American Association for the Advancement of Science (AAAS) maintains a library of conceptual items for all sciences (http://assessment.aaas.org/topics), and other organizations maintain specific archives for physics (Harvard, Interactive Learning Tooklkit, https://ilt.seas.harvard.edu/login/), earth science (SERC, ConcepTests, http://serc.carleton.edu/introgeo/interactive/ctestexm.html), and chemistry (ACS, Chemistry ConcepTests, http://www.jce.divched.org/JCEDLib/QBank/collection/ConcepTests/general.html) . Note: these databases are intended for college level instruction and age-appropriate questions will need to be selected.

Assessment Snapshot 9.7: ConcepTests for Explaining in Middle and High School
Students in Mrs. M's middle school class did a hands-on investigation of how sediment settles out from water to form layers (an example of process or '***function' determining structure [CCC-6]).*** She eventually wants them to be able to apply their model of layer formation to explain the extinction of the dinosaurs using accepted evidence from rock layers (MS-ESS1-4). She projects **Error! Reference source not found.** onto the screen and tells students, "We see this sequence of layers in the Earth. **Explain [SEP-6]** how they got to look the way they did. What processes happened and in what order? What's your evidence? If you think you have it figured out, answer the question about which layer is youngest." This is the first time she has ever shown them a problem like this. She has checked out a class set of iPads and she students click their answer on a free iPad app so that she sees a graph of their different responses updating in real-time. For this item, only 20 percent of the students offer the correct answer of F, with most students choosing A. Mrs. M anticipated that students may have missed a key concept and she has a contingent activity planned to help them understand a critical concept about layers that cut across other layers. She feels that they are ready to address the question and this time one third choose A and two thirds F. Students then pair up and discuss with the person next to them. She circulates around the class, listening to conversations. She then asks students to revote. Even though nearly 100 percent of the student responses are correct, she calls on specific students with some specific

questions, "Maria, you explained the whole geologic history to Lisa. Please repeat that briefly for us." After Maria, shares, Mrs. M continues with another inquiry, "Bryan, I was listening in and heard that you changed your thinking from A to F, and you had a really good reason that you told to Cliff. Please share how your explanation of the sequence changed." Mrs. M does not 'score' any of these items (including clicker responses), but she is implementing the assessment/instruction cycle many times during this simple interaction. Mrs. M constantly assesses and gives feedback to her students orally and adapts by delivering additional instruction on-the-fly or through planned contingency activities.

Mrs. M then provides additional information about the picture, indicating that layer C dates from 65 million years ago, the age of the dinosaurs and that layer F is evidence of a giant volcano nearby. She asks students to **construct an argument [SEP-8]** with their answer to the question: Could layer F's volcano be evidence of a volcanic eruption that wiped out the dinosaurs? After peer discussion, she has students write out a complete argument in their science notebooks that she will score with a rubric.

Figure 9.14. Example ConcepTest

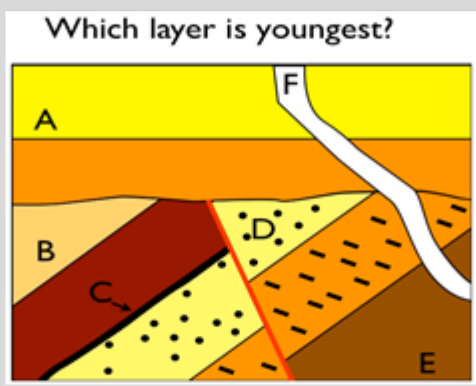


Diagram by M. d'Alessio

Commentary:

SEPs. This cross section of layers is a phenomena and students must **explain [SEP-6]** what caused this specific sequence. Students construct explanations in their conversations with one another, which the teacher listens to.

The multiple choice ConcepTest is primarily a frame that focuses these conversations, but it also provides instant feedback about common misconceptions that lead to flawed explanations (because the distractors in ConcepTests are specifically written or chosen to identify common misconceptions).

DCIs. This specific ConcepTest assesses students' ability to use rock strata to interpret the geologic history of an area (ESS1.C). To explain the relative position of different layers, students must apply knowledge of geoscience processes including as erosion and deposition (ESS2.C), the cycling of matter during volcanic eruptions (ESS2.A), and the motion of plates that causes rock layers to deform (ESS2.B).

CCCs. This specific ConcepTest assesses student understanding of the ***structure and function relationship [CCC-6]*** in geologic layers. Students cannot explain the structure without an understanding of the processes that cause these structures.

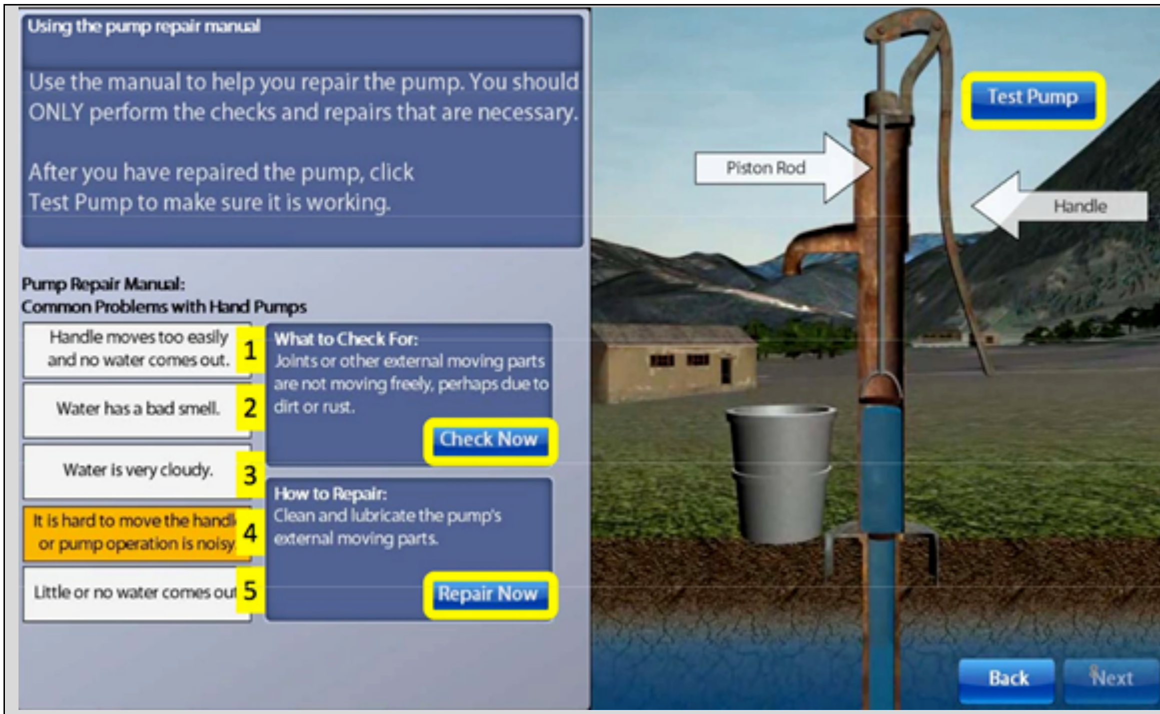
Designing Solutions

The practice of designing solutions is closely related to other SEPs through the stages of the engineering design process. The **designing solutions [SEP-6]** relies on **defining the problem [SEP-1]** and conducting **investigations [SEP-3]** to test the solutions. **Designing solutions [SEP-6]** also involves progressive iteration and refinement. Much like assessment of writing sometimes assigns value to how much students improved their writing from draft to draft, engineering design challenges can emphasize the iterative improvement of designs.

Assessment Snapshot 9.8: Designing Solutions for Middle and High School

Mrs. N wants her grade students to improve their iterative problem solving, an important part of **designing solutions [SEP-6].** Mrs. N introduces a performance task where students play the role of an engineer brought into a remote village to figure out why the local water well had stopped working. Mrs. N motivates the task: "Although we depend on plumbers, electricians, or car mechanics to help us when our technologies breakdown, we can be far more effective workers and citizens if we can fix at least some of our technologies ourselves." For this task, Mrs. N decides to assess **designing solutions [SEP-6]** separate from DCIs and she assumes that students have no prior knowledge of wells or hand pumps. An online instruction manual for the pump is embedded into the task, so the activity also assesses students' ability to **obtain information [SEP-8].** They use the manual to learn about the parts of the pump and create a mental **model [SEP-2]** for how the parts interact (figure 9.15). Mrs. N emphasizes that students will be able to develop a richer model if they consider how the shape and ***structure of each part relates to its function [CCC-6]*** or how each part acts like a component interacting with other parts as a ***system [CCC-4]***. Students then **perform investigations [SEP-3]** to gather evidence that help them isolate the pump's problems. The software gives them choices about how to troubleshoot the well (which is essentially testing for possible ***cause and effect relationships [CCC-2]).*** Since the task is self-paced within a computer, much of the feedback to students comes directly from the software program (automated formative assessment). When they choose a troubleshooting step that is not necessary, the computer invites them to determine why their choice was not the best one. Students end the computer task by developing a plan for maintaining the well that will prevent problems like this in the future. Mrs. N then has the students create a poster that **communicates [SEP-8]** their maintenance plan to villagers that may not speak English."

Figure 9.15. Sample Performance Task for Designing Solutions



Commentary:

SEPs. This published task is included because it illustrates how an interactive computer simulation can be used to assess an engineering challenge. This particular example emphasizes iterative problem solving, which is slightly different than iterative design refinement that is part of **designing solutions [SEP-6]**.

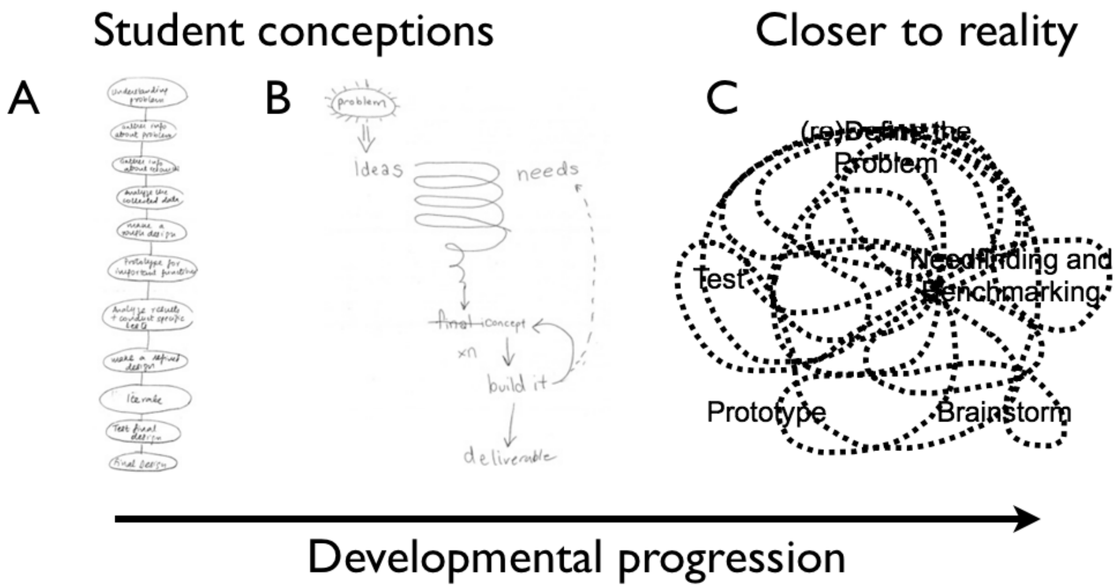
DCIs. This task involves an engineering DCI (ETS1.C: Optimizing Design Solutions) without coupling it to other content areas.

CCCs. Students must employ ***structure and function [CCC-6]***, ***systems [CCC-4]***, and ***cause and effect relationships [CCC-2]***, though this assessment has no explicit measurement of student understanding of these concepts.

Source
Adapted from the National Assessment of Educational Progress 2014.

Authentic engineering design has a built in assessment: since every engineering challenge has design constraints and criteria, teachers can assess student projects by whether or not they meet the criteria. While authentic, this approach fails to provide information about the developmental progression of skills. As students engage in engineering, their conception of the engineering design process progresses (figure 9.16) and they spend different amounts of time on each stage of the process (Atman et al. 2007). One formative assessment strategy is therefore to have students reflect on the different stages that they used during a design challenge.

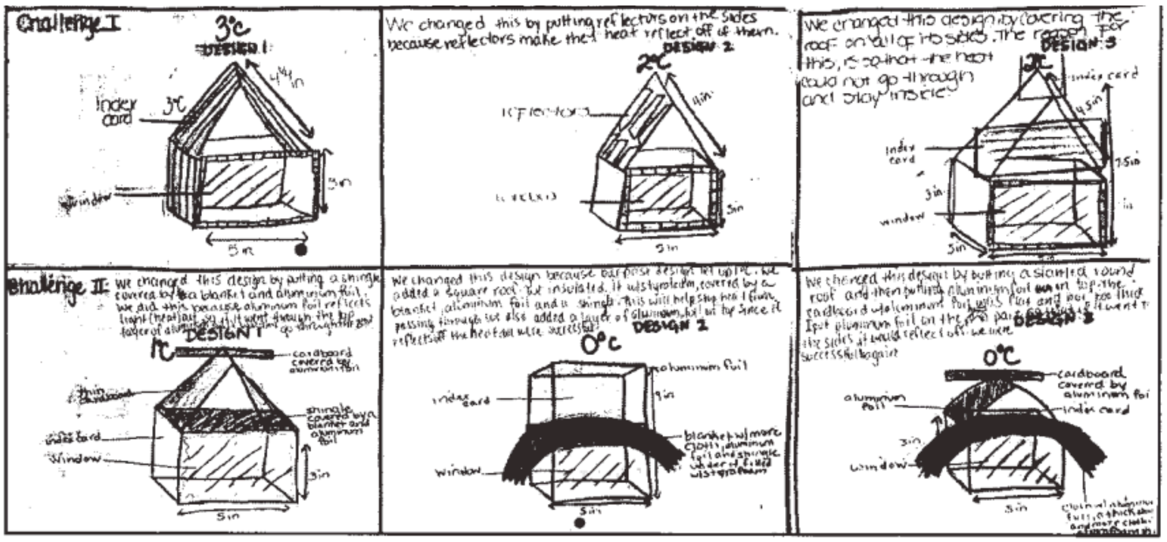
Figure 9.16. Developmental Progression of Conceptions of the Engineering Design Process



Student A conceives of the design process as a linear step while student B sees engineering as an iterative process. Both students are undergraduate engineering majors. Plot C is a theoretical illustration that more closely matches observations of practicing engineers. Source: Lande and Leifer 2010, Meinel and Leifer 2010.

As students work to iteratively improve their solutions, their testing and improvement strategies become more productive. Novices have trouble changing only a single variable during testing (Sadler, Coyle, and Schwartz 2000). Teachers can assess this ability by having students construct storyboards showing the evolution of their designs (Figure 9.12). A teacher can provide formative feedback by asking students to reflect on their drawings. Which change could they have done without? If they were to draw another frame, what test would they perform next? These diagrams are a powerful way for students to **communicate [SEP-8]** their solution design process.

Figure 9.12. Example Storyboard Illustrating Iterative Improvement during a Design Challenge



This diagram was produced by a pair of middle school students making a solar shelter. Source: Sadler, Coyle, and Schwartz 2000.

Engaging in Argument from Evidence

Arguments are the ‘currency’ used to exchange ideas in the scientific community. Over the course of their development, students learn how to formulate arguments that have value to the scientific community and practice evaluating arguments from others to determine if they have value and should be accepted. Arguments are, by definition, designed for external evaluation and are therefore more directly assessable than the related practice of **interpreting data [SEP-4]** (which can be entirely for private use to produce internal mental models).

Arguments can be broken down into three main components: a claim, evidence supporting the claim, and a chain of reasoning that links the evidence to the claim (fFigure 9.188; McNeill and Krajcik 2008)[2]. People internally base

[2] Claim-Evidence-Reasoning can also apply to **explanations [SEP-6]** where the claim is a description of how the phenomenon occurs.

their thoughts and decisions on evidence and prior knowledge about the way the world works, but they may not be consciously aware of those pieces. The "Claim, Evidence, Reasoning" framework helps students practice explicitly articulating what is initially automatic. Scientific communication relies on these components being presented publically so that they can be evaluated.

Figure 9.18. Graphic Organizer of a Claim, Evidence Reasoning Framework for Arguments

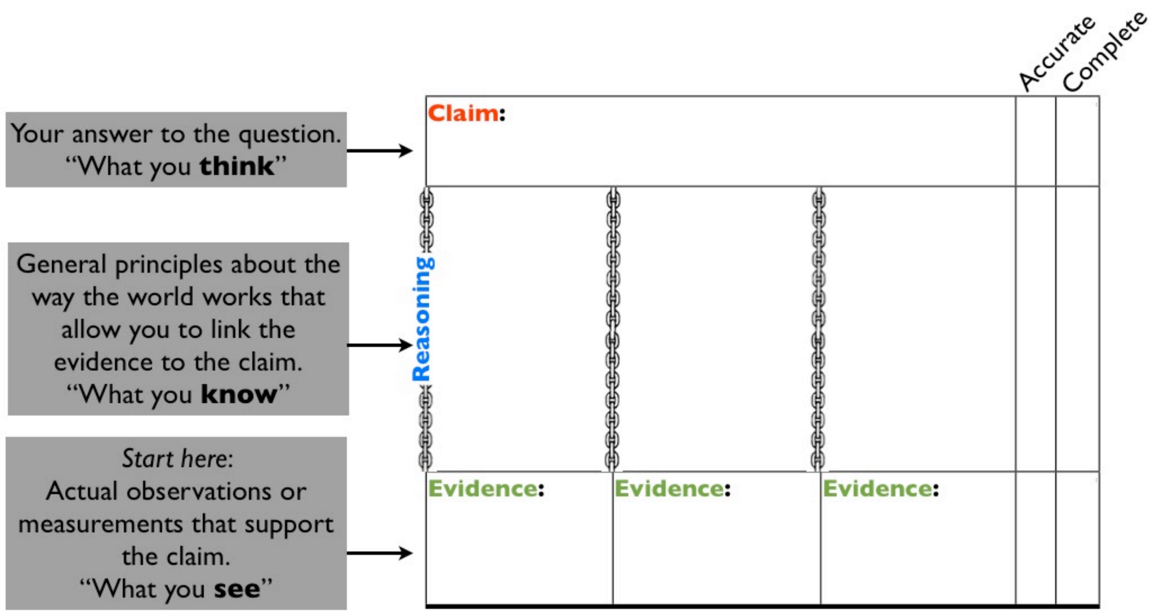


Diagram by M. d'Alessio

Scientists often evaluate arguments through the lens of crosscutting concepts: Does the data provide enough evidence to characterize a consistent ***pattern [CCC-1]***? Does the argument have sufficient evidence to justify a ***cause and effect relationship [CCC-2]***, or is the pattern just a simple correlation? Are there processes happening at a different ***scale [CCC-3]*** that the argument does not consider? Was the boundary of the ***system [CCC-4]*** chosen properly to encompass all the important interactions? Does the argument account for all the ***changes [CCC-6]*** with an appropriate ***flow of energy or matter [CCC-5]***? While scientists usually have discipline-specific ways of talking about them, the CCCs are essentially a generic checklist for evaluating the validity of an argument.

Assessing students' ability to construct or evaluate arguments can therefore draw on their understanding of CCCs.

McNeill and Krajcik (2008) suggest that the parts of a claim must be accurate, appropriate, and sufficient. Figure 9.18 has two columns on the right side for a 'checklist' to remind students of these features, though it combines the ideas of 'appropriate' and 'sufficient' into a single concept of 'complete'.

Table 5 illustrates one example of how these concepts can be evaluated for the three components of an argument. When teachers assess arguments, they often uncover student preconceptions that they can address through instruction. Deeply held student preconceptions are often at the root of inaccurate parts of an argument. Preconceptions can cloud perception so that students see evidence that isn't there (e.g., students claim that ice cubes will melt faster in saltwater than in freshwater and 'see' evidence to support that claim early in an experiment comparing the two while an objective observer cannot yet tell the difference in the melt rate). Similarly, students can use accurate evidence to support a misconception by generating faulty reasoning (e.g., a student claims that cats can see in the dark and has evidence that the cat's eyes appear to glow sometimes at night. The student wants to create a bridge from this evidence to their misconception and creates faulty reasoning that organisms see by producing light from their eyes). Asking students to explicitly spell out their evidence and reasoning exposes student beliefs to both teachers and students. According to conceptual change theory, students themselves need to be aware of their beliefs before they can modify them, and they won't change these ideas until they encounter new ideas that directly challenge them. Teachers, however, can design experiences that give students new evidence that specifically conflicts with those beliefs. When students have time to reflect on the conflict between an explicitly stated belief and new information, they are more likely to abandon a misconception. Formatively assessing arguments helps facilitate this process.

Table 9.5. Rubric for Scientific Arguments

	3	2	1
Claim (1 pt only)	X	X	Claim is scientifically correct and complete
Evidence	Provides appropriate and sufficient evidence to support claim.	Provides appropriate but insufficient evidence to support claim or also includes some inappropriate evidence.	Does not provide evidence, or only provides inappropriate evidence (evidence that does not support claim).
Reasoning (completeness)	All of the ideas necessary to link the evidence to the claim are included AND there are no "extra" ideas that are irrelevant to the claim.	Some attempt is made to relate evidence to underlying principles, but there are missing pieces or additional irrelevant pieces.	Does not provide reasoning, or only provides reasoning that does not link evidence to claim.
Reasoning (accuracy)	The evidence is tied to the claim by established scientific principles, AND there are no "extra" ideas that are incorrect.	The evidence is tied to the claim by established scientific principles, but there are also "extra" ideas that are incorrect.	The links between the evidence and the claim are based on incorrect ideas.

Reasoning receives the most weight in this rubric while the claim only receives one point out of ten. The rubric could be simplified for early elementary grades where selecting appropriate evidence is highlighted rather than reasoning. Source: Inspired by McNeill and Krajcik 2012.

Assessment Snapshot 9.9: Engaging in Argument during Science Talk for Elementary Students
Students in Mr. V's first grade class observed their shadow several times over the course of the day and also constructed a map of their schoolyard as part of their social studies work (CA History–Social Studies Content Standards 1.2.3). Mr. V presents students with a scenario: "The principal asked our class to find a good place on our schoolyard for a plant that needs sunlight in the morning and shade in the afternoon." Students examine their maps individually and come up with three ideas of where the plant could go and then discuss their proposals with a partner. Mr. V then gathers students around the classroom so that they all face one another in a circle for a "Science Talk" session where they will come to a consensus as a class about the best location. Students will use their DCI knowledge about shadows and ***patterns [CCC-1]*** of the Sun's movement (ESS1.A, ESS1.B) and construct **arguments using evidence [SEP-7]** that support specific **design solutions [SEP-6]**. Mr. V has prepared for the Science Talk by making a list of key concepts that he hopes students will mention and by reviewing the expectations about the practice of arguments in this grade span (*CA Science Framework* Appendix 3). Once students are quiet, Mr. V refers to a poster on the wall that shows the classroom norms for Science Talks. He reads to the key question and a sentence frame he has written on the board: "The plant should go _____. I think this because _____." He then invites students to share their ideas. During the discussion, Mr. V encourages students to talk to one another and not to him. He tries to speak as little as possible, intervening only to reinforce classroom norms and help maintain the focus. He also discretely keeps track of student contributions by taking notes on a simple checklist that provides him evidence of student mastery of the DCI and effective implementation of the practice. At the end of the session, he spends five minutes reflecting on patterns in what students said. On the back of his paper, he jots down a few ideas about what he will do during their next session to clarify problems. *(For more implementation about promoting discourse, see the Instructional*

Strategies Chapter.)

Commentary:

SEPs. Students engage in **argument [SEP-7]** where peers present competing arguments. Mr. V assesses the quality of the argument as he takes notes in his checklist.

DCIs. Students must integrate their knowledge of the systematic pattern of the Sun's movement across the sky during the course of a day (ESS1.A) and how certain objects cast shadows (PS4.B). Mr. V records student mastery of the DCIs in his checklist and notes common misconceptions during his reflection at the end of the session.

CCCs. Mr. V will need to be particularly attentive to how students are thinking about ***patterns [CCC-1]*** and ***stability and change [CCC-7]*** as he listens. Do students recognize that the Sun's position changes throughout the day, but that repeats a consistent pattern from one day to the next?

Adapted from an activity by Oakland Unified School District.

Obtaining, Evaluating, and Communicating Information

Obtaining information, evaluating it, and communicating it are all based on related competencies, but the specific behaviors are very different and need to be assessed differently. In elementary and middle school, the PEs that define the standards in the CA NGSS focus on obtaining and evaluating information, but generating communications products should be assessed in combination with the other practices in all grade bands. There is strong overlap between **evaluating information [SEP-8]** and **evaluating arguments [SEP-7]**, but to assess **evaluating information [SEP-8]**, teachers might include components of media literacy such as the ability to distinguish credible sources from less credible ones. Assessments of **communicating information [SEP-8]** may emphasizes criteria about the mechanics of written, oral, and visual communication, but should be assessed in parallel with other practices such as **scientific explanations [SEP-**

6] and **arguments [SEP-7]**. DCIs and CCCs can be assessed simultaneously with **communication [SEP-8]** by examining the content of the communications product.

Communication occurs in a range of media and modalities (including written text in both print and digital, oral communication, items that communicate visually such as drawings and graphs, and rich multimedia products). When the CA NGSS PEs incorporate **communications [SEP-8]**, they rarely specific the media in which competency must be demonstrated or that assessment must occur. The modalities teachers chose should be consistent with the vision of NGSS that students "engage in public discussions on science-related issues" and "be critical consumers of scientific information related to their everyday lives." (NRC 2012, 9). As such, teachers should assess using a range of modalities that go beyond classroom reading and writing and reflect the nature of 21st century communications such as panel discussions and debates, infographics, websites, social media, videos, etc.

While many of ELA/ELD strategies for assessing communication skills apply to science, the *NRC Framework* (NRC 2012) identifies several ways in which science communication is unique:

- *Science and engineering communications are "multimodal"* (they use an interconnected mix of words, diagrams, graphs, and mathematics). Teachers can assess how well students can relate these modalities by presenting students with a piece of information in one mode and asking them to produce complementary information in another. For example, students can be given a diagram and asked to write a text caption or select the most appropriate caption from a few examples. The Achieve (2015) evidence statements for high school suggest that a communication product does not demonstrate mastery of **communication [SEP-8]** unless it uses at least two modalities.
- *Science and engineering frequently use unfamiliar and specialized words* ('jargon'). The NRC (2000, 133) and American Association for the Advancement of Science (1993, 312) strongly discourage the

overemphasis on jargon and vocabulary in science education. Assessments that focus on the one dimensional understanding of vocabulary terms are *not* consistent with the goals of the CA NGSS. Students should be able to use and apply age-appropriate scientific vocabulary, but the assessment should usually be in the context of applications to other SEPs. If teachers specifically want to assess vocabulary, they can do so by having students rewrite a passage by eliminating scientific vocabulary and replacing it with everyday language (or to do the reverse).

- *In science and engineering, the details matter.* Students therefore need to pay constant attention to every word when obtaining scientific or engineering information. The process is sometimes complicated by a mismatch between the level of importance an idea has within the grammatical structure of a sentence and its importance for the scientific meaning of a sentence. For example, short introductory phrases and prepositions can have a dramatic impact on the scientific meaning of a sentence (e.g., 'assuming a frictionless surface'). Students must learn to read differently in order to notice all these pieces (CA CCSSM MP.6, CA CCSS for ELA/Literacy RI.3.4).

Assessment Snapshot 9.10: Communicating Information for Middle and High School
In the grade eight vignette 5.4 "Student Capstone Projects in chapter 5, Ms. S organizes a student capstone project where students document human impacts on Earth's systems. The project is very rich, so Ms. S needs an assessment strategy that will allow students to organize and present all their ideas. She decides to give students the experience of designing a website about their problem. It allows them to mix a wide variety of modalities including text and graphs, and even animations. Students must identify a specific purpose and target audience for their communication product. For example, the group studying a nearby stream decided that their target audience would be residents

of the neighborhood around the school. The team studying the school's energy consumption designed their site for the members of the student council. The students studying the possibility of deflecting an asteroid approaching the planet had seen a popular movie where the president ignored a scientist's claims about an oncoming asteroid until it was too late. They wanted to make their website useful to members of congress considering funding a new technology to protect the planet. After consulting the evidence statements for MS-ESS3-4, Ms. S integrates task-specific criteria into a generic rubric for project-based websites (table 9.6). This one rubric serves multiple purposes. The first two criteria are primarily for her classroom assessment to make sure that students have mastered key elements of the CCCs and DCIs. The intended purpose for the majority of the rubric scales is to provide her students specific feedback about website design, a skill that they are likely to use beyond this capstone project at the end of eighth grade.

Table 9.6. Rubric for a Website

Criterion	Beginning	Developing	Emerging	Mastering
Cause and Effect relationship [ESS2.A, ESS3.C] ***[CCC-2] [CCC-4]*** CA CCSS for ELA/Literacy RI.3	Describes the general functioning of Earth's systems but does not identify a clear cause and effect relationship related to human activities.	Accurately describes the relevant components of the Earth system and how they interact. Describes a cause and an effect, but fails to link them with coherent reasoning about interactions in the Earth system.	Accurately describes the relevant components of the Earth system and how they interact. Links a specific cause to a specific effect through coherent reasoning about interactions in the Earth system.	Accurately describes the relevant components of the Earth system and how they interact. Describes how specific human technologies cause changes to those the systems, and how technology can be used to mitigate, minimize, or reverse those

				changes.
Evidence and Interpretation *[CCC-2]* **[SEP-4]** **[SEP-7]** **[SEP-8]** CA CCSS for ELA/Literacy RI.1	No data or evidence are presented, the data are not reliable, or the data do not relate to the cause-effect relationship.	Accurate data and evidence are presented. The relationship between the data and the cause-effect relationship is not well defined.	Accurate data and evidence are presented and text explains how data are related to the cause-effect relationship.	Accurate data and evidence are clearly presented and text precisely and concisely explains how data are related to the cause-effect relationship. Data are sufficient to establish that there is a causal relationship and not just a correlation. Text argues against alternative interpretations of the data.
Target and Purpose **[SEP-8]**	Site lacks a sense of purpose. No indication that the site was created for a target audience other than teacher-as-grader.	Purpose may be somewhat unclear. Target audience is identified, and some choices are appropriate for this audience.	Site has a clear purpose. Major elements of the site are appropriate for the target audience.	Very strong understanding of who the site was created for. All elements of the site are engaging and appropriate for this audience.
Language and Conventions **[SEP-8]**	Errors in grammar and usage interfere with meaning. Many punctuation and spelling errors.	Errors in grammar and usage are noticeable, but do not interfere with meaning. Writing style is appropriate for	Few errors in grammar, usage, spelling or punctuation give clear evidence of careful editing.	Site has been fully edited to be free of errors in grammar, usage and mechanics. Writing style is deeply

	Writing style is not effective for the purpose. Site requires extensive editing.	the purpose.	Writing style is interesting and effective.	engaging.
Organization and Layout of Web Pages **[SEP-8]**	Layout and organization of pages is confusing, or cluttered or dull. Organization does not reflect ideas and content, but seems arbitrary.	Page layout may be 'busy' or unimaginative. Unreflective use of a template. Organization of pages does not obscure the content.	Page layout is interesting and appropriate for content. Layout and organization are appropriate for the content.	Page layout is creative and effective. Layout and organization helps provide structure to the ideas and content.
Credit and Sources ***[CCC-NoS]*** **[SEP-8]**	No reference of original sources. Information is copied without permission.	Sources of information are acknowledged. Most permissions have been secured. reference	Sources of information are credited in standard formats. All permissions are secured.	Sources of information are credited in standard formats. All permissions are secured and organized for future reference.

Source: Adapted from Galileo Educational Network n.d.

Commentary:

This is a rich assessment of a capstone project for all of middle school where the task requires students to integrate all three dimensions of the *CA NGSS*. In the evaluation of the task, some rubric criteria are one dimensional (especially those that focus on the mechanics of **communication [SEP-8]**), and some emphasize the integration of two dimensions at a time. Each criteria indicates the elements being assessed in the left column.

Conclusion

Assessments provide information to students about how well they are performing; to teachers about how well their students are learning and if modification to the instruction is necessary; to parents about their child's achievements; to districts about the effectiveness of instructional programs; and to policymakers about the effects of their policies. No single assessment can serve all these needs; an assessment system is needed to inform all stakeholders. The intent is to allow everyone within the educational system to make informed decisions regarding improved student learning, teacher development, instructional program modifications, and changes in policy (Popham 2000). The *CA NGSS* significantly alter the way science is taught in schools by making science education, grades K-12, resemble the way scientists work and think. Assessment must align with this vision, measuring not only what students know (DCIs), but how well they can generate new knowledge (SEPs), how well their knowledge relates to other understandings (CCCs), and how well they can combine these three dimensions together to understand phenomena and solve problems.

References

Achieve. 2015. NGSS Evidence Statements. http://www.nextgenscience.org/evidence-statements

American Association for the Advancement of Science. 1993. *Benchmarks for Science Literacy*. New York: Oxford University Press.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, and Joint Committee on Standards for Educational and Psychological Testing. 2014. *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.

Atman, Cynthia J., Robin S. Adams, Monica E. Cardella, Jennifer Turns, Susan Mosborg, and Jason Saleem. 2007. "Engineering Design Processes: A Comparison of Students and Expert Practitioners." *Journal of Engineering Education* 96 (4): 359–379.

Chambers, David Wade. 1983. "Stereotypic Images of the Scientist: The Draw-a-Scientist Test." *Science Education* 67 (2): 255–265.

d'Alessio, Matthew A. 2014. "What Kinds of Questions do Students Ask? Results from an Online Question Ranking Tool." *Electronic Journal of Science Education* 18 (5).

d'Alessio, Matthew A. and Loraine Lundquist. 2013. "Computer Supported Collaborative Rocketry: Teaching Students to Distinguish Good and Bad Data Like an Expert Physicist." *The Physics Teacher* 51 (7): 424–427.

Dasgupta, Annwesa P., Trevor R. Anderson, and Nancy Pelaez. 2014. "Development and Validation of a Rubric for Diagnosing Students' Experimental Design Knowledge and Difficulties." *CBE - Life Sciences Education* 13 (2): 265–284.

Duschl, Richard, and Drew Gitomer. 1997. "Strategies and Challenges to Changing the Focus of Assessment and Instruction in Science Classrooms." *Educational Assessment* 4 (1).

Galileo Educational Network. N.d. Inquiry and Assessment, Website Rubric. http://galileo.org/teachers/designing-learning/resources/inquiry-and-assessment/

Hamilton, Laura, and Brian Stecher. 2002. “Improving Test-Based Accountability.” In *Making Sense of Test Based Accountability in Education*, edited by Laura Hamilton, Brian Stecher, and Stephen Klein. Santa Monica, CA: RAND.

Herman, Joan L. 2010. *Coherence: Key to Next Generation Assessment Success (AACC Report)*. Los Angeles, CA: University of California.

Herman, Joan L., and Margaret Heritage. 2007. “Moving from Piecemeal to Effective Formative Assessment Practice: Moving Pictures on the Road to Student Learning.” Presentation given at the Council of Chief State School Officers (CCSSO) National Conference on Large Scale Assessment, Session 143. Nashville, TN.

Jirout, Jamie, and David Klahr. 2011. *Children’s Question Asking and Curiosity: A Training Study*. Evanston, IL: Society for Research on Educational Effectiveness.

Lasry, N. 2008. “Clickers or Flashcards: Is There Really a Difference?” *The Physics Teacher* 46 (4): 242–244.

Lande, Micah, and Leifer, Larry 2010. “Difficulties Student Engineers Face Designing the Future.” *International Journal of Engineering Education* 26 (2): 271–277.

Lehrer, Richard. 2011. “Learning to Reason about Variability and Chance by Inventing Measures and Models.” Presentation given at the National Association for Research in Science Teaching, Orlando, FL.

Mazur, Eric. 2009. “Farewell, Lecture?” *Science* 323 (5910): 50–51.

McConnell, David A., David N. Steer, Katharine D. Owens, Jeffrey R. Knott, Stephen Van Horn, Walter Borowski, Jeffrey Dick, Annabelle Foos, Michelle Malone, Heidi McGrew, Lisa Greer, and Peter J. Heaney. 2006. “Using ConcepTests to Assess and Improve Student Conceptual Understanding in Introductory Geoscience Courses.” *Journal of Geoscience Education* 54 (1): 61–68.

McNeill, Katherine L., and Joseph Krajcik. 2008. “Inquiry and Scientific Explanations: Helping Students use Evidence and Reasoning.” In *Science as Inquiry in the Secondary Setting*, edited by Julie Luft, Randy Bell, and Julie Gess-Newsome. Arlington, VA: National Science Teachers Association Press.

Meinel, Cristoph and Larry J. Leifer. 2010. "Design Thinking Research." In *Design Thinking: Understand–Improve–Apply*, edited by Hasso Plattner, Cristoph Meinel, and Larry J. Leifer. Heidelberg, Germany: Springer.

NASA Jet Propulsion Laboratory. N.d. Soda-Straw Rockets. http://www.jpl.nasa.gov/edu/teach/activity/straw-rocket/.

National Assessment of Educational Progress. 2014. Sample TEL Task. https://nces.ed.gov/nationsreportcard/tel/wells_item.aspx .

National Research Council (NRC). 2000. *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning*. Washington, D.C.: The National Academies Press.

———. 2012. *A Framework for K–12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, D.C.: The National Academies Press.

———. 2014. *Developing Assessments for the Next Generation Science Standards.* Washington, D.C.: The National Academies Press.

Sadler, D. Royce. 1989. "Formative Assessment and the Design of Instructional Strategies." *Instructional Science* 18: 119–144.

Sadler, Philip M., Harold P. Coyle, and Marc Schwartz. 2000. "Engineering Competitions in the Middle School Classroom: Key Elements in Developing Effective Design Challenges." *The Journal of the Learning Sciences* 9 (3): 299–327.

Sirum, Karen, and Jennifer Humburg. 2011. "The Experimental Design Ability Test (EDAT)." *Bioscene: Journal of College Biology Teaching* 37 (1): 8–16.

State Board of Education. 2016. *Proposed Design for California's Next Generation Science Standards General Summative Assessments***.** February 16, 2016. http://www.cde.ca.gov/be/ag/ag/yr16/documents/mar16item02.doc

Technology Enhanced Learning in Science. 2011. Sample Scoring Rubric Using the Knowledge Integration Framework. http://telscenter.org/projects/tels/assessments/rubric.

Vandergon, Virginia Oberholzer, John Reveles, Norman Herr, Dorothy Nguyen-Graff, Mike Rivas, Matthew d'Alessio, and Brian Foley. 2016. "Engaging Students in Conducting Data Analysis: The Whole-Class Data Advantage." In *Handbook of Research on Cloud-Based STEM Education for Improved Learning Outcomes*, edited by Lee Chao. Hershey, PA: IGI Global.

WISE. 2015. WISE Overview. http://wise4.org/wise-overview.pdf

Wiliam, Dylan. 2006. “Formative Assessment: Getting the Focus Right.” *Educational Assessment: Special Issue* 11 (3-4): 265–294.