Creating and Using Instructionally Supportive Assessments

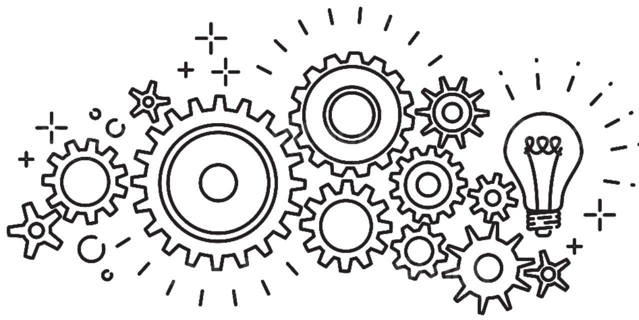
in *NGSS* Classrooms

EDITED BY

Christopher J. Harris
Joseph Krajcik
James W. Pellegrino

nsta Press
National Science Teaching Association

Table of Contents



CHAPTER 2

Creating Assessment Tasks for *NGSS* Classrooms: An Overview of the Design Process

Brian D. Gane, University of Kansas • Samuel Severance, Northern Arizona University • Joseph Krajcik, Michigan State University

The *Framework for K–12 Science Education* (*Framework;* National Research Council, 2012) and *Next Generation Science Standards* (*NGSS*; NGSS Lead States, 2013) require broad shifts in how we conceptualize science learning and how we assess students' progress. Perhaps the most noteworthy shift is from a view of science learning as primarily a process of acquiring science content knowledge toward a view of learning as a process of using and applying disciplinary core ideas in concert with science and engineering practices and crosscutting concepts to make sense of phenomena or solve real world problems (National Research Council, 2012). Referred to as three-dimensional learning, this view emphasizes using and applying the three dimensions in an integrated manner as the means for building the proficiencies expected by the *NGSS* Performance Expectations (PEs). To support and sustain teachers and students in this type of learning, assessment tasks for *NGSS* classrooms must also be three-dimensional so that they can provide information about how students are progressing toward achieving the PEs (National Research Council, 2014). This change in assessment practice represents another important shift for science education. As underscored in Chapter 1, this shift is a different way of thinking about assessment where what matters for measuring students' performance is not just what students know, but also how they use and apply what they know. In order to ensure that we measure what matters for science instruction, we need to begin with a foundational question: *What approach can be used to create assessments that will help you and* *other teachers know whether instructional experiences make a difference for students in building their three-dimensional learning?*

In this chapter, we introduce the *Next Generation Science Assessment* (*NGSA*) design process. The process is a systematic multi-step approach for designing classroom-based assessment tasks that provide evidence of three-dimensional learning so that teachers can gauge students' progress with the *NGSS* PEs. As you will learn, the process reflects an accessible vision for how to design three-dimensional assessment tasks. It also provides a structure that will add consistency for developing a variety of tasks that share common design elements and that align with their target PEs. Using this approach will enable you to fulfill the important requirements for assessment of three-dimensional learning.

We begin by casting a critical eye on an example assessment task and use this task as a starting point for considering what is needed for assessment tasks in general to be instructionally supportive for *NGSS* classrooms. Then, we provide an overview of the *NGSA* design process, briefly describing its six major steps for moving from a PE or bundle of PEs to a set of three-dimensional tasks for classroom use. We end the chapter by summarizing the primary takeaways of using this systematic approach for creating assessment tasks that align with PEs and that provide actionable information about how students are progressing toward meeting them.

Critical Examination of a Three-Dimensional Assessment Task

Imagine an assessment task that was created to align with the three-dimensional view of science learning and could be used to judge students' progress toward meeting an *NGSS* performance expectation. *What might this assessment task look like? How could science teachers use this assessment task in their classrooms to get a sense of what their students know and can do? Would an assessment task that measures three-dimensional learning need to fully address all components of a PE?* Figure 2.1 shows an example assessment task, *Miranda's Mystery Liquids*. The task relates to the physical science topic of Matter and its Interactions and aims to measure students' progress toward meeting MS-PS1-2, a PE from the middle school grade band (see Figure 2.2). *Does responding to this task provide information to the teacher and students as to whether students are building knowledge toward the PE?* In this task, students are presented with a brief scenario-based problem in which they need to use data analysis and interpretation to determine which, if any, of the unknown liquids in a collection might be the same substance. The task requires students to identify relevant patterns in the data table and apply knowledge about the characteristic properties of matter. Students also need to provide evidence and reasoning for what led them to make their determination.

Take a moment to read through the task and then respond to it.

FIGURE 2.1. Example assessment task: Miranda's Mystery Liquids

Miranda was responsible for cleaning up her work area and putting the materials away safely. During class, she used three different liquid substances, but after class, she found four unlabeled bottles of liquid by her desk. To put them away safely, she needs to know which liquids are the same and which are different.

To figure this out, Miranda measured the volume and mass of the liquids, which she used to calculate the density of each. She then tested the boiling point of the liquids. Table 1 shows the data from her investigation.

Table 1. Data of four liquids in different bottles.

Sample	Boiling Point	Mass	Volume	Density
1	100°C	6.10 g	6.10 cm^3	1.00 g/cm^3
2	126°C	5.39 g	6.10 cm^3	0.883 g/cm^3
3	78.4°C	8.05 g	10.2 cm^3	0.789 g/cm^3
4	126°C	9.01 g	10.2 cm^3	0.883 g/cm^3

A) Which information in the table would you use to tell Miranda whether any liquids could be the same substance? Be sure to tell why.

B) Based on the information in the table, which, if any, of the liquids are the same? Support your answer with what you know about the properties of matter.

After working through this task, consider how it relates to the PE from which it was developed, and how you would respond to the following questions:

1. To what extent does the task match with the dimensional elements of the PE?
2. What type of information can this task provide about students' proficiency with the PE?
3. How would this task be helpful in monitoring students' progress in building toward the PE?

FIGURE 2.2. Performance expectation MS-PS1-2 with foundation boxes (NGSS Lead States, 2013). Performance expectations integrate the three dimensions; the foundation boxes provide further information about each dimension.

MS-PS1-2. Analyze and interpret data on the properties of substances before and after the substances interact to determine if a chemical reaction has occurred. [Clarification Statement: Examples of reactions could include burning sugar or steel wool, fat reacting with sodium hydroxide, and mixing zinc with hydrogen chloride.] [*Assessment boundary: Assessment is limited to analysis of the following properties: density, melting point, boiling point, solubility, flammability, and odor.*]		
The performance expectation above was developed using the following elements from the NRC document *A Framework for K–12 Science Education*:		
Science and Engineering Practices	**Disciplinary Core Ideas**	**Crosscutting Concepts**
Analyzing and Interpreting Data • Analyzing data in 6–8 builds on K–5 and progresses to extending quantitative analysis to investigations, distinguishing between correlation and causation, and basic statistical techniques of data and error analysis. • Analyze and interpret data to determine similarities and differences in findings.	**PS1.A: Structure and Properties of Matter** • Each pure substance has characteristic physical and chemical properties (for any bulk quantity under given conditions) that can be used to identify it. **PS1.B: Chemical Reactions** • Substances react chemically in characteristic ways. In a chemical process, the atoms that make up the original substances are regrouped into different molecules, and these new substances have different properties from those of the reactants.	**Patterns** • Macroscopic patterns are related to the nature of microscopic and atomic-level structure.

When responding to the first question, you might have noticed a mismatch between the knowledge and capabilities that the assessment task asks students to demonstrate and what is fully required in the PE. For instance, *Miranda's Mystery Liquids* requires students to use only some elements of the disciplinary core idea (DCI). Students are only applying knowledge about the properties of matter in order to distinguish substances from one another. The task does not involve applying another important element of the DCI, which is focused on chemical reactions and the processes underlying them. Moreover, while both the PE and the task share the same science and engineering practice (SEP) of Analyzing and Interpreting Data, they differ somewhat in their application of the SEP. In *Miranda's Mystery Liquids*, the SEP is applied to determine similarities and differences between characteristic properties of substances. This is in contrast to the requirement of the PE, which is to determine similarities and differences before and after substances interact. The crosscutting concept (CCC) of Patterns is also the same for the task and the PE, but the task emphasizes patterns at the macroscopic scale whereas the PE places the emphasis on relating macroscopic patterns to the atomic level. Overall, the task matches with a number of important dimensional elements of the PE, but it does not fully cover all the PE's terrain.

The second and third questions might be a little harder to answer. *What type of information does this task stand to provide about students' proficiency with the PE? How would this task be helpful in monitoring students' progress in building toward the PE?* Because *Miranda's Mystery Liquids* addresses some of the key demands of the PE, the task at least stands to provide some valuable insight into students' developing proficiency. Noteworthy is that the task elicits some of the knowledge and capabilities students need to demonstrate in order to attain proficiency with the PE. For instance, being able to determine whether substances are the same or different based upon patterns in characteristic properties, as called for in the task, is a smaller three-dimensional performance that is needed for achieving the more comprehensive PE. Accordingly, it may not be a useful task for summative assessment purposes, but it holds promise for use during instruction at a time point when students would be expected to demonstrate proficiency with this essential part of the PE. The value of assessing students on a smaller three-dimensional performance like the one addressed in *Miranda's Mystery Liquids* is that it can help us to see whether students are on a path for building toward and successfully achieving the comprehensive PE. Tasks like these can provide teachers with "just-in-time" information that they can use to determine next steps for teaching and learning while in the midst of instruction. As will be described later in this chapter and elaborated on in Chapters 3 and 4, two features used in our *NGSA* design process—focusing on purposefully selected elements of the DCI, SEP, and CCC, and on occasion swapping in a different but closely related SEP or CCC—are done intentionally to create tasks that cohere and build with instruction over time. As part of this process, we break down PEs into comprehensive sets of smaller performances that in turn can be used to develop assessment tasks suitable for *NGSS* instruction. This enables us to create *instructionally supportive* tasks that can serve as markers to help teachers and students gauge how students are building in their proficiencies.

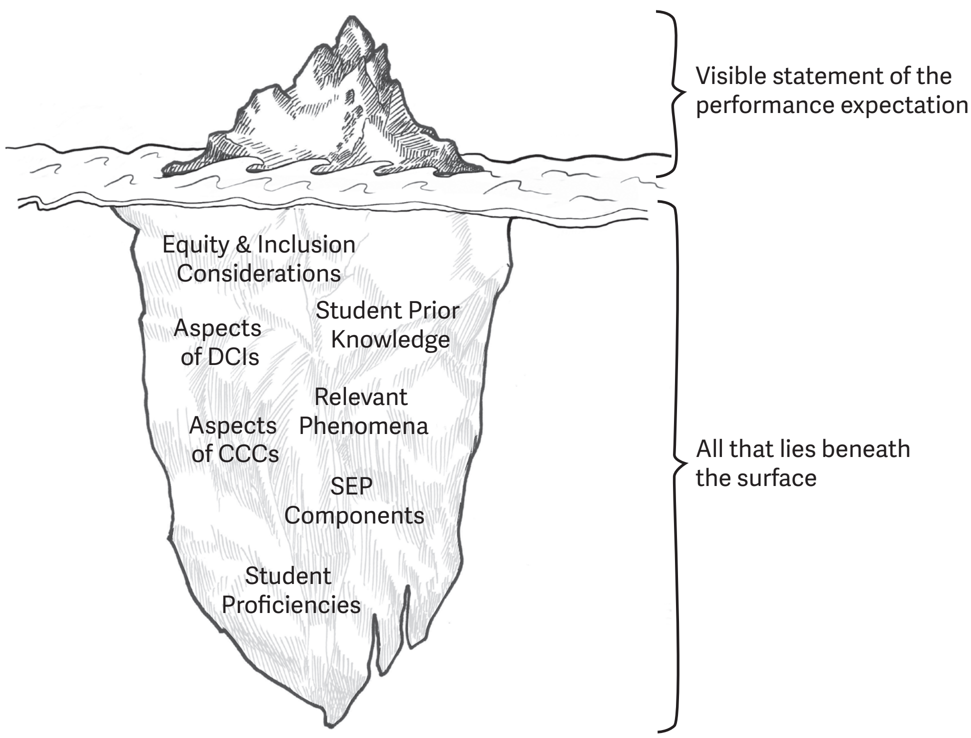
The Importance of Assessing Students' Progress Toward Achieving a PE

To understand how to assess a PE, one needs a deep understanding of all the elements encompassed within the PE. This is an important and worthwhile undertaking for anyone planning to develop three-dimensional tasks. If you have spent some time examining the *NGSS* PEs, you might have noticed several overarching characteristics:

- PEs are *statements* that describe what students should know and be able to do at the end of a grade or grade band. They are very broad in scope and were developed from the *Framework* (National Research Council, 2012). Each PE takes the form of a single statement that describes competencies at a large grain size, rather than going into the details that underlie them.
- PEs are *three-dimensional*: A PE always includes a DCI, SEP, and CCC. The dimensions are described in the *Framework*, and their elements (i.e., essential aspects of each dimension) are further described in the *NGSS*. The elements of the dimensions vary from PE to PE within a grade level or grade band, and they become more sophisticated as students progress through K–12.
- The three dimensions in the PEs are *integrated*: They describe how students use DCIs and CCCs by engaging in SEPs to make sense of phenomena or solve problems. The dimensions are not intended to be used in isolation from one another.

Expanding on these characteristics, a PE is analogous to an iceberg. The small portion of the iceberg that shows above the water's surface seems easy to navigate, but below the water's surface lies the remainder: a massive portion of the overall iceberg (see Figure 2.3). Similarly, a succinct PE statement such as MS-PS1-2, "Students who demonstrate understanding can analyze and interpret data on the properties of substances before and after the substances interact to determine if a chemical reaction has occurred," is really just the tip of the iceberg; students need to develop multiple proficiencies as they move toward achieving this PE. Just as it is unwise to ignore the portion of an iceberg below the water's surface when traveling through icy waters, we should not ignore the details that underlie each dimension within a PE. This becomes readily apparent when we consider the surface-level features of MS-PS1-2. For example, the DCI portion of the PE indicates that students will apply their knowledge about the properties of substances to determine if a chemical reaction has occurred. But what is expected of middle school students' knowledge of properties, substances, and chemical reactions? These details are what lies below the tip of the iceberg. The CCC portion, Patterns, is not explicit in the one-sentence PE statement, yet it is essential for students to identify patterns and reason with patterns in order to achieve this PE. *What are the types of patterns that students will need to use?* The SEP is clearly stated in the PE, with a focus on analysis of data to identify similarities and differences before and after substances interact. *But what grade level skills and abilities related to the SEP will students be expected to use?* By examining all that lies beneath the tip of the iceberg—or in this case the PE statement—we can more readily grasp the vast expanse. Once this expanse is mapped, it becomes clear that students' learning toward meeting a PE or a PE bundle must encompass a surprising amount.

FIGURE 2.3. Immediately visible portion of an *NGSS* Performance Expectation and what is called for beneath the surface



The *NGSA* design process is intended to assist task designers, whether classroom teachers, state and district educational leaders, or curriculum and assessment developers, in navigating through the PE icebergs with an eye toward understanding all of what lies beneath the surface. PEs are comprehensive and considered summative goals, and therefore need to be learned over time and through a sequence of lessons and instructional units. Alongside instruction, assessment tasks are needed that can be used—over time and in a formative manner—to determine where students are in their progress toward meeting the complex PEs. The *NGSA* design process can be used to unpack and identify the meaningful parts of the PEs that will be suitable for classroom-based assessment. The process emphasizes using the meaningful parts to construct comprehensive sets of smaller performance statements that we call *learning performances*. Learning performances are intermediary performance targets for instruction and assessment that can signal whether students are moving along a productive path to proficiency with a PE or PE bundle.

Learning Performances Help Instruction and Assessment to Work Together

Both instruction and assessment must work together to effectively help students build proficiency with the *NGSS* PEs over time (National Research Council, 2014). Instruction should help students build toward the PEs so that they can achieve the PEs for their grade level or grade band by the end of instruction. In parallel, assessment should provide a window into students' progress in building toward them. In the *NGSA* approach to assessment design, we use learning performances as three-dimensional learning goals that are smaller in scope than the performance expectations. These learning performances take the form of knowledge-in-use statements that incorporate aspects of DCIs, SEPs, and CCCs that students need to develop understanding of as they progress toward achieving a single PE or bundle. Learning performances are expressed using similar language as PEs and emphasize knowledge-in-use, but they are crafted at a more specified and manageable grain-size for classroom assessment purposes (Harris, Krajcik, Pellegrino & DeBarger, 2019). For example, the learning performance from which *Miranda's Mystery Liquids* (Figure 2.1) was developed is: *Students analyze and interpret data to determine whether substances are the same or different based upon patterns in characteristic properties*. What is foremost to call attention to is that this learning performance covers part of the multidimensional terrain that resides under the surface of the PE. Learning performances are useful in part because they are assessable in the midst of instruction, meaning that they provide opportunities to assess students as they are developing proficiency toward the larger PE.

Reexamination of the Example Task

Now that we have considered the role and utility of learning performances as classroom-based markers for monitoring students' progress toward meeting complex multidimensional PEs, consider *Miranda's Mystery Liquids* and those three questions again:

1. To what extent does the task match with the dimensional elements of the PE?
2. What type of information does this task stand to provide about students' proficiency with the PE?
3. How would this task be helpful in monitoring students' progress in building toward the PE?

Look carefully at the features of the task and how you responded to the task's prompts in light of the PE and also the learning performance. Although the task does not align fully with all parts of the PE, it does align with the parts of the PE that are represented in the learning performance. This learning performance describes an important three-dimensional requirement of the PE that students would need to achieve at some point during instruction as they progress toward meeting all that underlies the PE. Accordingly, the task stands to provide evidence on whether and how students are able to use their knowledge as specified in the learning performance. Such evidence is of high value because it enables us to assess "building toward" the PE.

Now that we have described our rationale for creating intermediary targets (i.e., learning performances) around which assessment tasks can be developed, we are ready to describe our process. In the remainder of this chapter, we provide an overview of the *NGSA* design process which can be used to construct instructionally supportive tasks that assess for three-dimensional learning. The process is introduced here and then explained in further detail in Chapters 3–7.

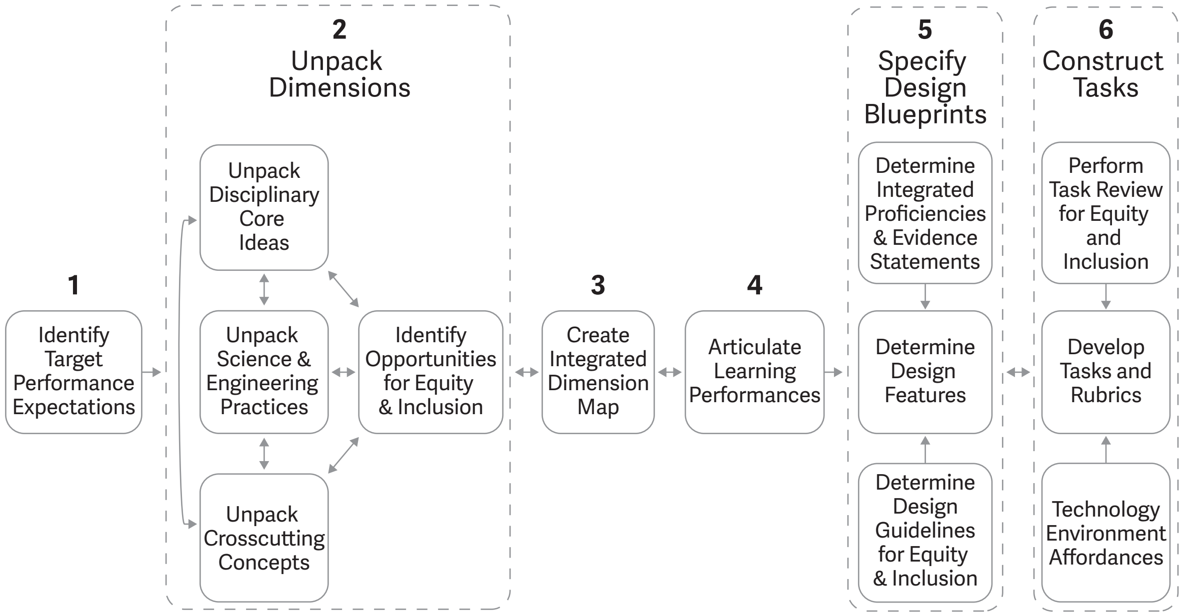
The *NGSA* Design Process: An Overview

Drawing conclusions about student performance from an assessment is essentially an effort of evidence-based argumentation. Assessment experts Robert Mislevy and colleagues (Mislevy, Steinberg & Almond, 2003; Mislevy & Haertel, 2006) used this idea as the basis for creating an assessment design framework called evidence-centered design (ECD). The ECD framework emphasizes the value of starting with a learning goal and determining the evidence that you would look for to make a judgment about students' performance of that learning goal, then specifying the features of tasks that will best bring out the evidence of performance. From the ECD perspective, the end goal of assessment is to make a claim about what students know and can do. In this endeavor, you must gather evidence to support that claim. This evidence typically takes the form of what students say, write, draw, or do in response to an assessment task.

Since it was proposed two decades ago, ECD has gained widespread attention in education as a worthwhile framework for principled assessment design. Noteworthy is that shortly after the release of the *NGSS*, the National Research Council released a comprehensive report on developing assessments for the *NGSS* (NRC, 2014). The report recommends that assessment design approaches follow the reasoning of ECD to ensure that tasks measure what matters for three-dimensional learning. The argumentative reasoning of ECD is integral to the *NGSA* design process for developing tasks that will provide evidence of students' three-dimensional performance as they build toward meeting the PEs.

The *NGSA* design process, illustrated in Figure 2.4, provides guidance on how to use PEs as the starting point for developing three-dimensional assessment tasks that can be used in the classroom to inform *NGSS* teaching and learning. This process allows us to create a set of learning performances from a PE or bundled group of PEs in a principled way. We then use the learning performances to guide the development of assessment tasks and accompanying rubrics. The process involves six major steps across three phases. The first phase, Steps 1–3, involves selecting a PE or PE bundle and systematically unpacking the dimensions to understand the assessable components. The elaborations from the unpacking are used to create a visual representation in the form of a map that lays out the dimensional "terrain" for fully achieving the PE or bundle. We refer to the map as an *integrated dimension map*. The second phase, Step 4, entails using the integrated dimension map in tandem with the unpacking to articulate and refine a set of learning performances that describe the proficiencies that students will need to demonstrate over time as they progress toward achieving the more comprehensive PE or PE bundle. The third phase, Steps 5–6, involves an organizational strategy called *design blueprints* for using learning performances to construct assessment tasks. Design blueprints provide the essential technical information for developing tasks that assess for three-dimensional learning. A single blueprint describes all of the major design decisions for creating one or more assessment tasks along with rubrics that each align to a learning performance.

FIGURE 2.4. The six steps of the *NGSA* task design process



The Important Role of Equity and Inclusion in the *NGSA* Design Process

It is widely recognized today that designing and enacting three-dimensional assessments that promote equity and inclusion are a responsibility of all science educators. In *NGSS* classrooms, instruction and assessment should work together to attract all students to science learning and support and sustain their motivation and engagement. In the case of assessment, tasks should be designed and used in a way that both leverages and values students' background knowledge and experiences and connects them to rigorous science learning. When we accomplish this, we create the opportunity for the full range of students to demonstrate their three-dimensional learning, particularly those whose backgrounds have been underrepresented in science education and careers including girls, students who identify as LGBTQ, students with disabilities, underrepresented students of color, students with low-income family backgrounds, and emerging multilingual learners.

Equity and inclusion considerations are woven throughout the *NGSA* design process, beginning with the initial steps of selecting and unpacking the PEs and continuing through to the final products of tasks and rubrics. To support three-dimensional assessment design that promotes equity, guidance is provided for developing tasks that use scenarios where the phenomena are likely to have broad interest or universal relevance for students, explicitly attend to language, and include scaffolds to make expectations explicit for students. To promote inclusion, guidance is provided for how to reduce bias in tasks, eliminate barriers that may interfere with student sensemaking, and support engagement so that students will be more likely to persist in reading and responding to tasks. When both equity and inclusion considerations are made, it becomes clear that assessment must allow for multiple ways for students to demonstrate their developing proficiencies and that using traditional or narrow assessment techniques and formats will not create an opportunity for students' multiple modes of performance to be recognized. The *NGSA* design process brings equity and inclusion considerations to the forefront so that tasks are accessible and fair for a wide range of students with varying backgrounds, skills, and abilities.

The Six Steps of the *NGSA* Design Process

What follows is a run-through of the six steps of the *NGSA* design process. Table 2.1 summarizes the key components of each step.

Step 1: Select Performance Expectation(s)

The first step in the design process is to select a target PE or a coherent bundle of PEs appropriate for classroom instruction. An appropriate PE or PE bundle should match with instruction such that the sequence of lessons and activities will provide an opportunity for students to build the knowledge and capabilities required by the PE or PE bundle over time. When there is a strong match between the PEs and the instruction, then it is suitable for the PE or PE bundle to serve as the focal point for developing three-dimensional assessment tasks.

Step 2: Unpack the NGSS *Dimensions*

One of the major pitfalls in designing assessment tasks for the *NGSS* is mistakenly believing that the one-sentence PE statement can be fully understood just by reading it. Similar to the iceberg (Figure 2.3), the PE in its entirety is very deep, and as designers we need to understand all the parts that are involved. Unpacking is a foundational step for detailing all that lies beneath the surface of a PE or PE bundle. The unpacking step is of high value because it enables designers to identify and elaborate on all that is involved for students to successfully demonstrate integrated proficiency. Importantly, unpacking focuses attention on the specific and meaningful aspects of the three dimensions, as well as the knowledge and capabilities that students need to develop along each of those dimensions. When information from a careful and thorough unpacking is documented, it becomes an important resource that can be used time and again during the design process to support and verify key design decisions.

Unpacking helps designers develop deep knowledge of the aspects of proficiency inherent within each dimension. It also focuses attention on identifying potential intersections between dimensions, thereby providing the foundation for integrating them later in the design process. Noteworthy too is that when unpacking the dimensions, it is of value to consider other SEPs and CCCs that could productively work together along with the DCI to build toward the PE or PE bundle. These considerations play an important role in the upcoming steps of mapping the dimensions and articulating learning performances.

When unpacking the dimensions, it is of paramount importance to always unpack with the student in mind. This includes describing students' prior knowledge and identifying likely student challenges with the dimensions; defining boundaries of what students should know and be able to do; earmarking issues of equity and inclusion that are relevant to the dimensions; identifying candidate phenomena relevant to both the PE and students' everyday lives and interests; and sketching out possible realistic scenarios that can provide a motivating context for making sense of phenomena.

As an example of unpacking the dimensions, consider PE MS-PS1-2 shown in Figure 2.2, which uses the SEP Analyzing and Interpreting Data, the elements from DCI PS1.A Structure and Properties of Matter, and the CCC Patterns. The DCI dimension is expressed within the portion of the PE focusing on the "properties of substances before and after the substances interact to determine if a chemical reaction has occurred." This part of the PE contains a considerable amount of disciplinary knowledge about the structure and properties of matter, thereby requiring that students must grasp a number of major ideas such as the following:

- Substances and their properties.
- How substances can interact.
- That new substances can be made from original substances.
- That changes in properties of substances from before they interact to afterward can serve as evidence that new substances are formed.

If we take a deeper dive and examine the two DCI elements underlying this PE (see the foundational boxes in Figure 2.2), several more ideas become visible, including:

- Pure substances have characteristic physical and chemical properties that can be observed and/or measured.
- Characteristic properties can be used to identify a substance and distinguish it from other substances.
- Substances react chemically in characteristic ways with other substances.
- During a chemical process, atoms in a substance can rearrange or regroup in new ways to form different molecules.

Important to note is that unpacking is much more involved than simply listing the various ideas found in the elements of the DCI. Unpacking also involves determining what students at the grade level are expected to know about the idea. In this case, we are considering middle school students. As an example, let's further unpack the last bullet, "During a chemical process, atoms in a substance can rearrange or regroup in new ways to form different molecules." To grasp this sub-idea, a middle school student would have to know that a molecule of a substance is always made up of the same type and number of atoms and that during a chemical reaction, the atoms can rearrange into new molecules to form new substances. These new molecules will always have the same type and number of atoms, just like the substances before the chemical reaction occurred. In this way, unpacking helps to establish the boundaries for the range and depth of DCI knowledge that students must attain as they build toward the PE. Also important is determining the prerequisite knowledge students should be expected to already possess, as well as any knowledge that goes beyond what students should know at the grade level. Closely related, it is important to identify known student challenges. In the case of DCI PS1.A, students typically have a difficult time visualizing how the arrangement and motion of atoms at the microscopic level can explain what is observed at the macroscopic level.

Another activity during the unpacking of a DCI is to identify phenomena that students can engage with that align with the DCI elements. Determining candidate phenomena and sketching possible scenarios to use in assessment tasks is an important part of unpacking that will directly inform task design. An engaging phenomenon or complex problem in an assessment task sets the stage for how students will engage with the task and provides a meaningful context for students to make sense of the task requirements. Importantly, a rich phenomenon and scenario can activate the appropriate knowledge and capabilities required for working through the task and demonstrating performance. To accomplish this, the phenomenon, or problem, needs to be relevant and engaging for students. Students should be able to use elements of the DCI with SEPs and CCCs to reason through and figure out the phenomenon or problem at hand, just as any scientist would. Ideally, the phenomenon or problem will allow students to engage not with "someone else's" science, but with science that they see as relevant to their own lived experiences. Further, these phenomena should be contextualized in compelling scenarios so that they create a "need-to-know" for students, and they should be complex enough that students must use all three dimensions to make sense of them. Achieving these aims, particularly while also ensuring the phenomenon or problem can be explained or solved using grade-appropriate knowledge and practices, is no small feat.

In tandem with the unpacking of the DCI, the SEP and CCC are also unpacked. When unpacking an SEP, the focus is on clearly articulating the essential grade band appropriate performance for the SEP. This includes specifying the aspects of the SEP that students are to perform as well as specifying the evidence required for students to demonstrate a high level of proficiency, identifying prior knowledge that is required of students to demonstrate the SEP, and identifying common challenges that students may encounter as they are developing sophistication with it. We also identify productive intersections between the SEP and other SEPs that are relevant for building toward the PE or PE bundle. When unpacking the CCCs, the focus is on identifying the important aspects of the CCC, as well as how the CCC intersects with both the SEP and particular sub-ideas of the DCI. Similar to unpacking SEPs, it is important to identify common challenges and to specify the evidence required for a student to demonstrate a high level of proficiency with the CCC.

The unpacking of all three dimensions is foundational for designing good assessment tasks and accompanying rubrics. Importantly, the unpacking step can promote consistency in the use of the dimensions throughout the design process. Detailed guidance for how to unpack is presented in Chapter 3.

Step 3: Map the Dimensions

The third step in the *NGSA* design process entails using the dimension elaborations from the unpacking to create what we call an *integrated dimension map* that provides a visual representation of the target PE or bundle. Mapping synthesizes the information from unpacking to lay out the DCI terrain and visually represent the most salient and productive intersections across the DCI, SEP, and CCC. The end result of mapping is an integrated dimension map that expresses the key relationships between the DCI elements that were elaborated in the unpacking and identifies how aspects of the SEP and CCC (also elaborated in the unpacking) can work with these disciplinary relationships to promote students' integrated proficiency.

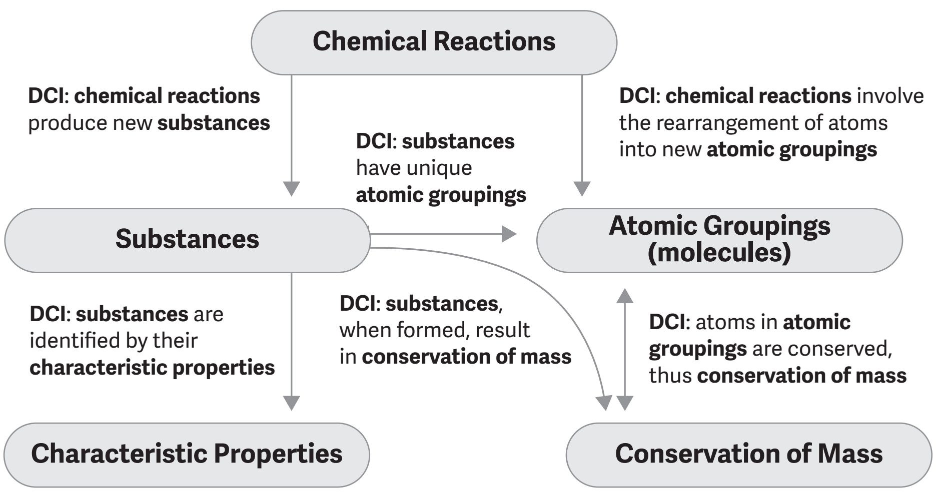
TABLE 2.1. Key Components of the *NGSA* Design Process

Key Design Components
1. Performance Expectations
• Comprehensive knowledge-in-use learning goals to be achieved *by the end* of instruction • Starting point for the *NGSA* Design Process • An appropriate PE or PE bundle will match with instruction that aims to build the knowledge and capabilities required by the PE or PE bundle over time.
2. Unpacking
• Lays out the essential aspects of each dimension and identifies productive intersections within and across dimensions • Promotes consistency in the use of dimensions during task design • Documents equity and inclusion considerations such as students' prior knowledge, anticipated challenges, and language and literacy demands • Identifies phenomena relevant to both the PE and students' interests as well as relatable scenarios that connect with students' everyday experiences
3. Integrated Dimension Maps
• Visual representation of the essential dimension elements and their sub-idea relationships • Shows the area and boundaries of a performance expectation or bundle
4. Learning Performances
• Knowledge-in-use statements that take on the three-dimensional structure of a PE but are smaller in scope and align to only a portion of the PE • Function in relation to other learning performances to cover the entire "terrain" of a PE or bundle • Provide intermediary performance goals for creating assessment tasks for classroom use
5. Design Blueprints
• Guide the principled development of tasks aligned to learning performances • Identify what should be included or not included in tasks and ensure that critical specifications are used consistently • Specify the essential features that all tasks must include and the variable features that can vary among tasks • Set down equity and inclusion guidelines so that tasks are accessible and fair for a wide range of students • Include the integrated proficiencies and evidence statements that demonstrate performance
6. Three-Dimensional Tasks & Rubrics
• One or more tasks can be created from a single design blueprint • Task should include a phenomenon that is contextualized in a scenario that encourages students to engage with the phenomenon and work through the task • Responding to the task should require students to use the aspects of the SEP, DCI, and CCC targeted by the learning performance • An integrated three-dimensional assessment task requires an integrated three-dimensional rubric

To develop an integrated dimension map, we start by laying out the major science sub-ideas of the DCI that was unpacked and specifying the relationships between them. Similar to a concept map, the mapping initially takes the form of a diagram that depicts the key sub-ideas with arrows that show the relationships among those sub-ideas. Figure 2.5 illustrates the mapping of the sub-ideas and grade-appropriate relationships for two bundled PEs, MS-PS1-2 and MS-PS1-5. These PEs are ideal to bundle together because they are complementary in their focus on chemical reactions and would likely be taught and assessed within the same instructional sequence or unit.

- MS-PS1-2. Analyze and interpret data on the properties of substances before and after the substances interact to determine if a chemical reaction has occurred.
- MS-PS1-5. Develop and use a model to describe how the total number of atoms does not change in a chemical reaction and thus mass is conserved.

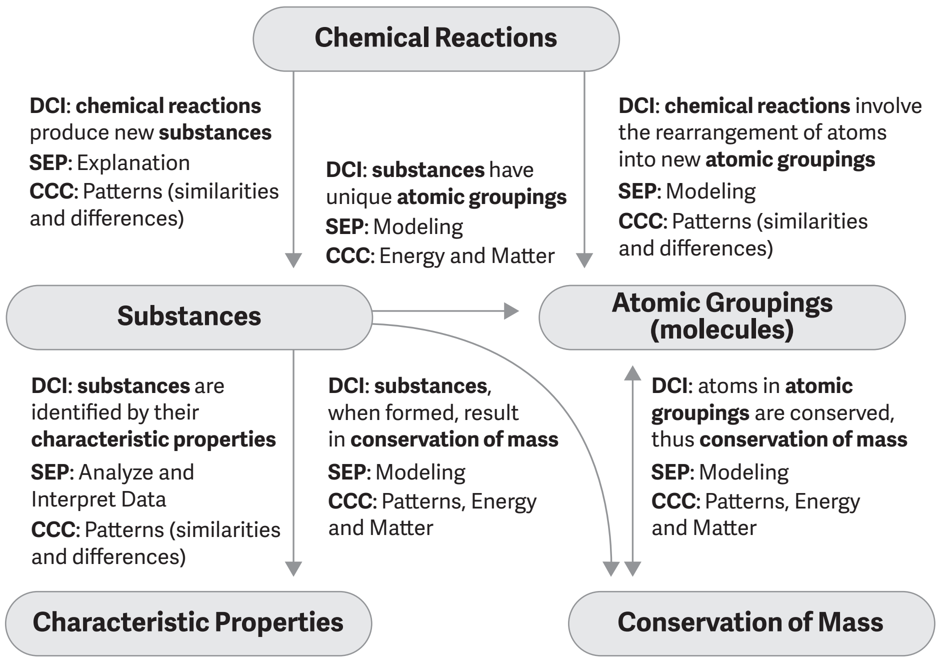
FIGURE 2.5. A dimension map showing the elements of the disciplinary core idea(s), and how they relate to each other, for PEs MS-PS1-2 and MS-PS1-5



Once the key science sub-ideas of the DCI are laid out, we overlay the relevant SEPs and CCCs onto the map as shown in Figure 2.6. We add only the SEPs that students could use to demonstrate their understanding of the DCIs and CCCs. For instance, the SEP Computational Thinking, while an important practice, does not work as well with these sub-ideas as it might with others. Next, we add CCCs that students could engage with and connect to the DCIs, or that might arise from their engagement with the SEP. We explicitly represent and consider each of the three dimensions and how they relate to one another in this step.

This theme of integration occurs throughout the design process to ensure that all three dimensions are incorporated into the tasks and rubrics.

FIGURE 2.6. An integrated dimension map showing the relationships between the elements of the disciplinary core idea(s), crosscutting concepts, and science and engineering practices for the bundled PEs MS-PS1-2 and MS-PS1-5



Step 4: Articulate Learning Performances With Multiple Opportunities for Access

The fourth step involves using the integrated dimension map as the starting point for articulating and refining a set of learning performances that collectively describe the proficiencies that students need to demonstrate in order to meet a PE or PE bundle. A single learning performance is crafted as a knowledge-in-use statement that is smaller in scope and covers a designated area of an integrated dimension map. Like a PE, each learning performance has a three-dimensional structure that helps maintain the requirement that students use their knowledge to make sense of phenomena or solve problems.

The integrated dimension map shown in Figure 2.6 was used to create the learning performance we introduced earlier in the chapter. This learning performance was created from the lower left region of the map where the major ideas of *substances* and *characteristic* *properties* are linked. The two major ideas are linked by a DCI sub-idea, the SEP Analyzing and Interpreting Data, and an aspect of the CCC Patterns. The components from this area of the map were brought together to develop a three-dimensional learning performance. We started with the action of the SEP and then added the DCI elements and the CCC to arrive at the learning performance: Students analyze and interpret data to determine whether substances are the same or different based upon patterns in characteristic properties.

You likely noticed that the learning performance we specified covers a relatively modest portion of the integrated dimension map. This same map can be used to articulate multiple learning performances for the PE bundle. All told, we articulated six learning performances for the bundled MS-PS1-2 and MS-PS1-5, one of which is the learning performance we have been exploring. As a rule, the set of learning performances should collectively describe the proficiencies that students need to demonstrate in order to meet a PE bundle. In this way, learning performances are intended to function together *as a set* that covers all areas of an integrated dimension map. An individual learning performance covers a smaller area; all the learning performances together address the entire PE bundle. Chapter 4 provides pointers on how to create sets of learning performances.

Another noteworthy feature of a learning performance is that it can use different SEPs and CCCs in concert with elements of the DCI. For instance, the target performance expectation might use one SEP (e.g., Analyzing and Interpreting Data) and one CCC (e.g., Patterns) but the learning performances could use alternate SEPs (e.g., Constructing Explanations) and CCCs (e.g., Cause and Effect). This integration of other complementary SEPs and CCCs is aimed at providing students with varied opportunities to engage with elements of the DCI in meaningful ways. A benefit of this feature is that it allows for multiple modes of performance by students to demonstrate their proficiencies. Also, there can be a real learning benefit for varying the SEPs. For example, some SEPs have greater linguistic demands (e.g., Constructing Explanations, Engaging in Argument from Evidence) and stand to provide students, including those with emerging language and comprehension skills, with increased opportunities to practice and develop their linguistic capacities, thereby promoting all students' learning (NGSS Lead States, 2013, Appendix D). Further, integrating DCI elements with various SEPs and CCCs supports students in flexible use of all three of the dimensions, providing for increased opportunities to develop their capabilities for using and applying knowledge to make sense of phenomena or solve problems.

A single learning performance is a claim that we make about what students know and can do regarding an essential part of a PE or PE bundle. Because learning performances specify what students should know and do, they serve as the keystone of our assessment argument. We use a learning performance to design one or more assessment tasks, which provide observable evidence (in the form of student responses) that we can use to judge whether the student can use the knowledge in the learning performance. For example, students' responses to the task *Miranda's Mystery Liquids* can be used as evidence of proficiency in using and applying the knowledge of the learning performance.

Step 5: Specify Design Blueprints

In this step, we use an organizational strategy called *design blueprints* to guide the principled development of tasks. A design blueprint is a document that brings together all that must be taken into account when creating tasks for each learning performance. The value of using blueprints is that they can make implicit design decisions *explicit* for those who will be doing the actual constructing of tasks and rubrics. Importantly, blueprints set the boundaries for what should be included or not included in tasks and they ensure that critical specifications, like task features, scaffolding levels, and format types are used consistently. When designers construct one or more assessment tasks for the same learning performance, using design blueprints increases the likelihood that the tasks will include all of the essential features from the unpacking and mapping, and that they will elicit the intended evidence from students. On the other hand, when task designers do not follow a set of specifications, important features may be overlooked and the end result may be tasks that do not measure what was intended. Also, when tasks are constructed primarily based on the designer's intuitions about what makes "good" tasks, that knowledge resides with the individual and may not always be readily recalled or consistently applied. Moreover, the designer's intuition cannot be efficiently taught nor reliably replicated by others who are not aware of the tacit knowledge held by the assessment designer.

When specifying a blueprint, we concern ourselves primarily with documenting and making explicit our thinking about what observable evidence the task will provide, why and how the task will provide that evidence, and why that evidence is relevant to the claim (i.e., the learning performance). To document this thinking, we create a design blueprint that is used to specify features needed in a task so that it aligns with the learning performance. In the end, each learning performance will have a design blueprint that outlines various features of a task that a designer needs to consider when developing any task that aligns to the learning performance. Think of a design blueprint as a guide with directions to follow in developing tasks. A value of design blueprints is that they can be used to make not just one task, but multiple tasks that are each aligned to the same learning performance.

When producing blueprints, we organize the specifications around the following five practical design questions:

1. What should students know and be able to do in order to demonstrate this learning performance?
2. What could students say or do that would provide evidence that they can use the knowledge described in this learning performance?
3. What are the essential features that all tasks constructed for this learning performance will need to include?
4. What are the variable features that some tasks (but not all tasks) constructed for this learning performance will need to include?
5. What equity and inclusion considerations are necessary to include to ensure that tasks constructed for this learning performance will be accessible and fair for a wide range of students?

As we answer these questions, we document our responses and label them as *integrated proficiencies*, *evidence statements*, *essential task features*, *variable task features*, and *equity and inclusion considerations* respectively. *Integrated proficiencies* refer to the abilities that students must employ to demonstrate the learning performance. For each integrated proficiency, we identify the observable evidence that students need to provide in order to show that they can use and apply the knowledge in the learning performance. We refer to these as *evidence statements* and they are the second key component of our "assessment argument." Importantly, they provide direct guidance for what we should look for in a student's performance. Because learning performances are multidimensional, our evidence statements are also multidimensional. An evidence statement succinctly describes the information we need to observe in a student's response to a task so that we can defend the claim made about what that student knows and can do.

An assessment should prompt students to say, do, or create something with the aim of getting them to provide observable data evidence that demonstrates their current level of integrated knowledge use of the DCIs, CCCs, and SEPs. For example, for the learning performance, *Students analyze and interpret data to determine whether substances are the same or different based upon patterns in characteristic properties,* the integrated proficiencies might include:

A. Ability to analyze data on substances to identify patterns in characteristic properties.

B. Ability to describe that two or more unknown substance samples are the same substance or different substances based on patterns in characteristic properties.

The corresponding evidence statements might be:

A. Students identify patterns in data regarding the characteristic properties (e.g., density, melting point, boiling point) of substances.

B. Students use evidence from data to write a conclusion with reasoning that if the same properties are found in multiple unknown samples, then those samples are the same substance because they share all of the same characteristic properties.

You can learn more about integrated proficiencies and evidence statements, as well as how to write them, in Chapter 4.

Next, we describe essential and variable task features. Think of *essential task features* as the features that all tasks need to include to meet a particular learning performance. For instance, all tasks constructed for the learning performance we are discussing must have a scenario that presents samples of substances that can be identified using characteristic properties at the middle school level. On the other hand, *variable task features* specify the important features that can vary in some form or by option across tasks. For meeting our learning performance, the type of substance samples and their states—whether gases, liquids, or solids—could vary across different tasks and yet still meet the requirements of the evidence statements. A helpful analogy that can further distinguish essential and variable task features is car manufacturing. When manufacturing cars, all cars that come off the assembly line must have an engine that will provide the energy for motion. Accordingly, the engine is an essential feature that all cars must share. However, the type of engine (e.g., fuel engine or electric motor) can differ and we still have a car. So, the engine type is a variable feature among cars. Describing the essential and variable task features provides the third key component of a coherent assessment argument. This component enables the development of multiple tasks that all share the same essential features yet may still differ in ways that matter for when and what purposes you might use them. Chapter 5 provides further details on identifying and using essential and variable task features.

Design blueprints also include *equity and inclusion considerations* that make explicit how tasks will leverage and value students' background knowledge, capabilities, and experiences. Our emphasis is on developing tasks that enable all students to show how they can use and apply what they know. To achieve this, designers must give attention to a range of matters, including considering language and sentence structure, including ways to reduce bias and increase inclusiveness, identifying phenomena that are relevant to students' lives, and incorporating features that enable a mixture of modes for demonstrating performance. Chapter 5 provides guidance on determining equity and inclusion guidelines and incorporating them into design blueprints.

Step 6: Construct Tasks and Rubrics

Once all the specifications (i.e., integrated proficiencies, evidence statements, essential task features, variable task features, and equity and inclusion considerations) are organized into a design blueprint, the final step is to construct the assessment tasks and accompanying rubrics (see Chapters 5 and 7). Writing a task involves selecting a compelling phenomenon or problem, describing or representing that phenomenon or problem in a scenario, writing the task prompts, and defining how students with different levels of proficiency will respond to the task (part of which involves constructing a rubric). It is imperative to select a phenomenon or problem that will allow students to make use of their knowledge and demonstrate the extent to which they can use their knowledge as required by the learning performance. A well-written scenario will introduce the phenomenon and help students home in on what they need to figure out or explain. Together, the scenario and phenomenon should create a compelling need for students to engage in reasoning about that phenomenon and work through the task.

The design blueprint serves as a guide in writing tasks. When constructing a single task, referring back to the design blueprint is critical to ensure the task can provide the evidence needed to make an argument that the learner has met the knowledge encompassed in the learning performance. When designing multiple tasks from a blueprint, each task should fully stand alone in representing the learning performance.

The task prompts should be written in a manner that elicits an integrated three-dimensional response, and responding to the task should require students to use the target aspects of the SEP, DCI, and CCC of the learning performance. For instance, *Miranda's Mystery Liquids* (see Figure 2.1) is a task that requires students to use all three dimensions in an integrated manner. In this task, students are first asked to consider what data they could use to identify which samples are from the same substance. To be able to do so, students need to know that density and boiling point are characteristic properties, that mass and volume are not characteristic properties, and that characteristic properties are unique to a substance. These ideas are related to the DCI and necessary for applying the SEP. The second prompt asks students to make a conclusion about which substances are the same. To successfully respond, students need to analyze the data in the table (the SEP), look for relevant patterns (the CCC), and identify which samples could be the same substance because of their properties (the DCI). They must use the dimensions in an integrated manner to make a conclusion that liquids in samples 2 and 4 are the same because they have similar densities and boiling points. As such, students make use of all three dimensions of scientific knowledge, providing evidence that they can use the knowledge in the ways specified in the learning performance.

Once the task is written in a manner that requires students to use all three dimensions, we develop a rubric that can be used to evaluate student responses (see Chapter 7). In short, an integrated three-dimensional assessment task requires an integrated three-dimensional rubric. The rubric needs to provide information on all three dimensions. A good rubric provides the teacher and student with useful, multidimensional information about a student's proficiency. To develop rubrics, we need to make use of the integrated proficiencies and evidence statements identified as part of writing the task design blueprints (Step 5). The rubrics are constructed using multiple parts, with each part representing a different integrated proficiency. Crucially, these rubrics preserve the multidimensional nature of the integrated proficiencies, meaning that the three *NGSS* dimensions are not separated into different rubrics.

Another important activity in this design step is to conduct a careful review of the task before it is finalized and deemed ready for classroom use. This review includes making sure that the task meets all of the design specifications in the blueprint and that it is accessible and fair to a broad range of students. Regarding the latter, we conduct an equity and inclusion review to ensure that the assessment task values students' background knowledge and experiences, and that the task enables all students to show how they can use and apply what they know. Following the review, we modify tasks as needed to address any problematic issues that were flagged.

When using assessments that are intended to be instructionally supportive, it is generally not useful to only determine whether students can or cannot do something. It is most useful to know what their performance looks like, the varying levels of performance among students, and how you might build on these performances over time (see Chapters 7 and 8). To this end, the *NGSA* design process can be used to develop task rubrics that determine the range of students' three-dimensional performances within a given classroom and provide guidance for next instructional steps (see Chapter 7).

Primary Takeaways

In this chapter, we introduced a process for systematically designing assessment tasks that can be used in the classroom to inform *NGSS* teaching and learning. The assessment design process is of high value because it enables various designers—whether classroom teachers, state and district educational leaders, or curriculum and assessment developers—to develop three-dimensional tasks and accompanying rubrics that are instructionally supportive. The primary takeaways from this chapter are:

1. Both instruction and assessment must work together to effectively help students build proficiency with the *NGSS* PEs over time. Instruction should help students build toward the PEs, and assessment tasks for *NGSS* classrooms need to be three-dimensional so that they can provide information about how students are progressing toward achieving them.

2. The *Next Generation Science Assessment* (*NGSA*) design process provides step-by-step guidance on how to use PEs as the starting point for developing three-dimensional assessment tasks. The process involves six major steps with equity and inclusion considerations woven throughout. It brings equity and inclusion considerations to the forefront so that tasks are accessible and fair for a wide range of students with varying backgrounds, skills, and abilities.

3. The *NGSA* design process can be used to unpack and identify the meaningful parts of PEs that will be suitable for classroom-based assessment. The process emphasizes using the meaningful parts to construct comprehensive sets of smaller performance statements that are called *learning performances*. Learning performances are intermediary performance targets for instruction and assessment that can signal whether students are moving along a productive path to proficiency with a PE or PE bundle.

4. Design blueprints are an organizational strategy for the principled development of tasks aligned to learning performances. Clearly specified design blueprints can bring coherence and consistency to the work of creating three-dimensional tasks. They serve as the ground plans for task designers, providing clear information for creating tasks and common reference points for checking work.

5. An integrated three-dimensional assessment task requires a corresponding integrated three-dimensional rubric. A well-designed rubric can illustrate what

three-dimensional science performance looks like and provide guidance to inform subsequent instruction that will help move students forward in building toward the PEs.

In Chapters 3–7, we delve deeper into the *NGSA* design process and provide step-by-step guidance for constructing well-aligned tasks that can be used in the classroom to follow and support students' progress in three-dimensional learning.

References

Harris, C. J. et al. 2019. Designing knowledge-in-use assessments to promote deeper learning. *Educational Measurement: Issues and Practice,* 38(2), 53–67.

Mislevy, R. J. & G. D. Haertel. 2006. Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, 25(4), 6–20.

Mislevy, R.J., L. S. Steinberg & R. G. Almond. 2003. On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3–62.

National Research Council (NRC). 2012. *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

National Research Council (NRC). 2014. *Developing Assessments for the Next Generation Science Standards*. Washington, DC: National Academies Press.

NGSS Lead States. 2013. *Next Generation Science Standards: For states, by states*. Washington, DC: National Academies Press.