DEVELOPING ASSESSMENTS FOR THE NEXT GENERATION SCIENCE STANDARDS

Committee on Developing Assessments of Science Proficiency in K-12
James W. Pellegrino, Mark R. Wilson, Judith A. Koenig, and Alexandra S. Beatty, Editors

Board on Testing and Assessment
Board on Science Education
Division of Behavioral and Social Sciences and Education

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

Copyright © National Academy of Sciences. All rights reserved.
THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. C. D. Mote, Jr., is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. C. D. Mote, Jr., are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org
4

CLASSROOM ASSESSMENT

Assessments can be classified in terms of the way they relate to instructional activities. The term classroom assessment (sometimes called internal assessment) is used to refer to assessments designed or selected by teachers and given as an integral part of classroom instruction. They are given during or closely following an instructional activity or unit. This category of assessments may include teacher-student interactions in the classroom, observations, student products that result directly from ongoing instructional activities (called “immediate assessments”), and quizzes closely tied to instructional activities (called “close assessments”). They may also include formal classroom exams that cover the material from one or more instructional units (called “proximal assessments”). 1 This category may also include assessments created by curriculum developers and embedded in instructional materials for teacher use.

In contrast, external assessments are designed or selected by districts, states, countries, or international bodies and are typically used to audit or monitor learning. External assessments are usually more distant in time and context from instruction. They may be based on the content and skills defined in state or national standards, but they do not necessarily reflect the specific content that was covered in any particular classroom. They are typically given at a time that is determined by administrators, rather than by the classroom teacher. This category includes such assessments as the statewide science tests required by the No Child

1This terminology is drawn from Ruiz-Primo et al. (2002) and Pellegrino (2013).
Left Behind Act or other accountability purposes (called “distal assessments”), as well as national and international assessments: the National Assessment of Educational Progress and the Programme for International Student Assessment (called “remote assessments”). Such external assessments and their monitoring function are the subject of the next chapter.

In this chapter, we illustrate the types of assessment tasks that can be used in the classroom to meet the goals of A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas (National Research Council, 2012a, hereafter referred to as “the framework”) and the Next Generation Science Standards: For States, By States (NGSS Lead States, 2013). We present example tasks that we judged to be both rigorous and deep probes of student capabilities and also to be consistent with the framework and the Next Generation Science Standards (NGSS). We discuss external assessments in Chapter 5 and the integration of classroom and external assessments into a coherent system in Chapter 6. The latter chapter argues that an effective assessment system should include a variety of types of internal and external assessments, with each designed to fulfill complementary functions in assessing achievement of the NGSS performance objectives.

Our starting point for looking in depth at classroom assessment is the analysis in Chapter 2 of what the new science framework and the NGSS imply for assessment. We combine these ideas with our analysis in Chapter 3 of current approaches to assessment design as we consider key aspects of classroom assessment that can be used as a component in assessment of the NGSS performance objectives.

**ASSESSMENT PURPOSES: FORMATIVE OR SUMMATIVE**

Classroom assessments can be designed primarily to guide instruction (formative purposes) or to support decisions made beyond the classroom (summative purposes). Assessments used for formative purposes occur during the course of a unit of instruction and may involve both formal tests and informal activities conducted as part of a lesson. They may be used to identify students’ strengths and weaknesses, assist educators in planning subsequent instruction, assist students in guiding their own learning by evaluating and revising their own work, and foster students’ sense of autonomy and responsibility for their own learning (Andrade and Cizek, 2010, p. 4). Assessments used for summative purposes may be administered at the end of a unit of instruction. They are designed to provide evidence of achievement that can be used in decision making, such as assigning grades; making promotion
or retention decisions; and classifying test takers according to defined performance categories, such as “basic,” “proficient,” and “advanced” (levels often used in score reporting) (Andrade and Cizek, 2010, p. 3).

The key difference between assessments used for formative purposes and those used for summative purposes is in how the information they provide is to be used: to guide and advance learning (usually while instruction is under way) or to obtain evidence of what students have learned for use beyond the classroom (usually at the conclusion of some defined period of instruction). Whether intended for formative or summative purposes, evidence gathered in the classroom should be closely linked to the curriculum being taught. This does not mean that the assessment must use the formats or exactly the same material that was presented in instruction, but rather that the assessment task should directly address the concepts and practices to which the students have been exposed.

The results of classroom assessments are evaluated by the teacher or sometimes by groups of teachers in the school. Formative assessments may also be used for reflection among small groups of students or by the whole class together. Classroom assessments can play an integral role in students’ learning experiences while also providing evidence of progress in that learning. Classroom instruction is the focus of the framework and the NGSS, and it is classroom assessment—which by definition is integral to instruction—that will be the most straightforward to align with NGSS goals (once classroom instruction is itself aligned with the NGSS).

Currently, many schools and districts administer benchmark or interim assessments, which seem to straddle the line between formative and summative purposes (see Box 4-1). They are formative in the sense that they are used for a diagnostic function intended to guide instruction (i.e., to predict how well students are likely to do on the end-of-year tests). However, because of this purpose, the format they use resembles the end-of-year tests rather than other types of internal assessments commonly used to guide instruction (such as quizzes, classroom dialogues, observations, or other types of immediate assessment strategies that are closely connected to instruction). Although benchmark and interim assessments serve a purpose, we note that they are not the types of formative assessments that we discuss in relation to the examples presented in this chapter or that are advocated by others (see, e.g., Black and Wiliam, 2009; Heritage, 2010; Perie et al., 2007). Box 4-1 provides additional information about these types of assessments.
Currently, many schools and districts administer benchmark or interim assessments, which they treat as formative assessments. These assessments use tasks that are taken from large-scale tests given in a district or state or are very similar to tasks that have been used in those tests. They are designed to provide an estimate of students’ level of learning, and schools use them to serve a diagnostic function, such as to predict how well students will do on the end-of-year tests.

Like the large-scale tests they closely resemble, benchmark tests rely heavily on multiple-choice items, each of which tests a single learning objective. The items are developed to provide only general information about whether students understand a particular idea, though sometimes the incorrect choices in a multiple-choice item are designed to probe for particular common misconceptions. Many such tasks would be needed to provide solid evidence that students have met the performance expectations for their grade level or grade band.

Teachers use these tests to assess student knowledge of a particular concept or a particular aspect of practice (e.g., control of variables), typically after teaching a unit that focuses on specific discrete learning objectives. The premise behind using items that mimic typical large-scale tests is that they help teachers measure students’ progress toward objectives for which they and their students will be held accountable and provide a basis for deciding which students need extra help and what the teacher needs to teach again.

**CHARACTERISTICS OF NGSS-ALIGNED ASSESSMENTS**

Chapter 2 discusses the implications of the NGSS for assessment, which led to our first two conclusions:

- Measuring the three-dimensional science learning called for in the framework and the Next Generation Science Standards requires assessment tasks that examine students’ performance of scientific and engineering practices in the context of crosscutting concepts and disciplinary core ideas. To adequately cover the three dimensions, assessment tasks will generally need to 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 (Conclusion 2-1).
The Next Generation Science Standards require that assessment tasks be designed so that they can accurately locate students along a sequence of progressively more complex understandings of a core idea and successively more sophisticated applications of practices and crosscutting concepts (Conclusion 2-2).

Students will likely need repeated exposure to investigations and tasks aligned to the framework and the NGSS performance expectations, guidance about what is expected of them, and opportunities for reflection on their performance to develop these proficiencies, as discussed in Chapter 2. The kind of instruction that will be effective in teaching science in the way the framework and the NGSS envision will require students to engage in science and engineering practices in the context of disciplinary core ideas—and to make connections across topics through the crosscutting ideas. Such instruction will include activities that provide many opportunities for teachers to observe and record evidence of student thinking, 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 (possibly in a personal science journal).

The products of such instruction form a natural link to the characteristics of classroom assessment that aligns with the NGSS. We highlight four such characteristics:

1. the use of a variety of assessment activities that mirror the variety in NGSS-aligned instruction;
2. tasks that have multiple components so they can yield evidence of three-dimensional learning (and multiple performance expectations);
3. explicit attention to the connections among scientific concepts; and
4. the gathering of information about how far students have progressed along a defined sequence of learning.

**Variation in Assessment Activities**

Because NGSS-aligned instruction will naturally involve a range of activities, classroom assessment that is integral to instruction will need to involve a corresponding variation in the types of evidence it provides about student learning. Indeed, the distinction between instructional activities and assessment activities may be blurred, particularly when the assessment purpose is formative. A classroom
assessment may be based on 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.

Science and engineering practices lend themselves well to assessment activities that can provide this type of evidence. For instance, when students are developing and using models, they may be given the opportunity to explain their models and to discuss them with classmates, thus providing the teacher with an opportunity for formative assessment reflection (illustrated in Example 4, below). Student discourse can give the teacher a window into students' thinking and help to guide lesson planning. A classroom assessment may also involve a formal test or diagnostic quiz. Or it may be based on artifacts that are the products of classroom activities, rather than on tasks designed solely for assessment purposes. These artifacts may include student work produced in the classroom, homework assignments (such as lab reports), a portfolio of student work collected over the course of a unit or a school year (which may include both artifacts of instruction as well as results from formal unit and end-of-course tests), or activities conducted using computer technology. A classroom assessment may occur in the context of group work or discussions, as long as the teacher ensures that all the students that need to be observed are in fact active participants. Summative assessments may also take a variety of forms, but they are usually intended to assess each student's independent accomplishments.

**Tasks with Multiple Components**

The NGSS performance expectations each blend a practice and, in some cases, also a crosscutting idea with an aspect of a particular core idea. In the past, assessment tasks have typically focused on measuring students' understanding of aspects of core ideas or of science practices as discrete pieces of knowledge. Progression in learning was generally thought of as knowing more or providing more complete and correct responses. Similarly, practices were intentionally assessed in a way that minimized specific content knowledge demands—assessments were more likely to ask for definitions than for actual use of the practice. Assessment developers took this approach in part to be sure they were obtaining accurate measures of clearly definable constructs. However, although understanding the language and termi-

---

2“A construct” is generally used to refer to concepts or ideas that cannot be directly observed, such as “liberty.” In the context of educational measurement, the word is used more specifically to refer to a particular body of content (knowledge, understanding, or skills) that an assessment
nology of science is fundamental and factual knowledge is very important, tasks that demand only declarative knowledge about practices or isolated facts would be insufficient to measure performance expectations in the NGSS.

As we note in Chapter 3, the performance expectations provide a start in defining the claim or inference that is to be made about student proficiency. However, it is also important to determine the observations (the forms of evidence in student work) that are needed to support the claims, and then to develop tasks or situations that will elicit the needed evidence. The task development approaches described in Chapter 3 are commonly used for developing external tests, but they can also be useful in guiding the design of classroom assessments. Considering the intended inference, or claim, about student learning will help curriculum developers and classroom assessment designers ensure that the tasks elicit the needed evidence.

As we note in Chapter 2, assessment tasks aligned with the NGSS performance expectations will need to have multiple components—that is, be composed of more than one kind of activity or question. They will need to include opportunities for students to engage in practices as a means to demonstrate their capacity to apply them. For example, a task designed to elicit evidence that a student can develop and use models to support explanations about structure-function relationships in the context of a core idea will need to have several components. It may require that students articulate a claim about selected structure-function relationships, develop or describe a model that supports the claim, and provide a justification that links evidence to the claim (such as an explanation of an observed phenomenon described by the model). A multicomponent task may include some short-answer questions, possibly some carefully designed selected-response questions, and some extended-response elements that require students to demonstrate their understandings (such as tasks in which students design an investigation or explain a pattern of data). For the purpose of making an appraisal of student learning, no single piece of evidence is likely to be sufficient; rather, the pattern of evidence across multiple components can provide a sufficient indicator of student understanding.

is to measure. It can be used to refer to a very specific aspect of tested content (e.g., the water cycle) or a much broader area (e.g., mathematics).
Making Connections

The NGSS emphasize the importance of the connections among scientific concepts. Thus, the NGSS performance expectations for one disciplinary core idea may be connected to performance expectations for other core ideas, both within the same domain or in other domains, in multiple ways: one core idea may be a prerequisite for understanding another, or a task may be linked to more than one performance expectation and thus involve more than one practice in the context of a given core idea. NGSS-aligned tasks will need to be constructed so that they provide information about how well students make these connections. For example, a task that focused only on students' knowledge of a particular model would be less revealing than one that probed students' understanding of the kinds of questions and investigations that motivated the development of the model. Example 1, “What Is Going on Inside Me?” (in Chapter 2), shows how a single assessment task can be designed to yield evidence related to multiple performance expectations, such as applying physical science concepts in a life science context. Tasks that do not address these connections will not fully capture or adequately support three-dimensional science learning.

Learning as a Progression

The framework and the NGSS address the process of learning science. They make clear that students should be encouraged to take an investigative stance toward their own and others’ ideas, to be open about what they are struggling to understand, and to recognize that struggle as part of the way science is done, as well as part of their own learning process. Thus, revealing students' emerging capabilities with science practices and their partially correct or incomplete understandings of core ideas is an important function of classroom assessment. The framework and the NGSS also postulate that students will develop disciplinary understandings by engaging in practices that help them to question and explain the functioning of natural and designed systems. Although learning is an ongoing process for both scientists and students, students are emerging practitioners of science, not scientists, and their ways of acting and reasoning differ from those of scientists in important ways. The framework discusses the importance of seeing learning as a trajectory in which students gradually progress in the course of a unit or a year, and across the whole K-12 span, and organizing instruction accordingly.

The first example in this chapter, “Measuring Silkworms” (also discussed in Chapter 3), illustrates how this idea works in an assessment that is embedded in a larger instructional unit. As they begin the task, students are not competent data
analysts. They are unaware of how displays can convey ideas or of professional conventions for display and the rationale for these conventions. In designing their own displays, students begin to develop an understanding of the value of these conventions. Their partial and incomplete understandings of data visualization have to be explicitly identified so teachers can help them develop a more general understanding. Teachers help students learn about how different mathematical practices, such as ordering and counting data, influence the shapes the data take in models. The students come to understand how the shapes of the data support inferences about population growth.

Thus, as discussed in Chapter 2, uncovering students' incomplete forms of practice and understanding is critical: NGSS-aligned assessments will need to clearly define the forms of evidence associated with beginning, intermediate, and sophisticated levels of knowledge and practice expected for a particular instructional sequence. A key goal of classroom assessments is to help teachers and students understand what has been learned and what areas will require further attention. NGSS-aligned assessments will also need to identify likely misunderstandings, productive ideas of students that can be built upon, and interim goals for learning.

The NGSS performance expectations are general: they do not specify the kinds of intermediate understandings of disciplinary core ideas students may express during instruction nor do they help teachers interpret students' emerging capabilities with science practices or their partially correct or incomplete understanding. To teach toward the NGSS performance expectations, teachers will need a sense of the likely progression at a more micro level, to answer such questions as:

- For this unit, where are the students expected to start, and where should they arrive?
- What typical intermediate understandings emerge along this learning path?
- What common logical errors or alternative conceptions present barriers to the desired learning or resources for beginning instruction?
- What new aspects of a practice need to be developed in the context of this unit?

Classroom assessment probes will need to be designed to generate enough evidence about students' understandings so that their locations on the intended pathway can be reliably determined, and it is clear what next steps (instructional activities) are needed for them to continue to progress. As we note in Chapter 2,
only a limited amount of research is available to support detailed learning progressions: assessment developers and others who have been applying this approach have used a combination of research and practical experience to support depictions of learning trajectories.

SIX EXAMPLES

We have identified six example tasks and task sets that illustrate the elements needed to assess the development of three-dimensional science learning. As noted in Chapter 1, they all predate the publication of the NGSS. However, the constructs being measured by each of these examples are similar to those found in the NGSS performance expectations. Each example was designed to provide evidence of students’ capabilities in using one or more practices as they attempt to reach and present conclusions about one or more core ideas: that is, all of them assess three-dimensional learning. Table 1-1 shows the NGSS disciplinary core ideas, practices, and crosscutting ideas that are closest to the assessment targets for all of the examples in the report.³

We emphasize that there are many possible designs for activities or tasks that assess three-dimensional science learning—these six examples are only a sampling of the possible range. They demonstrate a variety of approaches, but they share some common attributes. All of them require students to use some aspects of one or more science and engineering practices in the course of demonstrating and defending their understanding of aspects of a disciplinary core idea. Each of them also includes multiple components, such as asking students to engage in an activity, to work independently on a modeling or other task, and to discuss their thinking or defend their argument.

These examples also show how one can use classroom work products and discussions as formative assessment opportunities. In addition, several of the examples include summative assessments. In each case, the evidence produced provides teachers with information about students’ thinking and their developing understanding that would be useful for guiding next steps in instruction. Moreover, the time students spend in doing and reflecting on these tasks should

³The particular combinations in the examples may not be the same as NGSS examples at that grade level, but each of these examples of classroom assessment involves integrated knowledge of the same general type as the NGSS performance expectations. However, because they predate the NGSS and its emphasis on crosscutting concepts, only a few of these examples include reference to a crosscutting concept, and none of them attempts to assess student understanding of, or disposition to invoke, such concepts.
be seen as an integral part of instruction, rather than as a stand-alone assessment task. We note that the example assessment tasks also produce a variety of products and scorable evidence. For some we include illustrations of typical student work, and for others we include a construct map or scoring rubric used to guide the data interpretation process. Both are needed to develop an effective scoring system.

Each example has been used in classrooms to gather information about particular core ideas and practices. The examples are drawn from different grade levels and assess knowledge related to different disciplinary core ideas. Evidence from their use documents that, with appropriate prior instruction, students can successfully carry out these kinds of tasks. We describe and illustrate each of these examples below and close the chapter with general reflections about the examples, as well as our overall conclusions and recommendations about classroom assessment.

**Example 3: Measuring Silkworms**

The committee chose this example because it illustrates several of the characteristics we argue an assessment aligned with the NGSS must have; in particular, it allows the teacher to place students along a defined learning trajectory (see Figure 3-13 in Chapter 3), while assessing both a disciplinary core idea and a crosscutting concept.\(^4\) The assessment component is formative, in that it helps the teacher understand what students already understood about data display and to adjust the instruction accordingly. This example, in which 3rd-grade students investigated the growth of silkworm larvae, first assesses students’ conceptions of how data can be represented visually and then engages them in conversations about what different representations of the data they had collected reveal. It is closely tied to instruction—the assessment is embedded in a set of classroom activities.

The silkworm scenario is designed so that students’ responses to the tasks can be interpreted in reference to a trajectory of increasingly sophisticated forms of reasoning. A construct map displayed in Figure 3-13 shows developing conceptions of data display. Once the students collect their data (measure the silkworms) and produce their own ways of visually representing their findings, the teacher uses the data displays as the basis for a discussion that has several objectives.

---

\(^4\)This example is also discussed in Chapter 3 in the context of using construct modeling for task design.
The teacher uses the construct map to identify data displays that demonstrate several levels on the trajectory. In a whole-class discussion, she invites students to consider what the different ways of displaying the data “show and hide” about the data and how they do so. During this conversation, the students begin to appreciate the basis for conventions about display.\(^5\) For example, in their initial attempt at representing the data they have collected, many of the students draw icons to resemble the organisms that are not of uniform size (see Figure 3-14 in Chapter 3). The mismatches between their icons and the actual relative lengths of the organisms become clear in the discussion. The teacher also invites students to consider how using mathematical ideas (related to ordering, counting, and intervals) helped them develop different shapes to represent the same data.

The teacher's focus on shape is an assessment of what is defined as the crosscutting concept of patterns in the framework and the NGSS. These activities also cultivate the students' capacity to think at a population level about the biological significance of the shapes, as they realize what the different representations of the measurements they have taken can tell them. Some of the student displays make a bell-like shape more evident, which inspires further questions and considerations in the whole-class discussion (see Figure 3-15 in Chapter 3): students notice that the tails of the distribution are comparatively sparse, especially for the longer larvae, and wonder why. As noted in Chapter 3, they speculate about the possible reasons for the differences, which leads to a discussion and conclusions about competition for resources, which in turn leads them to consider not only individual silkworms, but the entire population of silkworms. Hence, this assessment provides students with opportunities for learning about representations, while also providing the teacher with information about their understanding of a crosscutting concept (pattern) and disciplinary core concepts (population-level descriptions of variability and the mechanisms that produce it).

**Example 4: Behavior of Air**

The committee chose this example to show the use of classroom discourse to assess student understanding. The exercise is designed to focus students' attention on a particular concept: the teacher uses class discussion of the students' models of air particles to identify misunderstandings and then support students in collaboratively resolving them. This task assesses both students' understanding of the concept and their proficiency with the practices of modeling and developing oral

---

\(^5\)This is a form of meta-representational competence; see diSessa (2004).
arguments about what they have observed. This assessment is used formatively and is closely tied to classroom instruction.

Classroom discussions can be a critical component of formative assessment. They provide a way for students to engage in scientific practices and for teachers to instantly monitor what the students do and do not understand. This example, from a unit for middle school students on the particle nature of matter, illustrates how a teacher can use discussions to assess students’ progress and determine instructional next steps.6

In this example, 6th-grade students are asked to develop a model to explain the behavior of air. The activity leads them to an investigation of phase change and the nature of air. The example is from a single class period in a unit devoted to developing a conceptual model of a gas as an assemblage of moving particles with space between them; it consists of a structured task and a discussion guided by the teacher (Krajcik et al., 2013; Krajcik and Merritt, 2012). The teacher is aware of an area of potential difficulty for students, namely, a lack of understanding that there is empty space between the molecules of air. She uses group-developed models and student discussion of them as a probe to evaluate whether this understanding has been reached or needs further development.

When students come to this activity in the course of the unit, they have already reached consensus on several important ideas they can use in constructing their models. They have defined matter as anything that takes up space and has mass. They have concluded that gases—including air—are matter. They have determined through investigation that more air can be added to a container even when it already seems full and that air can be subtracted from a container without changing its size. They are thus left with questions about how more matter can be forced into a space that already seems to be full and what happens to matter when it spreads out to occupy more space. The students have learned from earlier teacher-led class discussions that simply stating that the gas changes “density” is not sufficient, since it only names the phenomenon—it does not indicate what actually makes it possible for differing amounts of gas to expand or contract to occupy the same space.

In this activity, students are given a syringe and asked to gradually pull the plunger in and out of it to explore the air pressure. They notice the pressure

---

6This example was drawn from research conducted on classroom enactments of the IQWST curriculum materials (Krajcik et al., 2008; Shwartz et al., 2008). In field trials of IQWST, a diverse group of students responded to the task described in this example: 43% were white/Asian and 57% were non-Asian/minority; and 4% were English learners (Banilower et al., 2010).
against their fingers when pushing in and the resistance as they pull the plunger out. They find that little or no air escapes when they manipulate the plunger. They are asked to work in small groups to develop a model to explain what happens to the air so that the same amount of it can occupy the syringe regardless of the volume of space available. The groups are asked to provide models of the air with the syringe in three positions: see Figure 4-1. This modeling activity itself is not used as a formal assessment task; rather, it is the class discussion, in which students compare their models, that allows the teacher to diagnose the students’ understanding. That is, the assessment, which is intended to be formative, is conducted through the teacher’s probing of students’ understandings through classroom discussion.

Figure 4-2 shows the first models produced by five groups of students to depict the air in the syringe in its first position. The teacher asks the class to discuss the different models and to try to reach consensus on how to model the behavior of air to explain their observations. The class has agreed that there should be “air particles” (shown in each of their models as dark dots) and that the particles are moving (shown in some models by the arrows attached to the dots).

Most of their models are consistent in representing air as a mixture of different kinds of matter, including air, odor, dust, and “other particles.” What is not consistent in their models is what is represented as between the particles: groups 1 and 2 show “wind” as the force moving the air particles; groups 3, 4, and 5 appear to show empty space between the particles. Exactly what, if anything, is in between the air particles emerges as a point of contention as the students discuss their models. After the class agrees that the consensus model should include air particles shown with arrows to demonstrate that the particles “are coming out in different directions,” the teacher draws several particles with arrows and asks what to put next into the model. The actual classroom discussion is shown in Box 4-2.

The discussion shows how students engage in several scientific and engineering practices as they construct and defend their understanding about a disciplinary core idea. In this case, the key disciplinary idea is that there must be empty space between moving particles, which allows them to move, either to become more densely packed or to spread apart. The teacher can assess the way the students have drawn their models, which reveals that their understanding is not complete. They have agreed that all matter, including gas, is made of particles that are moving, but many of the students do not understand what is in between these moving particles. Several students indicate that they think there is air between the air par-
FIGURE 4-1 Models for air in a syringe in three situations for Example 4, “Behavior of Air.”
SOURCE: Krajcik et al. (2013). Reprinted with permission from Sangari Active Science.

ticles, since “air is everywhere,” and some assert that the particles are all touching. Other students disagree that there can be air between the particles or that air particles are touching, although they do not yet articulate an argument for empty space between the particles, an idea that students begin to understand more clearly in subsequent lessons. Drawing on her observations, the teacher asks questions
and gives comments that prompt the students to realize that they do not yet agree 
on the question of what is between the particles. The teacher then uses this obser- 
vation to make instructional decisions. She follows up on one student’s critique of 
the proposed addition to the consensus model to focus the students on their dis- 
agreement and then sends the class back into their groups to resolve the question.

In this example, the students’ argument about the models plays two roles: 
it is an opportunity for students to defend or challenge their existing ideas, and it 
is an opportunity for the teacher to observe what the students are thinking and to 
decide that she needs to pursue the issue of what is between the particles of air. It 
is important to note that the teacher does not simply bring up this question, but 
instead uses the disagreement that emerges from the discussion as the basis for the 
question. (Later interviews with the teacher reveal that she had in fact anticipated 
that the empty space between particles would come up and was prepared to take 
advantage of that opportunity.) The discussion thus provides insights into stu-
students' thinking beyond their written (and drawn) responses to a task. The models themselves provide a context in which the students can clarify their thinking and refine their models in response to the critiques, to make more explicit claims to explain what they have observed. Thus, this activity focuses their attention on key explanatory issues (Reiser, 2004).

This example also illustrates the importance of engaging students in practices to help them develop understanding of disciplinary core ideas while also giving teachers information to guide instruction. In this case, the teacher's active probing of students' ideas demonstrates the way that formative assessment strategies can be effectively used as a part of instruction. The discussion of the models not only reveals the students' understanding about the phenomenon, but also allows the teacher to evaluate progress, uncover problematic issues, and help students construct and refine their models.

**Example 5: Movement of Water**

The committee chose this example to show how a teacher can monitor developing understanding in the course of a lesson. “Clicker technology” is used to obtain individual student responses that inform teachers of what the students have learned from an activity and which are then the basis for structuring small-group discussions that address misunderstandings. This task assesses both understanding of a concept as it develops in the course of a lesson and students' discussion skills. The assessments are used formatively and are closely tied to classroom instruction.

In the previous example (Example 4), the teacher orchestrates a discussion in which students present alternative points of view and then come to consensus about a disciplinary core idea through the practice of argumentation. However, many teachers may find it challenging to track students' thinking while also promoting the development of understanding for the whole class. The example on the movement of air was developed as part of a program for helping teachers learn to lead students in “assessment conversations” (Duschl and Gitomer, 1997). In the

---

7 Clicker technology, also known as classroom response systems, allows students to use handheld clickers to respond to questions from a teacher. The responses are gathered by a central receiver and immediately tallied for the teacher—or the whole class—to see.

8 This example is taken from the Contingent Pedagogies Project, which provides formative assessment tools for middle schools and supports teachers in integrating assessment activities into discussions for both small groups and entire classes. Of the students who responded to the task, 46 percent were Latino. For more information, see http://contingentpedagogies.org [October 2013].
**BOX 4-2**

**STUDENT-TEACHER DIALOGUE**

*Haley’s objection: air is everywhere*

Ms. B: OK. Now what?
S: Just draw like little . . .
Haley: I think you should color the whole circle in, because dust . . . I mean air is everywhere, so . . .
Miles: The whole circle?
Ms. B: So, I color the whole thing in.
Haley: Yeah.
Ms. B: So, if I do one like that, because I haven’t seen one up here yet. If I color this whole thing in . . .

[Mrs. B colors in the whole region completely to show the air as Haley suggests.]

Michael: Then how would you show that . . . ?
Ms. B: Then ask . . . ask Haley some questions.
Students: How could that be? How would you show that?
Ms. B: Haley, people have some questions for you.

*Some students object to Haley’s proposal:*

Frank: How would you show air?
Haley: Air is everywhere, so the air would be everything.
Ss: Yeah.
Alyssa: But then, how would you show the other molecules? I mean, you said air is everything, but then how would you show the other . . .?
Ss: Yeah, because . . . [Multiple students talking]
Haley: What? I didn’t hear your question.
Alyssa: Um, I said if . . . You said air is everywhere, right?
Haley: Yeah. . . . so, that’s why you wanted to color it in. But there’s also other particles other than air, like dust and etc. and odors and things like that, so, how would you show that?
Miles: How are we going to put in the particles?
Ms. B: Haley, can you answer her?
Haley: No.
Ms. B: Why?
Haley: I don’t know.
Other student: Because there is no way.
Ms. B: Why can’t you answer?
Haley: What? I don’t know.
Ms. B: Is what she’s saying making sense?
Haley: Yeah.
Ms. B: What is it that you’re thinking about?
Haley: Um . . . that maybe you should take . . . like, erase some of it to show the odors and stuff.
Addison: No, wait, wait!
Ms. B: All right, call on somebody else.

**Addison proposes a compromise, and Ms. B pushes for clarification**

Addison: Um, I have an idea. Like since air is everywhere, you might be able to like use a different colored marker and put like, um, the other molecules in there, so you’re able to show that those are in there and then air is also everywhere.
Jerome: Yeah. I was gonna say that, or you could like erase it. If you make it all dark, you can just erase it and all of them will be.
Frank: Just erase some parts of the, uh . . . yeah, yeah, just to show there’s something in between it.
Ms. B: And what’s in between it?
Ss: The dust and the particles. Air particles. Other odors.
Miles: That’s like the same thing over there.
Alyssa: No, the colors are switched.
Ms. B: Same thing over where?
Alyssa: The big one, the consensus.
Ms. B: On this one?
Alyssa: Yeah.
Ms. B: Well, what she’s saying is that I should have black dots every which way, like that. [*Ms. B draws the air particles touching one another in another representation, not in the consensus model, since it is Haley’s idea.*]
Students: No what? Yeah.
Ms. B: Right?
Ms. B: OK. Talk to your partners. Is this what we want? [*pointing to the air particles touching one another in the diagram*]

**Students discuss in groups whether air particles are touching or not, and what is between the particles if anything.**

SOURCE: Reiser et al. (2013). Copyright by the author; used with permission.
task, middle school students engage in argumentation about disciplinary core ideas in earth science. As with the previous example, the formative assessment activity is more than just the initial question posed to students; it also includes the discussion that follows from student responses to it and teachers’ decisions about what to do next, after she brings the discussion to a close.

In this activity, which also takes place in a single class session, the teacher structures a conversation about how the movement of water affects the deposition of surface and subsurface materials. The activity involves disciplinary core ideas (similar to Earth’s systems in the NGSS) and engages students in practices, including modeling and constructing examples. It also requires students to reason about models of geosphere-hydrosphere interactions, which is an example of the cross-cutting concept pertaining to systems and system models.9

Teachers use classroom clicker technology to pose multiple-choice questions that are carefully designed to elicit students’ ideas related to the movement of water. These questions have been tested in classrooms, and the response choices reflect common student ideas, including those that are especially problematic. In the course of both small-group and whole-class discussions, students construct and challenge possible explanations of the process of deposition. If students have difficulty in developing explanations, teachers can guide students to activities designed to improve their understanding, such as interpreting models of the deposition of surface and subsurface materials.

When students begin this activity, they will just have completed a set of investigations of weathering, erosion, and deposition that are part of a curriculum on investigating Earth systems.10 Students will have had the opportunity to build physical models of these phenomena and frame hypotheses about how water will move sediment using stream tables.11 The teacher begins the formative assessment activity by projecting on a screen a question about the process of deposition designed to check students’ understanding of the activities they have completed; see Figure 4-3 for a sample question. Students select their answers using clickers.

9The specific NGSS core idea addressed is similar to MS-ESS2.C: “How do the properties and movement of water shape Earth’s surface and affect its systems?” The closest NGSS performance expectation is MS-ESS2-c: “Construct an explanation based on evidence for how geoscience processes have changed Earth’s surface at varying time and spatial scales.”

10This curriculum, for middle school students, was developed by the American Geosciences Institute. For more information, see http://www.agiweb.org/education/ies [July 2013].

11Stream tables are models of stream flows set up in large boxes filled with sedimentary material and tilted so that water can flow through.
Pairs or small groups of students then discuss their reasoning and offer explanations for their choices to the whole class. Teachers help students begin the small-group discussions by asking why someone might select A, B, or C, implying that any of them could be a reasonable response. Teachers press students for their reasoning and invite them to compare their own reasoning to that of others, using specific discussion strategies (see Michaels and O'Connor, 2011; National Research Council, 2007). After discussing their reasoning, students again vote, using their clickers. In this example, the student responses recorded using the clicker technology are scorable. A separate set of assessments (not discussed here) produces scores to evaluate the efficacy of the project as a whole.

The program materials include a set of “contingent activities” for teachers to use if students have difficulty meeting a performance expectation related to an investigation. Teachers use students’ responses to decide which contingent activities are needed, and thus they use the activity as an informal formative assessment. In these activities, students might be asked to interpret models, construct explanations, and make predictions using those models as a way to deepen their understanding of Earth systems. In this example about the movement of air, students who are having difficulty understanding can view an animation of deposition and then make a prediction about a pattern they might expect to find at the mouth of a river where sediment is being deposited.
The aim of this kind of assessment activity is to guide teachers in using assessment techniques to improve student learning outcomes. The techniques used in this example demonstrate a means of rapidly assessing how well students have mastered a complex combination of practices and concepts in the midst of a lesson, which allows teachers to immediately address areas students do not understand well. The contingent activities that provide alternative ways for students to master the core ideas (by engaging in particular practices) are an integral component of the formative assessment process.

**Example 6: Biodiversity in the Schoolyard**

The committee chose this example to show the use of multiple interrelated tasks to assess a disciplinary core idea, biodiversity, with multiple science practices. As part of an extended unit, students complete four assessment tasks. The first three serve formative purposes and are designed to function close to instruction, informing the teacher about how well students have learned key concepts and mastered practices. The last assessment task serves a summative purpose, as an end-of-unit test, and is an example of a proximal assessment. The tasks address concepts related to biodiversity and science practices in an integrated fashion.

This set of four assessment tasks was designed to provide evidence of 5th-grade students’ developing proficiency with a body of knowledge that blends a disciplinary core idea (biodiversity; LS4 in the NGSS; see Box 2-1 in Chapter 2) and a crosscutting concept (patterns) with three different practices: planning and carrying out investigations, analyzing and interpreting data, and constructing explanations (see Songer et al., 2009; Gotwals and Songer, 2013). These tasks, developed by researchers as part of an examination of the development of complex reasoning, are intended for use in an extended unit of study.

---

12A quasi-experimental study compared the learning gains for students in classes that used the approach of the Contingent Pedagogies Project with gains for students in other classes in the same school district that used the same curriculum but not that approach. The students whose teachers used the Contingent Pedagogies Project demonstrated greater proficiency in earth science objectives than did students in classrooms in which teachers only had access to the regular curriculum materials (Penuel et al., 2012).

13The tasks were given to a sample of 6th-grade students in the Detroit Public School System, the majority of whom were racial/ethnic minority students (for details, see Songer et al., 2009).
Formative Assessment Tasks

Task 1: Collect data on the number of animals (abundance) and the number of different species (richness) in schoolyard zones.

Instructions: Once you have formed your team, your teacher will assign your team to a zone in the schoolyard. Your job is to go outside and spend approximately 40 minutes observing and recording all of the animals and signs of animals that you see in your schoolyard zone during that time. Use the BioKIDS application on your iPod to collect and record all your data and observations.

In responding to this task, students use an Apple iPod to record their information. The data from each iPod is uploaded and combined into a spreadsheet that contains all of the students’ data; see Figure 4-4. Teachers use data from individual groups or from the whole class as assessment information to provide formative information about students’ abilities to collect and record data for use in the other tasks.

Task 2: Create bar graphs that illustrate patterns in abundance and richness data from each of the schoolyard zones.

Task 2 assesses students’ ability to construct and interpret graphs of the data they have collected (an important element of the NGSS practice “analyzing and interpreting data”). The exact instructions for Task 2 appear in Figure 4-5. Teachers use the graphs the students create for formative purposes, for making decisions about further instruction students may need. For example, if students are weak on the practices, the teacher may decide to help them with drawing accurate bars or the appropriate labeling of axes. Or if the students are weak on understanding of the core idea, the teacher might review the concepts of species abundance or species richness.

Task 3: Construct an explanation to support your answer to the question: Which zone of the schoolyard has the greatest biodiversity?

Before undertaking this task, students have completed an activity that helped them understand a definition of biodiversity: “An area is considered biodiverse if it has both a high animal abundance and high species richness.” The students were also given hints (reminders) that there are three key parts of an explanation: a claim, more than one piece of evidence, and reasoning. The students are also
<table>
<thead>
<tr>
<th>Animal Name</th>
<th>Zone A</th>
<th>Zone C</th>
<th>Zone E</th>
<th>Microhabitat</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earthworms</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>In dirt</td>
<td>4</td>
</tr>
<tr>
<td>Ants</td>
<td>2</td>
<td>229</td>
<td>75</td>
<td>On something hard</td>
<td>306</td>
</tr>
<tr>
<td>Other insects</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>On grass</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Other microhabitat</td>
<td></td>
</tr>
<tr>
<td>Unknown beetle</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>On plant</td>
<td>3</td>
</tr>
<tr>
<td>Unknown insect</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>On dirt</td>
<td>2</td>
</tr>
<tr>
<td>Other leggy invertebrates</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>In dirt</td>
<td>1</td>
</tr>
<tr>
<td>American robin</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>On tree</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>In the sky</td>
<td></td>
</tr>
<tr>
<td>Black tern</td>
<td>200</td>
<td>0</td>
<td>0</td>
<td>On plant</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>On something hard</td>
<td></td>
</tr>
<tr>
<td>House sparrow</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>On tree</td>
<td>1</td>
</tr>
<tr>
<td>Mourning dove</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>On tree</td>
<td>3</td>
</tr>
<tr>
<td>Unknown bird</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>On tree</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>In the sky</td>
<td></td>
</tr>
<tr>
<td>Other birds</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>In water</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>On grass</td>
<td></td>
</tr>
<tr>
<td>E. fox squirrel</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>On something hard</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>On grass</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Other microhabitat</td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>10</td>
<td>21</td>
<td>1</td>
<td>On something hard</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>On grass</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Other microhabitat</td>
<td></td>
</tr>
<tr>
<td>Other mammal</td>
<td>3</td>
<td>0</td>
<td>16</td>
<td>On something hard</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Other microhabitat</td>
<td></td>
</tr>
<tr>
<td>Red squirrel</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>On tree</td>
<td>2</td>
</tr>
</tbody>
</table>

Number of Animals (Abundance) 237 263 104 Total 604
Number of Kinds of Animals (Richness) 11 8 9 Total 28

FIGURE 4-4 Class summary of animal observations in the schoolyard, organized by region (schoolyard zones), for Example 6, “Biodiversity in the Schoolyard.”
Instructions:

1. Use your zone summary to make a bar chart of your abundance data.

   Please remember to label your axes.

   ![Our Abundance Data]

   Result: According to the bar chart above, zone C has the highest abundance.

2. Use your zone summary to make a bar chart of your richness data.

   Please remember to label your axes.

   ![Our Richness Data]

**FIGURE 4-5** Instructions for Task 2 for Example 6, “Biodiversity in the Schoolyard.”

NOTE: See text for discussion.
given the definitions of relevant terms. This task allows the teacher to see how well students have understood the concept and can support their ideas about it. Instructions for Task 3 and student answers are shown in Box 4-3.

**Summative Assessment Task**

Task 4: Construct an explanation to support an answer to the question: Which zone of the schoolyard has the greatest biodiversity?

For the end-of-unit assessment, the task presents students with excerpts from a class data collection summary, shown in Table 4-1, and asks them to construct an explanation, as they did in Task 3. The difference is that in Task 4, the hints are removed: at the end of the unit, they are expected to show that they understand what constitutes a full explanation without a reminder. The task and coding rubric used for Task 4 are shown in Box 4-4.

**The Set of Tasks**

This set of tasks illustrates two points. First, using tasks to assess several practices in the context of a core idea together with a crosscutting concept can provide a wider range of information about students’ progression than would tasks that focused on only one practice. Second, classroom assessment tasks in which core ideas, crosscutting concepts, and practices are integrated can be used for both formative and summative purposes. Table 4-2 shows the core idea, crosscutting concept, practices, assessment purposes, and performance expectation targets for assessment for each of the tasks. Each of these four tasks was designed to provide information about a single performance expectation related to the core idea, and each performance expectation focused on one of three practices. Figure 4-6 illustrates the way these elements fit together to identify the target for assessment of Tasks 3 and 4.

Second, the design of each task was determined by its purpose (formative or summative) and the point in the curriculum at which it was to be used. Assessment tasks may, by design, include more or less guidance for students, depending on the type of information they are intended to collect. Because learning is a process that occurs over time, a teacher might choose an assessment task with fewer guides (or scaffolds) for students as they progress through a curriculum to gather evidence of what students can demonstrate without assistance. Thus, the task developers offered a practice progression to illustrate the different levels of
BOX 4.3
INSTRUCTIONS AND SAMPLE STUDENT ANSWERS FOR TASK 3 IN EXAMPLE 6, “BIODIVERSITY IN THE SCHOOLYARD”

Instructions: Using what you have learned about biodiversity, the information from your class summary sheet, and your bar charts for abundance and richness, construct an explanation to answer the following scientific question:

Scientific Question: Which zone in the schoolyard has the highest biodiversity?
My Explanation [figure or text box?]

Make a CLAIM: Write a complete sentence that answers the scientific question.

Zone A has the greatest biodiversity.

Hint: Look at your abundance and richness data sheets carefully.

Give your REASONING: Write the scientific concept or definition that you thought about to make your claim.

Hint: Think about how biodiversity is related to abundance and richness.

Biodiversity is related to abundance and richness because it shows the two amounts in one word.

Give your EVIDENCE: Look at your data and find two pieces of evidence that help answer the scientific question.

1. Zone A has the most richness.
2. Zone A has a lot of abundance.

Hint: Think about which zone has the highest abundance and richness.

NOTES: Student responses are shown in italics. See text for discussion.
TABLE 4-1 Schoolyard Animal Data for Example 6 Summative Task, “Biodiversity in the Schoolyard”

<table>
<thead>
<tr>
<th>Animal Name</th>
<th>Zone A</th>
<th>Zone B</th>
<th>Zone C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillbugs</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Ants</td>
<td>4</td>
<td>6</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Robins</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Squirrels</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Pigeons</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Animal abundance</td>
<td>6</td>
<td>14</td>
<td>16</td>
<td>36</td>
</tr>
<tr>
<td>Animal richness</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

Guidance that tasks might include, depending on their purpose and the stage students will have reached in the curriculum when they undertake the tasks.

Box 4-5 shows a progression for the design of tasks that assess one example of three-dimensional learning: the practice of constructing explanations with one core idea and crosscutting concept. This progression design was based on studies that examined students’ development of three-dimensional learning over time, which showed that students need less support in tackling assessment tasks as they progress in knowledge development (see, e.g., Songer et al., 2009).

Tasks 3 and 4, which target the same performance expectation but have different assessment purposes, illustrate this point. Task 3 was implemented midway through the curricular unit to provide formative information for the teacher on the kinds of three-dimensional learning students could demonstrate with the assistance of guides. Task 3 was classified as a Level 5 task (in terms of the progression shown in Box 4-5) and included two types of guides for the students (core idea guides in text boxes and practice guides that offer the definition of claim, evidence, and reasoning). Task 4 was classified as a Level 7 task because it did not provide students with any guides to the construction of explanations.

**Example 7: Climate Change**

The committee chose this flexible online assessment task to demonstrate how assessment can be customized to suit different purposes. It was designed to probe student understanding and to facilitate a teacher’s review of responses. Computer software allows teachers to tailor online assessment tasks to their purpose and to the stage of learning that students have reached, by offering more or less supporting information.
BOX 4.4

TASK AND CODING RUBRIC FOR TASK 4 IN EXAMPLE 6, “BIODIVERSITY IN THE SCHOOLYARD”

Write a scientific argument to support your answer for the following question.

Scientific Question: Which zone has the highest biodiversity?

Coding

4 points: Contains all parts of explanation (correct claim, 2 pieces of evidence, reasoning)
3 points: Contains correct claim and 2 pieces of evidence but incorrect or no reasoning
2 points: Contains correct claim + 1 piece correct evidence OR 2 pieces correct evidence and 1 piece incorrect evidence
1 point: Contains correct claim, but no evidence or incorrect evidence and incorrect or no reasoning

Correct Responses

Claim

Correct: Zone B has the highest biodiversity.

Evidence

1. Zone B has the highest animal richness.
2. Zone B has high animal abundance.

Reasoning

Explicit written statement that ties evidence to claim with a reasoning statement: that is, Zone B has the highest biodiversity because it has the highest animal richness and high animal abundance. Biodiversity is a combination of both richness and abundance, not just one or the other.

and guidance. The tasks may be used for both formative and summative purposes: they are designed to function close to instruction.

This online assessment task is part of a climate change curriculum for high school students. It targets the performance expectation that students use geoscience data and the results from global climate models to make evidence-based
TABLE 4-2 Characteristics of Tasks in Example 6, “Biodiversity in the Schoolyard”

<table>
<thead>
<tr>
<th>Core Idea</th>
<th>Crosscutting Concepts</th>
<th>Practices</th>
<th>Purpose of Assessment</th>
<th>Target for Assessment: Performance Expectation</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS4.D Biodiversity and Humans</td>
<td>Patterns</td>
<td>Planning and carrying out investigations</td>
<td>Formative</td>
<td>Task 1. Collect data on the number of animals (abundance) and the number of different species (richness) in schoolyard zones.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Analyzing and interpreting data</td>
<td>Formative</td>
<td>Task 2. Create bar graphs that illustrate patterns in abundance and richness data from each of the schoolyard zones.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Constructing explanations</td>
<td>Formative</td>
<td>Task 3. Construct an explanation to support your answer to the question: Which zone of the schoolyard has the greatest biodiversity?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Constructing explanations</td>
<td>Summative</td>
<td>Task 4. Construct an explanation to support your answer to the question: Which zone of the schoolyard has the greatest biodiversity?</td>
</tr>
</tbody>
</table>

forecasts of the impacts of climate change on organisms and ecosystems. This example illustrates four potential benefits of online assessment tasks:

1. the capacity to present data from various external sources to students;
2. the capacity to make information about the quality and range of student responses continuously available to teachers so they can be used for formative purposes;
3. the possibility that tasks can be modified to provide more or less support, or scaffolding, depending on the point in the curriculum at which the task is being used; and

14 This performance expectation is similar to two in the NGSS ones: HS-LS2-2 and HS-ESS3-5, which cover the scientific practices of analyzing and interpreting data and obtaining, evaluating, and communicating evidence.
4. the possibility that the tasks can be modified to be more or less active depending on teachers’ or students’ preferences.

In the instruction that takes place prior to this task, students will have selected a focal species in a particular ecosystem and studied its needs and how it is distributed in the ecosystem. They will also have become familiar with a set of model-based climate projections, called Future 1, 2, and 3, that represent more and less severe climate change effects. Those projections are taken from the Intergovernmental Panel on Climate Change (IPCC) data predictions for the year 2100 (Intergovernmental Panel on Climate Change, 2007): see Figure 4-7. The materials provided online as part of the activity include

- global climate model information presented in a table showing three different IPCC climate change scenarios (shown in Figure 4-7);
- geosciences data in the form of a map of North America that illustrates the current and the predicted distribution of locations of optimal biotic and abiotic\(^\text{15}\) conditions for a species, as predicted by IPCC Future 3 scenario: see Figure 4-8; and

\(^{15}\text{The biotic component of an environment consists of the living species that populate it, while the abiotic components are the nonliving influences such as geography, soil, water, and climate that are specific to the particular region.}
BOX 4-5

PROGRESSION FOR MULTIDIMENSIONAL LEARNING TASK DESIGN

This progression covers constructing a claim with evidence and constructing explanations with and without guidance. The + and ++ symbols represent the number of guides provided in the task.

**Level 7:**  Student is provided with a question and is asked to construct a scientific explanation (no guides).

**Level 6+:** Student is provided with a question and is asked to construct a scientific explanation (with core ideas guides only).

**Level 5++:** Student is provided with a question and is asked to construct a scientific explanation (with core ideas guides and guides defining claim, evidence and reasoning).

**Level 4:**  Student is provided with a question and is asked to make a claim and back it with evidence (no guides).

**Level 3+:** Student is provided with a question and is asked to make a claim and back it with evidence (with core ideas guides only).

**Level 2++:** Student is provided with a question and is asked to make a claim and back it with evidence (with core ideas guides and guides defining claim and evidence).

**Level 1:**  Student is provided with evidence and asked to choose appropriate claim OR student is provided with a claim and is asked to choose the appropriate evidence.

SOURCE: Adapted from Gotwals and Songer (2013).

- an online guide for students in the development of predictions, which prompts them as to what is needed and records their responses in a database that teachers and students can use. (The teacher can choose whether or not to allow students access to the pop-up text that describes what is meant by a claim or by evidence.)

The task asks students to make and support a prediction in answer to the question, “In Future 3, would climate change impact your focal species?” Students are asked to provide the following:
<table>
<thead>
<tr>
<th>Future 1</th>
<th>Future 2</th>
<th>Future 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Slow</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1862</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1499</td>
</tr>
<tr>
<td></td>
<td></td>
<td>983</td>
</tr>
</tbody>
</table>

FIGURE 4-7 Three simplified Intergovernmental Panel on Climate Change (IPCC)-modeled future scenarios for the year 2100.
SOURCE: Adapted from Peters et al. (2012).

- a claim (the prediction) as to whether or not they believe the IPCC scenario information suggests that climate change will affect their chosen animal;
- reasoning that connects their prediction to the model-based evidence, such as noting that their species needs a particular prey to survive; and
- model-based evidence that is drawn from the information in the maps of model-based climate projections, such as whether or not the distribution of conditions needed by the animal and its food source in the future scenario will be significantly different from what it is at present.

Table 4-3 shows sample student responses that illustrate both correct responses and common errors. Students 1, 3, and 4 have made accurate predictions, and supplied reasoning and evidence; students 2, 5, and 6 demonstrate common errors, including insufficient evidence (student 2), inappropriate reasoning and evidence (student 5), and confusion between reasoning and evidence (student 6). Teachers can use this display to quickly see the range of responses in the class and use that information to make decisions about future instruction.

**Example 8: Ecosystems**

The committee chose this example, drawn from the SimScientists project, to demonstrate the use of simulation-based modules designed to be embedded in a curriculum unit to provide both formative and summative assessment information. Middle school students use computer simulations to demonstrate their understanding of core ideas about ecosystem dynamics and the progress of their
FIGURE 4-8 Current and predicted Future 3 distribution for the red squirrel for Example 7, “Climate Change.” SOURCE: Songer et al. (2013). Copyright by the author; used with permission.

thinking as they move from exploring ecosystem components to interactions of those components to the way systems behave. Thus, the simulations also address the crosscutting concept of systems. The assessment components function close to classroom instruction.

In this set of classroom modules, students use simulated, dynamic representations of particular ecosystems, such as a mountain lake or grasslands, to investigate features common to all ecosystems. The students investigate the roles of and relationships among species within habitats and the effects of these interactions on population levels (Quellmalz et al., 2009). Simulations of these environments can be used both to improve students’ understanding of complex ecosystems and to
### TABLE 4-3 Sample Student Responses in Example 7, “Climate Change”

**Scientific Question:** In Future 3, would climate change impact your focal species?

<table>
<thead>
<tr>
<th>Student</th>
<th>Claim</th>
<th>Reasoning</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student 1</td>
<td>Claim</td>
<td>Climate change will effect my focal species.</td>
<td>The abiotic conditions will change, and the temperature will change; therefore, the habitat of my species will change.</td>
</tr>
<tr>
<td></td>
<td>Reasoning</td>
<td></td>
<td>The map shows it will move into the Western part, therefore the climate changed.</td>
</tr>
<tr>
<td>Student 2</td>
<td>Claim</td>
<td>Yes it will effect it, it will shorten the range.</td>
<td>When the climate changes the focal species will have to move north because it won’t be able to stand the warm weather.</td>
</tr>
<tr>
<td></td>
<td>Reasoning</td>
<td></td>
<td>The map.</td>
</tr>
<tr>
<td>Student 3</td>
<td>Claim</td>
<td>Yes, climate change would effect the red-backed salamander.</td>
<td>Abiotic and biotic factors can cause the red-backed salamander to relocate, such as temperature, precipitation, and invasive species.</td>
</tr>
<tr>
<td></td>
<td>Reasoning</td>
<td></td>
<td>If the temperatures increase, the red-backed salamander would have to live farther north where temperatures are suitable for its survival.</td>
</tr>
<tr>
<td>Student 4</td>
<td>Claim</td>
<td>I think that climate change in Future 3 will not impact my focal species.</td>
<td>Some abiotic features that could affect the focal species could be the climate, but it won’t move the focal species from the location.</td>
</tr>
<tr>
<td></td>
<td>Reasoning</td>
<td></td>
<td>According to the distribution map for Future 3 the American Kestrel does not move from the location.</td>
</tr>
<tr>
<td>Student 5</td>
<td>Claim</td>
<td>No because my focal species is a bird and it can migrate to a warmer area but if the climate gets warm earlier then it will migrate earlier and it could affect it’s normal time to migrate.</td>
<td>The food they eat might not be out of hibernation or done growing in the area it migrates to.</td>
</tr>
<tr>
<td></td>
<td>Reasoning</td>
<td></td>
<td>It eats mice and mice hibernate and and so do voles and if the climate changes to a cold climate too early then their food will be hidden and they will have to migrate early.</td>
</tr>
</tbody>
</table>

**NOTE:** Both correct and incorrect responses are shown.

**SOURCE:** Songer et al. (2013). Copyright by the author; used with permission.
<table>
<thead>
<tr>
<th>Model Level</th>
<th>Description</th>
<th>Disciplinary Core Ideas</th>
<th>Science Practices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component</td>
<td>What are the components of the system and their rules of behavior?</td>
<td>Every ecosystem has a similar pattern of organization with respect to the roles (producers, consumers, and decomposers) that organisms play in the movement of energy and matter through the system. (NGSS: LS2.A — Interdependent relationships in Ecosystems)</td>
<td>Analyze and interpret data to provide evidence for phenomena.</td>
</tr>
<tr>
<td>Interaction</td>
<td>How do the individual components interact?</td>
<td>Matter and energy flow through the ecosystem as individual organisms participate in feeding relationships within an ecosystem. (NGSS: LS2.B — Cycles of Matter and Energy Transfer in Ecosystems)</td>
<td>Develop a model to describe phenomena. Analyze and interpret data.</td>
</tr>
<tr>
<td>Emergent</td>
<td>What is the overall behavior or property of the system that results from many interactions following specific rules?</td>
<td>Interactions among organisms and among organisms and the ecosystem’s nonliving features cause the populations of the different organisms to change over time. (NGSS: LS2.C — Ecosystem Dynamics, Functioning, and Resilience)</td>
<td>Use a model to plan and carry out investigations. Analyze and interpret data to provide evidence.</td>
</tr>
</tbody>
</table>

FIGURE 4-9 Ecosystems target model for Example 8, “Ecosystems.”
SOURCE: SimScientists Calipers II project (2013). Reprinted with permission.

assess what they have learned. The simulated environments provide multiple representations of system models at different scales. They require students to apply core ideas about ecosystems and to carry out such practices as building and using models, planning and conducting investigations (by manipulating the system elements), and interpreting patterns.

Figure 4-9 shows a model of the characteristics of and changes in ecosystems as it would appear on the screen. The model would be very difficult for students to observe or investigate using printed curriculum materials.16 For example, Figure 4-10 shows part of a simulated mountain lake environment. Students observe animations of the organisms’ interactions and are then asked to draw a food web directly on the screen to represent a model of the flow of matter and energy in the ecosystem. If a student draws an arrow that links a food consumer to the wrong source of matter and energy, a feedback box coaches the student to observe again by reviewing the animation, thus providing formative feedback.

---

16These same features also make it difficult to display the full impact of the simulation in this report.
FIGURE 4-10 Screenshot of a curriculum-embedded assessment of student constructing a food web to model the flow of matter and energy in the ecosystem (with feedback and coaching); part of Example 8, “Ecosystems.”

In the subsequent curriculum-embedded assessment, students investigate what happens to population levels when relative starting numbers of particular organisms are varied; see Figure 4-11. The interactive simulation allows students to conduct multiple trials to build, evaluate, and critique models of balanced ecosystems, interpret data, and draw conclusions. If the purpose of the assessment is formative, students can be given feedback and a graduated sequence of coaching by the program. Figure 4-11 shows a feedback box for this set of activities, which not only notifies the student that an error has occurred but also prompts the student to analyze the population graphs and design a third trial that maintains the survival of the organisms. As part of the assessment, students also complete tasks that ask them to construct descriptions, explanations, and conclusions. They are guided in assessing their own work by judging whether their response meets specified criteria, and then how well their response matches a sample one, as illustrated in Figure 4-12.
Can you do better than Dr. A? Design three trials so that both the shrimp and alewife populations survive for 20 years.

- Use the sliders to change the starting numbers of shrimp and alewife.
- Click RUN to see what happens.
- When you have saved 3 trials in which shrimp and alewife survive for 20 years, click NEXT.

**FIGURE 4-11** Screenshot of a curriculum-embedded assessment of student using simulations to build balanced ecosystem population models (with feedback and coaching); part of Example 8, "Ecosystems."

**SOURCE:** SimScientists Calipers II project (2013). Reprinted with permission.

The SimScientists assessments are designed to provide feedback that addresses common student misconceptions about the ecosystem components, interactions that take place within them, or the way they behave, as well as errors in the use of science practices. The simulation generates reports to students about their progress toward goals for conceptual understanding and use of practices, and it also provides a variety of reporting options for teachers. Teachers can view progress reports for individual students as well as class-level reports (Quellmalz et al., 2012).

The SimScientists assessment system was also designed to collect summative assessment information after students complete a regular curriculum unit on ecosystems (which might have included the formative assessment modules described above). Figures 4-13 and 4-14 show tasks that are part of a benchmark assessment scenario in which students are asked to investigate ways to restore an Australian grasslands ecosystem—one that is novel to them—that has been affected by a significant fire. No feedback or coaching is provided. Students investigate the roles of
and relationships among the animals, birds, insects, and grass by observing animations of their interactions. Students draw a food web representing a model of the flow of energy and matter throughout the ecosystem, based on the interactions they have observed. Students then use the simulation models to plan, conduct, interpret, explain, and critique investigations of what happens to population levels when numbers of particular organisms are varied. In a culminating task, students present their findings about the grasslands ecosystem.

These task examples from the SimScientists project illustrate ways that assessment tasks can take advantage of technology to represent generalizable, progressively more complex models of science systems, present challenging scientific reasoning tasks, provide individualized feedback, customize scaffolding, and promote self-assessment and metacognitive skills. Reports generated for teachers and students indicate the level of additional help students may need and classify students into groups for which tailored, follow-on, reflection activities are recommended (to be conducted during a subsequent class period).
FIGURE 4-13 Screenshot of a benchmark summative assessment of a student constructing a food web to model the flow of matter and energy in the ecosystem (without feedback and coaching); part of Example 8, “Ecosystems.”
SOURCE: SimScientists Calipers II project (2013). Reprinted with permission.

FIGURE 4-14 Screenshot of a benchmark summative assessment of a student using simulations to build balanced ecosystem population models (without feedback and coaching); part of Example 8, “Ecosystems.”
SOURCE: SimScientists Calipers II project (2013). Reprinted with permission.
These formative assessments also have an instructional purpose. They are designed to promote model-based reasoning about the common organization and behaviors of all ecosystems (see Figure 4-9) and to teach students how to transfer knowledge they gain about how one ecosystem functions to examples of new ecosystems (Buckley and Quellmalz, 2013).17

LESSONS FROM THE EXAMPLES

The six examples discussed above, as well as the one in Chapter 2, demonstrate characteristics we believe are needed to assess the learning called for in the NGSS and a range of approaches to using assessments constructively in the classroom to support such learning. A key goal of classroom assessment is to elicit and make visible students’ ways of thinking and acting. The examples demonstrate that it is possible to design tasks and contexts in which teachers elicit student thinking about a disciplinary core idea or crosscutting concept by engaging them in a scientific practice. The examples involve activities designed to stimulate classroom conversations or to produce a range of artifacts (products) that provide information to teachers about students’ current ways of thinking and acting, or both. This information can be used to adjust instruction or to evaluate learning that occurred during a specified time. Some of the examples involve formal scoring, while others are used by teachers to adjust their instructional activities without necessarily assigning student scores.

Types of Assessment Activities

In “What Is Going on Inside Me?” (Example 1 in Chapter 2), students produce a written evidence-based argument for an explanation of how animals get energy from food and defend that explanation orally in front of the class. In “Measuring Silkworms” (Example 3, above, and also discussed in Chapter 3), students produce representations of data and discuss what they do and do not reveal about the data. In “Behavior of Air” (Example 4, above), models developed by groups of students are the stimulus for class discussion and argumentation that the teacher uses to diagnose and highlight discrepancies in students’ ideas. In “Movement of Water” (Example 5, above), multiple-choice questions that students answer using

---

17The system was designed using the evidence-centered design approach discussed in Chapter 3. Research on the assessments supports the idea that this approach could be a part of a coherent, balanced state science assessment system: see discussion in Chapter 6.
clickers are the stimulus for class discussion (assessment conversation). In each of these examples, students’ writing and classroom discourse provide evidence that can be used in decisions about whether additional activities for learning might be needed, and, if so, what kinds of activities might be most productive. In many of these examples, listening to and engaging with other students as they discuss and defend their responses is a part of the learning process, as students work toward a classroom consensus explanation or a model based on the evidence they have collected. The classroom discussion itself in these cases is the basis for the formative assessment process.

We note that when assessments are designed to be used formatively, the goal is sometimes not to assign scores to individual students but rather to decide what further instruction is needed for groups of students or the class as a whole. Thus, instead of scoring rubrics, criteria or rubrics that can help guide instructional decisions may be used. (When the goal includes assessment of both individuals and groups, both types of scoring rubrics would be needed.) Teachers need support to learn to be intentional and deliberative about such decisions. In the examples shown, designers of curriculum and instruction have developed probes that address likely learning challenges, and teachers are supported in recognizing these challenges and in the use of the probes to seek evidence of what their students have learned and not learned, along some continuum.

“Ecosystems” (Example 8, above) is a computer-based system in which students use simulations both to learn and to demonstrate what they have learned about food webs. It includes tasks that are explicitly designed for assessment. Other tasks may not be sharply distinguished from ongoing classroom activities. The data collection tasks in “Biodiversity in the Schoolyard” (Example 6, above) are part of students’ ongoing investigations, not separate from them, but they can provide evidence that can be used for formative purposes.

Similarly, in “Measuring Silkworms” (Example 3) students create displays as part of the learning process in order to answer questions about biological growth. Constructing these displays engages students in the practice of analyzing data, and their displays are also a source of evidence for teachers about students’ proficiencies in reasoning about data aggregations; thus they can be used formatively. These forms of reasoning also become a topic of instructional conversations, so that students are encouraged to consider additional aspects of data representation, including tradeoffs about what different kinds of displays do and do not show about the same data. As students improve their capacity to visualize data, the data discussion then leads them to notice characteristics of organisms or populations.
that are otherwise not apparent. This interplay between learning a practice (data representation as an aspect of data analysis) and learning about a core idea (variation in a population), as well as a crosscutting concept (recognizing and interpreting patterns), provides an example of the power of three-dimensional learning, as well as an example of an assessment strategy.

**Interpreting Results**

A structured framework for interpreting evidence of student thinking is needed to make use of the task artifacts (products), which might include data displays, written explanations, or oral arguments. As we discuss in Chapter 3, interpretation of results is a core element of assessment, and it should be a part of the assessment design. An interpretive framework can help teachers and students themselves recognize how far they have progressed and identify intermediate stages of understanding and problematic ideas. "Measuring Silkworms" shows one such framework, a learning progression for data display developed jointly by researchers and teachers. "Behavior of Air" is similarly grounded in a learning progressions approach. "Movement of Water" presents an alternative example, using what is called a facets-based approach\(^{18}\) to track the stages in a learning progression (discussed in Chapter 2)—that is, to identify ideas that are commonly held by students relative to a disciplinary core idea. Although these preconceptions are often labeled as misconceptions or problematic ideas, they are the base on which student learning must be built. Diagnosing students' preconceptions can help teachers identify the types of instruction needed to move students toward a more scientific conception of the topic.

What these examples have in common is that they allow teachers to group students into categories, which helps with the difficult task of making sense of many kinds of student thinking; they also provide tools for helping teachers decide what to do next. In "Movement of Water," for example, students' use of clickers

---

\(^{18}\)In this approach, a facet is a piece of knowledge constructed by a learner in order to solve a problem or explain an event (diSessa and Minstrell, 1998). Facets that are related to one another can be organized into clusters, and the basis for grouping can either be an explanation or an interpretation of a physical situation or a disciplinary core idea (Minstrell and Kraus, 2005). Clusters comprise goal facets (which are often standards or disciplinary core ideas) and problematic facets (which are related to the disciplinary idea but which represent ways of reasoning about the idea that diverge from the goal facet). The facets perspective assumes that, in addition to problematic thinking, students also possess insights and understandings about the disciplinary core idea that can be deepened and revised through additional learning opportunities (Minstrell and van Zee, 2003).
to answer questions gives teachers initial feedback on the distribution of student ideas in the classroom. Depending on the prevalence of particular problematic ideas or forms of reasoning and their persistence in subsequent class discussion, teachers can choose to use a “contingent activity” that provides a different way of presenting a disciplinary core idea.

The interpretive framework for evaluating evidence has to be expressed with enough specificity to make it useful for helping teachers decide on next steps. The construct map for data display in “Measuring Silkworms” meets this requirement: a representation that articulated only the distinction between the lowest and highest levels of the construct map would be less useful. Learning progressions that articulate points of transition that take place across multiple years—rather than transitions that may occur in the course of a lesson or a unit—would be less useful for classroom decision making (although a single classroom may often include students who span such a range) (Alonzo and Gearhart, 2006).

**Using Multiple Practices**

The examples above involve tasks that cross different domains of science and cover multiple practices. “What Is Going on Inside Me?,” for example, requires students to demonstrate their understanding of how chemical processes support biological processes. It asks students not only to apply the crosscutting concept of energy and matter conservation, but also to support their arguments with explicit evidence about the chemical mechanism involved. In “Measuring Silkworms” and “Biodiversity in the Schoolyard,” students’ responses to the different tasks can provide evidence of their understanding of the crosscutting concept of patterns. It is important to note, however, that “patterns” in each case has a different and particular disciplinary interpretation. In “Measuring Silkworms,” students must recognize pattern in a display of data, in the form of the “shapes” the data can take, and begin to link ideas about growth and variation to these shapes. In contrast, in “Biodiversity in the Schoolyard,” students need to recognize patterns in the distribution and numbers of organisms in order to use the data in constructing arguments.

Three of the examples—“Measuring Silkworms,” “Biodiversity in the Schoolyard,” and “Climate Change”—provide some classroom-level snapshots of emerging proficiency with aspects of the practices of analyzing and interpreting data and using mathematics and computational thinking. We note, though, that each of these practices has multiple aspects, so multiple tasks would be needed to provide a complete picture of students’ capacity with each of them. Although
assessment tasks can identify particular skills related to specific practices, evaluating students’ disposition to engage in these practices without prompting likely requires some form of direct observation or assessment of the products of more open-ended student projects.\(^{19}\)

In instruction, students engage in practices in interconnected ways that support their ongoing investigations of phenomena. Thus, students are likely to find that to address their questions, they will need to decide which sorts of data (including observational data) are needed; that is, they will need to design an investigation, collect those data, interpret the results; and construct explanations that relate their evidence to both claims and reasoning. It makes little sense for students to construct data displays in the absence of a question. And it is not possible to assess the adequacy of their displays without knowing what question they are pursuing. In the past, teachers might have tried to isolate the skill of graphing data as something to teach separately from disciplinary content, but the new science framework and the NGSS call for teachers to structure tasks and interpret evidence in a broad context of learning that integrates or connects multiple content ideas and treats scientific practices as interrelated. Similarly, assessment tasks designed to examine students’ facility with a particular practice may require students to draw on other practices as they complete the task.

We stress in Chapter 2 that a key principle of the framework is that science education should connect to students’ interests and experiences. Students are likely to bring diverse interests and experiences to the classroom from their families and cultural communities. A potential focus of classroom assessment at the outset of instruction is to elicit students’ interests and experiences that may be relevant to the goals for instruction. However, identifying interests has not often been a focus of classroom assessment research in science, although it has been used to motivate and design assessments in specific curricula.\(^{20}\)

One approach that could prove fruitful for classroom assessment is a strategy used in an elementary curriculum unit called *Micros and Me* (Tzou et al., 2007). The unit aims to engage students in the practice of argumentation to learn about key ideas in microbiology. In contrast to many curriculum units, however, this example provides students with the opportunity to pursue investigations related to issues that are relevant to them. The researchers adapted a qualitative

---

\(^{19}\)The phrase “disposition to engage” is used in the context of science education to refer to students’ degree of engagement with and motivation to persevere with scientific thinking.

\(^{20}\)One example is *Issues, Evidence, and You*: see Science Education for Public Understanding Program (SEPUP) (1995) and Wilson and Sloane (2000).
methodology from psychology, photo-elicitation, which is used to identify these issues. Research participants take photos that become the basis for interviews that elicit aspects of participants’ everyday lives (Clark-Ibañez, 2004). In *Micros and Me*, at the beginning of the unit, students take photos of things or activities they do to prevent disease and stay healthy. They share these photos in class, as a way to bring personally relevant experiences into the classroom to launch the unit. Their documentation also helps launch a student-led investigation focused on students’ own questions, which are refined as students encounter key ideas in microbiology.

In describing the curriculum, Tzou and Bell (2010) do not call out the practice of self-documentation of students’ personally relevant experiences as a form of assessment. At the same time, they note that a key function of self-documentation is to “elicit and make visible students’ everyday expertise” relevant to the unit content (Tzou and Bell, 2010, p. 1136). Eliciting and making visible prior knowledge is an important aspect of assessment that is used to guide instruction. It holds promise as a way to identify diversity in the classroom in science that can be used to help students productively engage in science practices (Clark-Ibañez, 2004; Tzou and Bell, 2010; Tzou et al., 2007).

**Professional Development**

The framework emphasizes that professional development will be an indispensable component of the changes to science education it calls for (see National Research Council, 2012a, Ch. 10). The needed changes in instruction are beyond our charge, but in the context of classroom assessment, we note that significant adaptation will be asked of teachers. They will need systematic opportunities to learn how to use classroom discourse as a means to elicit, develop, and assess student thinking. The Contingent Pedagogies Project (see Example 4, above) illustrates one way to organize such professional development. In that approach, professional development included opportunities for teachers to learn how to orchestrate classroom discussion of core disciplinary ideas. Teachers also learned how to make use of specific discussion strategies to support the practice of argumentation.

Eliciting student thinking through skillful use of discussion is not enough, however. Tasks or teacher questions also have to successfully elicit and display students’ problematic ways of reasoning about disciplinary core ideas and problematic aspects of their participation in practices. They must also elicit the interests and experiences students bring, so they can build on them throughout instruction. This is part of the process of integrating teaching and assessment. Thus, both
teachers and assessment developers need to be aware of the typical student ideas about a topic and the various problematic alternative conceptions that students are likely to hold. (This is often called pedagogical content knowledge.) In addition, teachers need a system for interpreting students’ responses to tasks or questions. That system should be intelligible and usable in practice: it cannot be so elaborate that teachers find it difficult to use in order to understand student thinking during instruction. (The construct map and its associated scoring guide shown in Chapter 3 are an example of such a system.)

CONCLUSIONS AND RECOMMENDATIONS

The primary conclusion we draw from these examples is that it is possible to design tasks and contexts in which teachers elicit students’ thinking about disciplinary core ideas and crosscutting concepts by engaging them in scientific practices. Tasks designed with the characteristics we have discussed (three dimensions, interconnections among concepts and practices, a way to identify students’ place on a continuum) produce artifacts, discussions, and activities that provide teachers with information about students’ thinking and so can help them make decisions about how to proceed or how to adjust subsequent instruction or to evaluate the learning that took place over a specified period of time.

Questions have been raised about whether students can achieve the ambitious performance expectations in the NGSS. The implementation of the NGSS is a complex subject that is beyond the scope of our charge; however, each of the examples shown has been implemented with diverse samples of students, and there have been students who succeeded on them (although there are also students who did not). The tasks in our examples assess learning that is part of a well-designed, coherent sequence of instruction on topics and in ways that are very similar to NGSS performance expectations. Each example offers multiple opportunities to engage in scientific practices and encourage students to draw connections among ideas, thus developing familiarity with crosscutting concepts.

CONCLUSION 4-1 Tasks designed to assess the performance expectations in the Next Generation Science Standards will need to have the following characteristics:

\[21\] Samples included students from rural and inner-city schools, from diverse racial and ethnic backgrounds, and English-language learners.
• multiple components that reflect the connected use of different scientific practices in the context of interconnected disciplinary ideas and crosscutting concepts;
• reflect the progressive nature of learning by providing information about where students fall on a continuum between expected beginning and ending points in a given unit or grade; and
• an interpretive system for evaluating a range of student products that is specific enough to be useful for helping teachers understand the range of student responses and that provides tools to helping them decide on next steps in instruction.

CONCLUSION 4-2 To develop the skills and dispositions to use scientific and engineering practices needed to further their learning and to solve problems, students need to experience instruction in which they (1) use multiple practices in developing a particular core idea and (2) apply each practice in the context of multiple core ideas. Effective use of the practices often requires that they be used in concert with one another, such as in supporting explanation with an argument or using mathematics to analyze data. Classroom assessments should include at least some tasks that reflect the connected use of multiple practices.

CONCLUSION 4-3 It is possible to design assessment tasks and scoring rubrics that assess three-dimensional science learning. Such assessments provide evidence that informs teachers and students of the strengths and weaknesses of a student’s current understanding, which can guide further instruction and student learning and can also be used to evaluate students’ learning.

We emphasize that implementing the conception of science learning envisioned in the framework and the NGSS will require teachers who are well trained in assessment strategies such as those discussed in this chapter. Professional development will be essential in meeting this goal.

CONCLUSION 4-4 Assessments of three-dimensional science learning are challenging to design, implement, and properly interpret. Teachers will need extensive professional development to successfully incorporate this type of assessment into their practice.
On the basis of the conclusions above, the committee offers recommendations about professional development and for curriculum and assessment development.

**RECOMMENDATION 4-1** State and district leaders who design professional development for teachers should ensure that it addresses the changes called for by the framework and the Next Generation Science Standards in both the design and use of assessment tasks and instructional strategies. Professional development must support teachers in integrating practices, crosscutting concepts, and disciplinary core ideas in inclusive and engaging instruction and in using new modes of assessment that support such instructional activities.

Developing assessment tasks of this type will require the participation of several different kinds of experts. First, for the tasks to accurately reflect science ideas, scientists will need to be involved. Second, experts in science learning will also be needed to ensure that knowledge from research on learning is used as a guide to what is expected of students. Third, assessment experts will be needed to clarify relationships among tasks and the forms of knowledge and practice that the items are intended to elicit. Fourth, practitioners will be needed to ensure that the tasks and interpretive frameworks linked to them are usable in classrooms. And fifth, as we discuss further in Chapter 6, this multidisciplinary group of experts will need to include people who have knowledge of and experience with population subgroups, such as students with disabilities and students with varied cultural backgrounds, to ensure that the tasks are not biased for or against any subgroups of students for reasons irrelevant to what is being measured.

We note also that curricula, textbooks, and other resources, such as digital content, in which assessments may be embedded, will also need to reflect the characteristics we have discussed—and their development will present similar challenges. For teachers to incorporate tasks of this type into their practice, and to design additional tasks for their classrooms, they will need to have worked with many good examples in their curriculum materials and professional development opportunities.

**RECOMMENDATION 4-2** Curriculum developers, assessment developers, and others who create resource materials aligned to the science framework and the Next Generation Science Standards should ensure that assessment activities included in such materials (such as mid- and end-of-chapter activi-
ties, suggested tasks for unit assessment, and online activities) require students to engage in practices that demonstrate their understanding of core ideas and crosscutting concepts. These materials should also reflect multiple dimensions of diversity (e.g., by connecting with students’ cultural and linguistic identities). In designing these materials, development teams need to include experts in science, science learning, assessment design, equity and diversity, and science teaching.