Learning Progressions and Embedded Assessment

Derek C. Briggs

Erin Marie Furtak

University of Colorado Boulder

November 6, 2018

Pre-print from S. Brookhart & J. McMillan (Eds) *Classroom Assessment and Educational Measurement*, NCME Book Series.

Abstract

Learning progressions have great potential as an organizing framework for classroom instruction and assessment. However, successful implementation of this framework hinges upon developing a curriculum-embedded system of student assessment. In this chapter, an approach to meeting this challenge is illustrated in the context of a learning progression in science that crosses the disciplinary boundaries of physics, chemistry and biology in a high school setting. Four key ingredients of our approach include (1) mapping and aligning the scientific content of the learning progression to the curricula of the participating teachers, (2) making the case that assessment activities targeted to the learning progression can provide teachers with relevant insights about their students, (3) bringing teachers together to discuss student ideas that emerge from assessment activities, and (4) linking the assessments within and across the courses taught by participating teachers.

Introduction

It has been nearly two decades since the publication of the seminal National Research Council report, *Knowing What Students Know* (NRC, 2001), and during this time there has been rapidly increasing interest in the topic of learning progressions (Alonzo, 2011; Shepard, Penuel & Pellegrino, 2018a; Wilson, 2018). Learning progressions are empirically grounded and testable hypotheses about how students' understanding of core concepts within a subject domain grows and become more sophisticated over time with appropriate instruction (Corcoran, Mosher, & Rogat, 2009). As such, research on learning progressions represents one tangible response to a key recommendation from *Knowing What Student Know*; namely, that all assessment activities should be motivated by, or at least motivate reflection about, theories for how students learn in a given subject domain. The presence of a learning theory is important because once established it is more general and comprehensive than any single assessment event. A good theory helps teachers to pose the questions best suited to their students, so ideally, it is learning theory that eventually drives student assessment, and not the other way around (Shepard, Penuel, & Pellegrino, 2018b). Moreover, theories about how students learn can help teachers discern instructionally relevant insights from the answers that students give on assessment items. Because learning progressions are premised on testable hypotheses, the learning theories that they embody can and should be modified and refined over time, something that is especially important when learning is viewed as a sociocognitive or sociocultural phenomenon (Penuel & Shepard, 2016), because the theory that may best explain changes in student understanding in one situated context may not have the same explanatory power in another.

There are numerous challenges to the use of learning progressions as an organizing framework for classroom instruction and assessment (c.f., Alonzo & Gotwals, 2009). In this chapter, we focus on one challenge in particular: the challenge of developing a curriculum-embedded assessment system. Designing good student assessments is always challenging, but it is especially so in a learning progression context for at least two related reasons. The first reason is that almost by definition, the greatest utility of a learning progression is an orientation to teaching that focuses on growth over status. A good learning progression marks out one (or more) likely path(s) that students are expected to traverse as they become more sophisticated in their understanding of a core concept. It follows that for teachers to gain insights about the actual path that students take over the course of an instructional period, it is necessary to organize *multiple* assessment events along the way, and it can be challenging to find a way to ensure that each assessment is appropriately targeted and aligned to the different levels of the progression. The second reason assessment design in this context is challenging (closely related to the first reason) is that the middle to top ends of most learning progressions are typically characterized by expectations of cognitive complexity that go beyond the recall of isolated facts or the application of these facts as part of standard procedures. As a consequence, distinguishing between a student, or a group of students, at different locations on a continuum from novice to expert can require more complex tasks (e.g., constructed-response items, performance-based tasks) that are time-consuming to design, administer, and evaluate. Taken together, to the extent that an ideal use of learning progressions could involve lengthy assessment events on multiple occasions, it is little wonder that prior research on learning progressions tends to involve cross-sectional data collected at one point in time, rather than longitudinal data collected at multiple time points. This is unfortunate because it limits the benefits teachers are likely to get from using the learning progression as a tool for formative assessment[1], and because it provides for a fairly weak test of the theory underlying the learning progression.

We argue that one way for a learning progression framework to realize its full potential is through the development of curriculum-embedded assessments (hereafter we refer to these simply as embedded assessments). Embedded assessments serve dual purposes as part of a comprehensive assessment system. On the one hand, they are proximal to a teacher's curriculum and can be used to provide immediate feedback that facilitates student learning. On the other hand, they include scorable tasks that can be used to reliably monitor growth in student understanding over time, and to evaluate what students have learned at some given point in time. Ideally, there will be coherence between assessments used for both formative and summative purposes. In this chapter, we describe work from an ongoing project in which we were faced with the challenge of building a system of embedded assessment in support of a learning progression in science. The learning progression at the heart of this project pertains to the modeling of energy flows, a "big idea" in science that crosses disciplinary boundaries, and which we are presently implementing and evaluating as part of structured professional development activities with classroom teachers of physics, chemistry and biology in a high

---

[1] We define formative assessment following Bennett (2011) as both the processes and instruments that elicit what students know and are able to do for the purpose of informing subsequent classroom instruction.

school setting. Following the development of a learning progression for the modeling of energy flows, four of the key ingredients of our approach include (1) mapping and aligning the scientific content of the learning progression to both the content standards and the curricula of the participating teachers, (2) building a system of assessments targeted to the learning progression that can provide teachers with relevant insights about their students, (3) bringing teachers together to discuss student ideas that emerge from embedded assessments, and (4) linking the assessments within and across the courses taught by participating teachers in physics, chemistry and biology with a subset of common tasks.

Motivating Context

The context for our illustration comes from the first two years of a research project funded by the National Science Foundation (NSF). This project is itself situated within an ongoing a Research-Practice Partnership between a medium sized Colorado school district (total enrollment of about 40,000 students) and researchers at the University of Colorado Boulder. The partnership was formed around the need, from the school district's perspective, to support secondary science teachers in developing, using and interpreting student assessments for a variety of purposes, ranging from those that were low-stakes (involving primarily formative classroom use by teachers), to those that could be higher-stakes (involving summative uses to grade and compare students, or even as a basis for teacher evaluations). From our perspective as researchers, we saw the partnership as an opportunity to both study and challenge what, in the United States at least, has become a conventional teacher view about student assessment. Two aspects of this conventional view are especially salient to our work. The first is that assessment is something that is "done" to students and that it comes from an external source outside a teacher's control. The second is that the point of assessing students is to find out if they get it or they don't (Otero, 2006). We view learning progressions, and the approach to student assessment that they require, as a promising way to challenge this conventional view to the benefit of both teachers and their students.

The intervention at the heart of our NSF-funded project was to directly engage high school science teachers in a process of using learning progressions as a framework for iteratively designing, enacting, and reflecting upon student assessment. This engagement took place during regularly scheduled meetings in teachers' *professional learning communities* (PLCs; McLaughlin & Talbert, 2001; Laughran, Mulhall & Berry, 2004; Gröschner et al., 2014). Over a course of two years, our team took over facilitation of a subgroup of six "focus" PLCs, two in each of the disciplinary content areas of physics, chemistry and biology. Historically, although the activities within each PLC varied, a major emphasis had been placed on what the district referred to as "data cycling," which involved teachers administering assessments, collecting data and analyzing results in rapid cycles lasting 2-4 weeks. To create these assessments, teachers were expected to draw upon a variety of resources, including their curriculum materials, test item banks from textbook publishers, and released state test items. One challenge then, was to demonstrate to the district that it was still possible to make "data-based decisions" using assessments designed using a learning progression framework. A second challenge was to make the case to teachers that this framework could help them to teach and assess the content of the NGSS more effectively and efficiently. We suspect that our district's context would be familiar to most researchers who engage with teachers around issues of classroom assessment.

There were a two other aspects of national and local context that informed our work. First, in 2012, the National Research Council released the report, *A Framework for K-12 Science Education*, and a year later a collaboration between the National Science Teachers Association and the American Association for the Advancement of Science, facilitated by the organization Achieve Inc, led to the release of *the Next Generation of Science Standards (NGSS)*. A defining feature of the NGSS relative to previous approaches such as the *National Science Education Standard* or the *Benchmarks for Scientific Literacy* is the view that the teaching and learning of science should be conceptualized as an interwoven three-dimensional enterprise in which students generate, or are presented with, a real-world phenomenon and then use some combination of disciplinary core ideas (DCIs), scientific and engineering practices (SEPs) and crosscutting concepts (CCs) to make sense of it. At the start of our project, Colorado had just adopted the NGSS, and the district was just beginning to grapple with the implications of this for its curriculum, instruction and assessment structures and activities.

A second salient aspect of our context was that our partner district uses a "Physics-First" curricular sequence, in which Physics is taught to 9th grade students, Chemistry to 10th grade students, and Biology to 11th grade students. The Physics-First structure to the science curriculum, which is contrasted with the traditional sequence of Biology-Physics-Chemistry, is intended to help students establish a foundation in core physical concepts such as energy and force, and then use these concepts to facilitate subsequent learning about chemical reactions and molecules, all of which are the foundations of modern biology (Popkin, 2009). An implicit premise here is the notion that one should revisit student understandings of certain core concepts even as they cross disciplinary boundaries, and this echoes a premise of the NGSS. In this sense adopting the NGSS gave the district an excuse to make their implicit premise an explicit hypothesis. The desire to find a scientific concept that would be central to all three discipline-specific courses was a key reason that we chose to develop a learning progression around energy. In the next section, we explain how we went about this development.

Development of a Learning Progression for the Modeling of Energy Flows

Prior Research on Energy as a Learning Progression

The law of energy conservation is deceivingly simple, requiring that initial energy is equal to final energy in any isolated system. The implications of this law can prove to be tremendously useful to scientific investigations of both natural and human-generated phenomena, since it introduces a fixed constraint. Whenever the energy of a system increases, we know that the additional energy had to have come from some other source outside the system. Whenever the energy in a system decreases, we know that the lost energy has to have gone to some other system. At the same time, as a scientific concept, the term “energy” is vague and abstract, as much a label that gets attached to a process as it is a specific thing that is tangible and observable at some fixed moment in time. The physicist Richard Feynman famously remarked that “It is important to realize that in physics today, we have no knowledge what energy is” (Feynman, Leighton & Sands, 1989, pp. 4–3).

To really understand what energy is and how it comes to be requires, among other things, an understanding of the particulate nature of matter, an understanding of force and the relationship between force and potential energy, an understanding of electric and magnetic fields, chemical reactions among molecules, and an understanding of the concept of a system (Chen et al., 2014; Nordine, 2016). Because of this, when children begin to receive formal instruction about energy, they come to understand it almost exclusively by where it comes from and what it does, leaving the lingering question of what it *is* to sit in a black box, to be revisited at some later date. At the same time, children enter school settings having already developed intuitive understandings about where energy comes from (e.g., the sun, a battery, food) and what it does (e.g., makes things move and grow). As Nordine (2016) points out, students are likely to perceive some cognitive dissonance when they first encounter the law of energy conservation, because it is likely to conflict with a previously established mental model that conceives of energy as a thing that gets acquired, used up, and reacquired. Because this mental model may well predict many observable phenomena with reasonable accuracy, teachers are faced with the challenge of helping students integrate this flawed (but useful) model with a more complex and seemingly counterintuitive account.

One approach to meeting this challenge is to conceptualize the understanding of energy as a learning progression, defined with respect to some combination of four interrelated big ideas (Herrmann-Abell & DeBoer, 2017; Neumann, Viering, Boone, & Fischer, 2013; Nordine, 2016).

1. Energy comes in different forms and manifestations

2. Energy can be transformed from one form to another, or transferred from one object to another.
3. Energy is conserved. It is never destroyed—only transformed or transferred.
4. Energy is degraded or dissipated in all macroscopic processes.

In one of the earliest studies to adopt this perspective and examine it empirically, Liu & McKeough (2005) conducted a secondary analysis of TIMMS multiple-choice items written to assess students' understanding of energy. Liu & McKeough coded these energy items to correspond to one of the four big ideas listed above, and subsequently found that comparisons among the items with respect to their difficulty for students to answer them correctly supported the hypothesis that the ideas had some hierarchical structure: items related to the identification of energy forms tended to be easier to solve than items related to energy transformation and transfer, which in turn tended to be easier to solve than items related to energy conservation and dissipation. The core findings from this study were subsequently replicated in follow-up studies involving performance assessments (Liu & Collard, 2005) and constructed-response items (Lee & Liu, 2010).

Neumann et al. (2013) built upon these results to develop a more elaborated learning progression that they sought to validate as part of a prospective study. A partial order was hypothesized to exist across the four big ideas about energy, along with a hierarchical order *within* each of the four big ideas. The "within big idea" order was to be related to the degree of scaffolding (in the form of hints) that a student would need to correctly solve a selected response item. The results from this study showed mixed support for the hypothesized partial order between the big ideas. While items associated with the identification of energy forms tended to be easiest for students to solve, and items associated with energy conservation tended to be hardest, there was no significant difference between items associated with energy transfer and transformation and those associated with energy dissipation. Beyond this, Neumann et al. found no evidence of hierarchies within the big idea items associated with scaffolding (though this may have been due to acknowledged confounds in their item design and administration).

The results from these early studies, and others that followed by Herrmann-Abell & DeBoer (2017) and Park & Liu (2016) were, perhaps not surprisingly, inconclusive, but they can be characterized as ground-breaking in the sense that they represented early, exploratory attempts to connect the design of student assessments to Piagetian or neo-Piagetian theories of how students become more sophisticated in their understanding of energy. To a great extent, these studies raised more questions than they answered about both the nature of a learning progression that could (or should) be posited for energy, the nature of the assessments that should be used to test the learning progression, and the nature of the evidence one would expect to find in order to validate—or invalidate—the learning progression. An important limitation of these early efforts is that they do not situate the assessment of a student's location on the learning progression within any particular context for curriculum and instruction. Neither was there any theory of action for how the information from these assessments could be used by teachers for formative or summative purposes. Instead, the assessment efforts were exclusively focused on high-level theory validation.

Energy in the Next Generation of Science Standards

Energy is the only concept in the NGSS that is named as both a Disciplinary Core Idea (DCI) and a Cross-Cutting Concept (CCC). That is, on the one hand energy is one of four major DCIs situated within the physical sciences, and the NGSS sketches out a rough learning progression for American students from Kindergarten through high school in terms of four smaller grain ideas about energy, depicted in the rows of Table 1 labelled PS3.A, PS3.B, PS3.C and PS3.D. On the other hand, the NGSS casts energy (together with matter) as one of seven CCCs that can play a role in understanding phenomena related to DCIs across not only physical science, but also across earth and space science and life science. The NGSS's suggested progression of energy and matter across grade bands is depicted in the last row of Table 1.

Insert Table 1 about here

The NGSS were written to support a laudable vision for K-12 science instruction, one that calls for students to actively engage in the practices of scientists to answer puzzling questions about the world around them. Rather than promoting the view that science is constituted by a set of facts and procedures that need to be memorized, the NGSS was designed to promote the goal of students graduating from high school with both a curiosity about the world around them, and the ability to use a small set of core ideas and practices about science that they have begun to master to investigate and understand novel phenomena. Few would argue that this is not a worthwhile ambition.

At the same time, the "three-dimensional" structure of the NGSS present a significant challenge to student assessment, and a little bit of arithmetic can illustrate the issue at hand. If making sense of a phenomenon always involves some combination of one or more interrelated DCIs, SEPs and CCCs, then prospectively, if we simply count up all the unique DCIs at the smallest available grain size (44) and cross them by unique SEPs (8) and CCCs (7), there are a total 2,464 combinations that might, in theory, be brought to the table to characterize the means by which a student makes sense of any given phenomenon. The proper construct for assessment, and the grain size of the construct thus becomes an open question. Is it the ability of a student to understand and explain a specific phenomenon? Is it some underlying DCI abstracted from a motivating phenomenon but specific to a subset of SEPs or CCCs? Is it some underlying DCI generalized across all SEPs or CCCs? In an attempt to mitigate this issue, the NGSS specifies *performance expectations*, organized by grade band and discipline, that represent a purposeful crossing of some subset of DCIs, SEPs and CCCs. Still, the number of unique performance expectations remain daunting from an assessment perspective. In grades K-2 there are 33 unique performance expectations, in grades 3-5 there are 45, in middle school there are 59, and in high school, there are 72. And since each performance expectation comes with a detailed set of evidence standards that stipulate what a student should know and be able to do to demonstrate mastery, the design and administration of an assessment for just one performance expectation is likely to be a time-intensive activity.

To the extent that teachers wish to assess their students for the purpose of gaining insights about their learning within the course of a semester or academic school year, the NGSS, if viewed in isolation, are unlikely to be sufficient. As argued at the outset of this chapter, when there is a desire to use student assessments to gain insights about learning, it helps to have a learning theory in mind. Although the NGSS does provide some learning progression markers for each DCI, SEP and CCC dimension across grade bands, the dimensions have not been integrated and there is no hypothesis available for what a progression might look like *within* a grade or course. In some sense then, each NGSS performance expectation could be cast as the "upper anchor" of a within or across grade learning progression, with the trajectory that leads to this upper anchor left unspecified. For example, in high school there are 14 unique performance expectations that include energy and matter as a CCC, 9 unique performance expectations that include energy as a DCI, and just one that includes energy as *both* a CCC and a DCI. Each of these 24 performance expectations could, in principle, be the basis for a learning progression related to the understanding of energy within and across the high school grades.

A Learning Progression for Modeling Energy Flows in High School

The learning progression (LP) we developed builds upon consensus positions (e.g., NGSS) and the extant research literature in science education (Neumann et al, 2013; Hermann-Abell, 2017) but also breaks new ground. Our proposed LP maintains a link to the big ideas about energy that have been the basis for previous large-scale investigations. That is, we posit that student conceptions of energy are some function of these big ideas, and that some of the big ideas are easier to grasp and interrelate than others. For example, students are likely to be able to identify different forms of energy that they encounter when presented with canonical cycles in the natural world (i.e., the rock cycle, the water cycle, the carbon cycle,) or with the motion of an object or objects in a closed system (i.e., the swinging of a pendulum). Following Herrmann-Abell & DeBoer (2017) we distinguish between five main types of energy forms: kinetic energy, gravitational potential energy, thermal energy, elastic potential energy and chemical energy. Prior to high school, we can expect that students have previously come into contact with these different labels and hence can recognize that energy comes in multiple forms. Forms of energy go hand in hand with the transformation and transfer of energy. For example, when a student sees a pendulum swinging, the reason a student identifies different forms of energy is the recognition that energy is changing as the pendulum swings. Being able to *connect* ideas about energy forms, transformation and transfer and the law of energy conservation represents an important conceptual demarcation, one that hinges upon the ability to recognize and distinguish between a *system* and its *surroundings*. Another important demarcation is an understanding of the *mechanism* through which energy is transferred and how this can lead to dissipation or degradation. These would include transfer by conduction, convection, radiation, forces, electrically or by sound.

Distinctions among levels of our LP depend upon the ability of a student to *develop and use a model* that interrelates the big ideas about energy for the purpose of explaining and predicting a *phenomenon*, where we define a phenomenon as an observable event or state that can be explained or predicted through scientific investigation. In including the scientific and engineering practice of developing and using models in our LP for energy, we draw upon a recent revision by Pierson, Clark, & Sherard (2017) to a well-known LP for modeling in science first developed by Schwarz, Reiser, Archer, Kenyon, & Fortus (2012). This modeling LP was defined with respect to five different categories, with each category further delineated with respect to a hierarchy of discrete levels. We pull from the "mechanistic-generative" category which distinguishes between models that are

- descriptive only (lowest level),
- illustrate patterns,
- represent a mechanism to explain a predicted phenomenon, and
- predict and generate questions about possible new phenomena (highest level).

Insert Figure 1 here

The center column in Figure 1 depicts the five levels of our Modeling Energy Flows LP[2], with the lowest entry level at the bottom, and the highest level at the top. Each level represents differences in the sophistication of a model a student could develop and/or use to make sense of a phenomenon of interest in terms of the flows of energy into, within and out of the system. Implicit is the scenario in which a student is presented with a phenomenon in the physical world that can be linked to a specific DCI, but the generic progression as specified here is, at this point, agnostic about the specific nature of the phenomenon and its associated DCI that would be required to illustrate the mechanism of energy transfer or transformation. What *is* assumed is that the student is receiving instruction and practice in using what Lacy, Tobin, Wiser & Crissman (2014) refer to as an "Energy Lens" when thinking about the phenomenon at

[2] The modeling energy flows learning progression was the product of the collaborative iterations of our research team, and in addition to the lead authors, involved contributions from Jason Buell, Kate Henson, Rajendra Chattergoon, Kelsey Tayne, Amy Burkhardt, Caitlin Fine and Borbala Mahr. A more detailed report on its development can be found at https://www.colorado.edu/cadre/report.

hand. Taking an Energy Lens means, that before developing and/or using a model to make sense of a phenomenon, students get used to asking themselves the following questions:

- What is the system of interest?
- What observable or measurable changes or other interesting behaviors are taking place?
- Where in the system are energy changes occurring?
- Where does the energy come from?
- Where does the energy go to?
- What is the evidence for our answers?

And to these questions we might add, for high school students, what are the limitations of the evidence we have available? Though Lacy et al. (2014) frame the questions that accompany an Energy Lens as an activity for students in elementary school grades, the same habits of mind surely apply to developing a good model of an energy flow in high school and beyond.

Returning to the learning progression in Figure 1, we focus on the critical distinctions between each level. At level 1, a student can develop a model to answer some of the questions above, but will generally only be able to do so by showing or identifying physical components of a phenomenon, specific energy forms, or transformations motivated by a change they have observed. The key change at level 2 of the progression is the ability to identify and distinguish the appropriate system and surrounding, and to show that there is a relationship between the increase in one form of energy and the decrease in another form. At this level students can use their model to show patterns that are suggestive of energy sources and destinations, even if they remain hazy about the evidence that connects one to the other. At level 3 a student can develop a model that shows the total energy of the system is conserved either by accounting for all transfers or transformations within the system or by dissipation out of the system. At level 4 a student can develop a model to illustrate a mechanism that can explain and predict the phenomena in question in terms of a transfer of energy. It is at this level that the student is able to use the law of energy conservation as a constraint on the system, and can explain the role of energy in a given phenomenon through an interaction between all the big ideas of energy, and describe limitations of the model. Finally at level 5, a student is able to generalize the model to other phenomena beyond that which spurred the need for a model, and to recognize limitations in the model with this novel purpose in mind.

With respect to the structure of the NGSS, the LP above weaves together many of the different dimensions that are used to characterize the core ideas and concepts of science. It clearly combines the idea of energy and matter as a CCC with designing and using models as a SEP. But it also incorporates aspects of others CCCs and SEPs. To be at levels 3 through 5 of the LP will typically require some students to rely on practices related to modeling, practices that include analyzing and interpreting data, using mathematical and computational thinking, constructing explanations and engaging in argument from evidence. Similarly, progress up the levels will typically invoke other crosscutting concepts, most notably patterns, cause and effect, scale, systems, and stability and change. The LP is not meant to be applied to *all* DCIs, only those that have been flagged by the NGSS as belonging within a performance expectation that includes some combination of energy & matter as a CCC, modeling as an SEP, and any of the four energy-specific DCIs. In the next section we show how this is used to both constrain the content domain to which the LP can be applied while also mapping to units of our participating school district's curriculum.

We conclude this section by pointing out that the Modeling of Energy Flows LP is intended to be used to support the development of assessments for a mixture of formative and summative purposes both within a particular grade and course in high school, and across courses in high school. A general hypothesis of this LP is that when it comes to modeling energy flows, the order of the five levels of sophistication remain the same irrespective of the scientific discipline of a high school course. This does not, however, imply an assumption that the progress across levels will be linear, or that it does not depend upon the course sequence. A linear progression would imply that once a student has demonstrated an ability to model the energy flow for some sample of phenomena by the end of a grade 9 course at a level 4, that they will be able to do so at a level 4 or 5 in their grade 10 course. This would be possible, but seems unlikely. More plausibly, practice with modeling energy flows in one disciplinary context should make it easier to do so in the next disciplinary context. When students follow a Physics-First Curriculum, one can track the implied longitudinal progression associated with modeling energy flows of phenomenon from a physical science perspective (courses in physics and chemistry), followed by life science perspective (course in biology). This is probably the ideal curricular sequence for the modeling of energy learning progression, because physics and chemistry give students the tools to model the mechanisms behind energy transfers at the particulate level, and this can then by gainfully applied to biological phenomena.

Mapping the Learning Progression to Standards and Curricular Units of Instruction

To be relevant and useful as basis for either formative or summative assessment, teachers need to be able to see that an LP is not only aligned with, but can help to bolster, the instructional activities that are part of planned curricular units. Although teachers in most school districts are often given considerable flexibility with respect to the structure of these units and their timing, they are expected to demonstrate that the units have been linked to the district's content standards for science. In this context, those standards are the performance expectations of the NGSS. Our goals was to show teachers the connections between the Modeling Energy Flows LP and the NGSS performance expectations, and to use this link to show teachers how the LP (and its associated assessment tasks) can be used to make connections across units that might not have been visible otherwise.

We established a manageable domain for this learning progression by filtering the performance expectations for grades 9-12 to include, with one exception, only those that include the SEP of modeling and either the CCC of energy and matter or one of the four DCIs associated with energy[3]. This resulted in a total of 11 unique performance expectations that could, in principle be matched to the disciplinary focus of high school courses of physics (4), chemistry (3) and biology (4). We chose two PEs per discipline as basis for focal curricular units and associated student assessments, and these are listed in Table 2. Each of these PEs can be readily associated with the modeling of energy flows for a given phenomenon, but they can

[3] The one exception was the performance expectation for Energy (PS3-4) which is linked to the SEP "Planning and Carrying Out Investigations". It is nonetheless clearly aligned with our modeling energy flow LP given its CC of systems and systems models its energy-specific DCIs.

differ with respect to the DCIs a student would encounter in coming to a sophisticated understanding of the mechanism behind energy transfer.

Insert Table 2 here

Each of the performance expectations in Table 2 is related to the Modeling Energy Flows LP in the following way: mastery of any of the performance expectations can always be associated with level 4 of the LP. More specifically, every performance expectation can be fleshed out into a DCI-specific LP with levels that characterize a student's most likely pathway to an understanding of energy flows that demonstrates an integration of the four big ideas about energy. Our conjecture is that, with respect to the ability of students to model energy flows, these levels would track with the ones specified in our general LP (see Figure 1). A great advantage of this perspective, if it can be validated, is that it lends greater coherence across NGSS performance expectations and associated curricular units by emphasizing the way that energy is a concept that cuts across them, and how models can be used as a tool for sense-making and explanation.

Designing a System of Assessments

As the Modeling Energy Flows LP has a three dimensional structure in keeping with the ethos of the NGSS, building a system of assessments aligned to it is challenging. In 2014, the

National Research Council released the report *Developing Assessments for the Next Generation of Science Standards*, and one of its principal conclusions underscores this challenge:

> Measuring the learning described in the NGSS will require assessments that are significantly different from those in current use. Specifically, the tasks designed to assess performance expectations in the NGSS will need to have the following characteristics:
>
> - Include multiple components that reflect the connected use of different scientific practices in the context of interconnected disciplinary ideas and cross-cutting concepts;
> - Address 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
> - Include an interpretive system for evaluating a range of student products that are specific enough to be useful for helping teachers understand the range of student responses and provide tools for helping teachers decide on next steps in instruction. (NRC, 2014, p. 3)

The necessary features of assessment tasks described in *Developing Assessments for the Next Generation* make clear the desirability of building a system of embedded assessment opportunities, wherein assessments are included at multiple junctures within a given curricular unit as part of planned classroom activities that promote learning. The assessment tasks themselves would be expected to vary with respect to their format, their duration, and their use. To this end, we envisioned and developed three types of assessment tasks: performance-based tasks and labs, phenomenon-based item clusters, and conceptually oriented multiple-choice items.

Performance-based Tasks and Labs

Performance-based tasks are to a great extent most closely aligned with the vision for science assessment sketched out by the National Research Council. A common feature of such a task is that it is always premised on a motivating question or scenario that presents students with a real-world phenomenon, and then poses questions about the phenomenon that ideally should lead students to invoke the three dimensions of disciplinary core ideas, cross-cutting concepts, and scientific practices in coming up with answers to these questions. Figure 2 provides an example of a performance-based task we developed to elicit evidence about students' ability to model energy flows across multiple disciplinary contexts. The task is premised on a scenario in which students are asked "How can corn provide energy to power a bus?" They are informed that corn is grown for many purposes, not just for food; it can also be made into fuel. Next they are asked to develop a diagrammatic model that traces energy as it flows from the sun to the corn, is processed into ethanol, and then flows from ethanol to the movement of the bus. Finally, they are asked to use the model to explain how energy flows through these systems, including all energy inputs, outputs, transfer and transformations (see Furtak, Binder & Henson, 2018, for more details on the development of this task).

The task has some notable characteristics. It includes three different stages that could correspond to distinct system models, and in each one a flow of energy could be depicted with respect to different transfers and transformations. At the same time, the mechanisms by which these energy transformations and transfers take place would potentially require a student to invoke DCIs specific to physics, chemistry and biology. One could argue that the ability to fully complete a task such as this would represent an ideal end goal for a student after three years of instruction in physics, chemistry and biology, provided that the instruction was able to consistently emphasize the role of energy as a cross-cutting concept and the role of modeling as a practice that can be used to depict and understand energy flows. In this kind of idealized scenario, we would still expect to see considerable variability in the sophistication of student responses, and these responses would be scorable with respect to the levels of the modeling energy flows LP previously depicted in Figure 1.

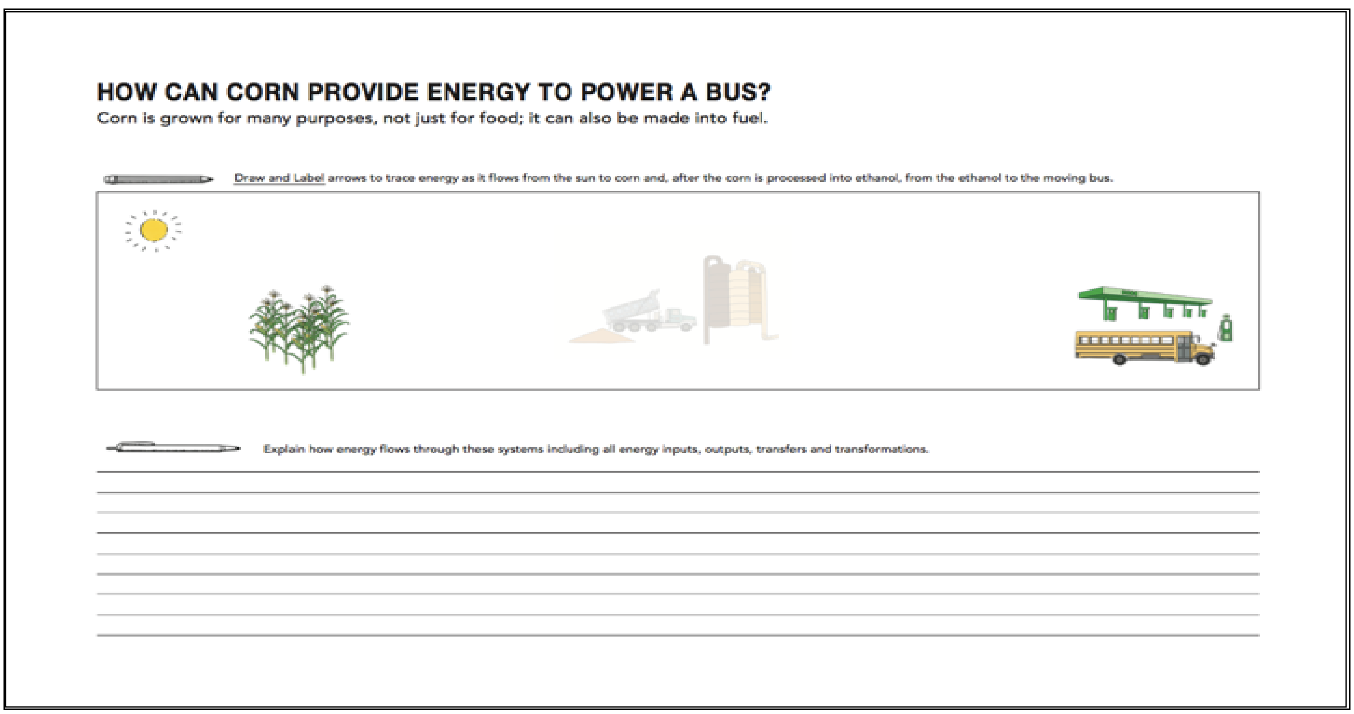


Figure 2. Biofuels Performance-Based Task

A problem that we soon discovered when piloting this task with high school students in physics and biology courses as a standalone assessment was that very few students were able to engage with it in the way that we had intended. One reason for this was that because it provides minimal scaffolding or points of entry for students who are just developing their understandings about energy flows and how to describe them with a model. As a classroom assessment, a task such as this works best in the context of an activity that could incorporated into a project or lab-based investigation that could span multiple days of class time.

An example of this that can serve as a template is provided by Eisenkraft (2014) with the "cheese puff lab." The motivation for the activity is showing students the food label showing the nutrition facts about a bag of cheese puffs and then asking them to speculate about how the number of calories associated with a single cheese puff is determined. How do calories provide a representation of energy content? From there, students participate in a lab in which they are asked to attach a cheese puff to a small apparatus that sits below a container of water, light it on fire, and then measure the change in water temperature before and after the cheese buff has finished burning. Eisenkraft's cheese puff lab is a great example of a phenomenon that could be readily connected to the modeling energy flows LP, because it invokes multiple DCIs that would be relevant to whether students were taking a course in physics, chemistry or biology, because it provides an opportunity for students to practice model development, and because the cross-cutting concept of energy provides the critical framing for answering the motivating question. In these sorts of lab settings, assessment is still at the center of the activity in that it is remains important for teachers to elicit and attend to the differences in student ideas about the energy flows both before and after the central lab activities. But the assessment is embedded within the larger lab activity, which could (and probably should) span multiple days. The responses students give to targeted questions about the data they have collected can become the basis for *student-work focus sessions* (described later).

Phenomenon-Based Item Clusters

Phenomenon-Based Item Clusters (PBICs) are similar to the performance-based tasks described above in that they are also associated with a specific motivating question related to some observable phenomenon, but they are broken into a sequence of items intended to provide students with scaffolding so they are better equipped to engage with the phenomenon even if they only have a limited understanding of the underlying energy concepts. In a sense they are intended to mimic an interaction with a teacher who is able to help the students see and make connections between the phenomenon and disciplinary core ideas related to energy flows. These ideas can be brought to the fore by helping students engage the task through the development and use of a model of energy flow, so item clusters are intended to provide students with the information and prompts necessary to set this in motion.

All PBICs are based on a common design template[4] that can be used to create an assessment with for any of the six PE-specific versions of the modeling energy LP shown previously in Table 2. Figures 3 and 4 provide two examples of a motivating phenomenon, one that invokes a DCI specific to biology (LS1.C Organization for Matter and Energy Flow in

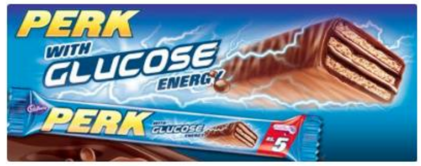
[4] The development of the PBIC template was spearheaded by Rajendra Chattergoon and Jason Buel. For details, see https://www.colorado.edu/cadre/report

Organisms), and one that invokes a DCI specific to physics (PS3.A Definitions of Energy)[5]. Each figure also includes the next two items that follow the opening scenario as students are asked to identify and distinguish between the system and surround. Not depicted in these figures are the next four items in the PBIC in which students are asked to identify the forms of energy in the phenomenon and to characterize the patterns that suggest energy is being transferred or transformed. The opening six items of each PBIC, which ask for selected responses from the student, probe the extent to which the student understands the mechanism of energy transfer underlying this scenario, and can use the law of the conservation of energy as a constraint on the system. These are items that help make distinctions primarily between levels 1 and 2 of the LP. In addition, these items provide students with the vocabulary they will need to develop of a model of the energy flow. The culmination of each PBIC are three constructed-response items that ask the student to (1) draw a model that shows the phenomenon (e.g., Draw a model that shows how an energy bar provides a runner with energy to move) and (2) describe in words how energy is being transferred or transformed (e.g., Use your model to describe in words how the energy bar provides a runner with energy to move.) and (3) characterize the limitations of the model as a way of explaining the phenomenon. These are the items that help to distinguish between levels 2 through 4 of the LP.

[5] We thank Knut Neumann and Jeffrey Nordine for their permission to use this “Stuntman Felix” scenario which they developed as part of a different research project.

How does an energy bar help a runner move?

Olivia runs on her high school's track and field team. Her coach suggests that she eat a glucose energy bar before her last race to help her when she is tired. Olivia asks her coach how glucose will help her run. To help explain the science to her, Olivia's coach has asked you for help.



*In biology, a **system** is the part of the world that is under investigation and the **surroundings** are anything outside of the investigation. Circle the best answer.*

1. Which of the following is part of **the system** where the cellular respiration reaction occurs?

   (a) air outside Olivia's body only (b) glucose only (c) Olivia's cells and O_2 in the cells only (d) glucose, Olivia's cells, and O_2 in the cells

2. Which of the following is part of **the surroundings**?

   (a) air outside Olivia's body only (b) glucose only (c) Olivia's cells and O_2 in the cells only (d) glucose, Olivia's cells, and O_2 in the cells

Figure 3. Scenario and First Two items from a Phenomenon-Based Item Cluster in Biology

Where does a skydiver's energy to break the sound barrier come from?

Stuntman Felix Baumgartner holds the world record for the fastest speed achieved by a human without an engine. He broke the sound barrier and reached a top speed of 377 m/s. He used a balloon to fly approximately 39,000 meters above Earth and, wearing a special suit, jumped down.

*In physics, a **system** is the part of the world that is under investigation and the **surroundings** are anything outside of the investigation. Circle the best answer.*

1. Which of the following is included in **the system** where Felix gets energy to fall?

   (a) the air around Felix's body only (b) Felix's body only (c) the Earth only (d) Felix's body and the Earth

2. Which of the following is part of **the surroundings**?

   (a) the air around Felix's body only (b) Felix's body only (c) the Earth only (d) Felix's body and the Earth

Figure 4. Scenario of Two items from a Phenomenon-Based Item Cluster in Physics

Conceptually Oriented Multiple-Choice Items

The last type of items that comprise our assessment system are multiple-choice (MC) items that focus on students' conceptual knowledge. Here we generally pull from pre-existing items with a focus on DCIs in the physical sciences. Many of these are described in the published studies by Herrmann-Abel & DeBoer (2017), Neumann et al. (2013) and Park & Liu (2016). Some of these items are publicly available from a website maintained by the *American Association for the Advancement of Science (AAAS)*[6]*.* We also use or adapt multiple-choice items for Chemistry from Jim Minstrell's Diagnoser assessment system (Thissen-Roe, Hunt & Minstrell, 2004). On the one hand, these items are more limited in the depth of information they can elicit about the ability of students to develop and use models to describe and explain energy flows in the three-dimensional manner envisioned by the NGSS. However, they can provide very relevant information about students' understandings of DCIs and sometimes certain SEPs and CCCs. In addition, in some cases the "distractors" (incorrect answer options) have been written to reflect common student misconceptions, so there may be more diagnostic information that can be gleaned beyond whether the student got the item correct or not (c.f., Briggs, Alonzo, Schwab & Wilson, 2006). Finally, they are easy to score, and with respect to the AAAS items, there is normative information available to compare the frequency distribution of students in a given classroom to a national sample of students in at least the same age range. An example of a conceptually oriented MC item, taken from the Diagnoser assessment system, is depicted in Figure 5.

[6] See http://assessment.aaas.org/topics/1/EG#/0

Jody left a half-filled glass of sweet (sugar added) tea with ice on her dresser for 2 days. The ice has melted and the tea is now at room temperature. If you could look inside the glass and see the molecules of sugar, tea and water, what would you see?

A. The sugar, tea, and water molecules are motionless. There is no movement or change at this point.
B. The sugar, tea and water molecules are in constant, random motion, even though the ice has entirely melted.
C. The water, tea and sugar molecules are reacting with each other, and will eventually form a new substance.

Figure 5. Example of a Conceptually Oriented Multiple Choice Item

Creating Assessment Events for Different Uses

The three assessment formats described above—performance-based tasks and labs, PBICs, and conceptually oriented MC items—are the raw ingredients that comprise a system of embedded assessments. In this system, different assessment events could be used in support of different purposes. The NRC panel responsible for the report *Developing Assessments for the Next Generation of Science Standards* makes the distinction between a *classroom assessment* and an *assessment for monitoring*. A classroom assessment is one that is selected by teachers and typically given to students at the culmination of a curriculum activity or unit. Two defining features are its timing (within or immediately following related instructional topics) and who controls it (teachers). In contrast to a classroom assessment, an assessment for monitoring is one that has typically not been developed by a teacher who is being asked to administer it, and is less likely to be as closely related to the curriculum and instruction that immediately preceded its administration. Two defining features of an assessment for monitoring are its standardization and reliability. Although both classroom assessments and assessments for monitoring *can* be used for formative and summative purposes, on the balance classroom assessments are better suited for formative use, and assessments for monitoring are better suited for summative use. In our ideal vision of an embedded assessment system, a learning progression provides a framework that promotes coherence between the two different types of assessment events. That is, whether an assessment is given for formative purposes with the timing and content at the local discretion of a classroom teacher, or given for summative purposes with the timing and content at the discretion of a school district or state, the assessment should be written to provide insights about student conceptions relative to the theory of learning embodied by the learning progression. To the extent that performance-based tasks and labs, PBICs, and conceptually oriented MC items have all been designed to elicit this information, any one of these assessment formats, or a mixture of them, could be used for either classroom assessment, or assessment for monitoring. However, as we discuss later, the evidence needed to validate an assessment created for these different uses is likely to differ.

Professional Development and Teacher Ownership

Formative Assessment Design Cycle

The LP, its connection to curricular units, and a system of assessment tasks are the key ingredients that support working collaboratively with teachers in their PLCs to enact the *Formative Assessment Design Cycle* (Furtak & Heredia, 2014; Furtak, Morrison & Kroog, 2014).

The cycle, illustrated in Figure 6, is intended as a sense-making space for teachers to iteratively work with the LP as they learn to design, enact, and use information from the LP to inform their instruction. The cycle begins with teachers *Setting Goals* and *Exploring Student Thinking*. In this initial phase, teachers use the LP as a model for how student learning can unfold in a given domain of interest. Next, teachers *Design* and *Revise Formative Assessment Tasks* using the LP to create prompts that specifically target particular levels of understanding. Next teachers *Collect Data* in their own classrooms, enacting the tasks as common formative assessments (Ainsworth & Viegut, 2009), and collect evidence of what students know and are able to do relative to the LP to later discuss in their PLC. The cycle concludes when the teachers *Reflect* on their classroom enactment and *Make Inferences* about what students know and are able to do. In this crucial final step, teachers' interpretations are guided by the LP as they categorize student responses before identifying the types of instructional feedback that will be most useful to help each cluster of students move forward.

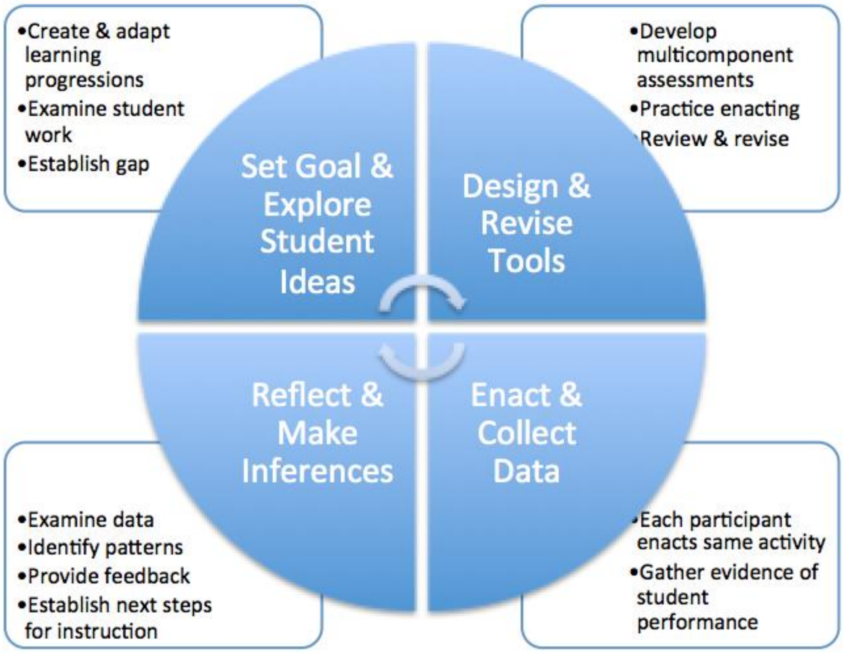


Figure 6. Stages of the Formative Assessment Design Cycle

In our research project we tailored this design cycle to a LP framework in the following ways. First, teachers were not expected to create a learning progression "from scratch." Instead, our starting point for the cycle was the general modeling energy flows learning progression developed by our research team, where (as described above) this development was informed by the research literature on energy in science education and the framework for science education established by the NGSS. In this case, the LP development went through several revisions informed by the results from piloting performance-based tasks, PBICs, and conceptually-oriented multiple choice items to teachers' students. To provide teachers an opportunity for ownership, we collaborated with them to develop a list of PE-specific indicators that help them easily identify the specific ideas that they will be expecting to see at each level of the learning progression for a given PE. Second, teachers are also provided with templates and examples of performance-based tasks and PBICs that they could use directly with their students as part of energy-related curricular units, or that they could use as basis for writing new tasks and items. The idea here was that while we wanted to empower teachers to write their own assessment tasks, they needed to be given a starting point and some guidelines for the principles that should inform these tasks. We also provided checklists derived from prior research on effective formative assessment to help teachers learn about scaffolds that can help students make their reasoning explicit to their teachers (e.g. Kang, Thompson, & Windschitl, 2014). One overarching principle is that assessment tasks should be chosen deliberately such that they can be used to elicit differences about the ways students model energy flows in their disciplinary context, and that this information should be useful in giving teachers ideas about what to do next. With this in mind, we have teachers engage in *student-work focus sessions* as part of the "collect data" and "reflect and make inferences" stages of the formative assessment design cycle.

Student-Work Focus Sessions

During student-work focus sessions[7], teachers meet together to discuss student responses to common assessment tasks with the goals of (a) making visible the qualitatively different ways that students make sense of energy as a cross-cutting scientific concept, (b) seeing the connections between the assessment tasks and the levels of the learning

[7] We have developed a guidebook for student focus sessions, which is available on the website for the Center for Assessment, Design, Research and Evaluation (https://www.colorado.edu/cadre/learning-progressions-project)

progressions, and (c) suggesting revisions that improving the assessment tasks and learning progression. The inputs for these sessions are a small number of student responses to assessment tasks that have been written to align to the LP. The responses are specifically selected by the organizer of the session to characterize the variability in the ways that students answer the questions that have been posed to them. There are two phases to a student-oriented focus session.

- In the first phase, all participating teachers make explicit connections between the scoring of the tasks, the student conceptions each task is expected to elicit, and how this relates to the levels of the LP. Next, teachers are asked to each score the same set of student responses where scoring requires the teacher to make a judgment about the sophistication of a student's ideas about energy flows and how they can be modeled. They then discuss any differences in their scores for the same student, come to a consensus score, and discuss ideas to modify the task and minimize score discrepancies in the future.
- In phase two, participants examine the consensus scores and student work to generate a better sense for the strengths and weaknesses in individual students as well as the groups of students they may represent. They then discuss possible next steps for instruction

A key to the success of the Formative Assessment Design Cycle is that the cycle needs to fit within the timeframe of a curricular unit emphasizing a known NGSS performance expectation. The challenge of coordinating this with teachers across different courses and different schools was considerable. To see if this could work as a proof of concept in our project, we limited ourselves in each discipline to one curricular unit related to energy in the fall/winter and another in the winter/spring demarcations of the academic calendar.

The Question of Validity

In line with recommendations from the National Research Council (NRC, 2014), a comprehensive and coherent assessment system should be able to support both formative and summative assessment purposes. But these represent two different use cases, and although some of the evidence needed to validate an assessment assembled for each use may overlap, much of it will be distinct, requiring the conduct of uniquely constituted studies. As a brief example of validity evidence that should overlap, the connection to a common learning progression implies that the information about students being elicited for a formative purpose should not conflict or be inconsistent with that which is elicited for a summative purpose. This requires evidence that the content and cognitive complexity of items in an assessment for monitoring learning (e.g., a district or state-administered interim assessment) is aligned with what is found in the items used for classroom assessments.

As an example of validity evidence specific to formative classroom use, consider the use case of a teacher including one of the performance-based tasks our team had developed as part of multi-day lab activity in her instructional unit. Students work in groups on the lab activity and in the process they discuss and write up responses to questions that focus on forms of energy, energy transformations, and creating diagramatic models that describe the phenomenon motivating the lab (e.g., the burning of a cheese-puff). A teacher may walk around different stations in the classroom to listen to student discussions and/or read written responses in real time. By noticing differences in how students are making sense of the energy-related phenomenon, and by connecting what is being heard or read to distinctions suggested by the learning progression, the teacher decides on the next instructional move to take (e.g., ask probing questions to a specific student or student group, convene the full class to have the different groups share their answers, etc.). In this hypothetical use case, assessment is happening in the moment, yet nothing is being formally scored, and if a teacher makes the wrong initial inference about what a student or group of students understands about the phenomenon in question and its relationship to the more general concept of energy flows, there will be other opportunities to adjust this inference by collecting additional information. For the assessment to be valid for this formative use, critical sources of evidence to gather are whether students find the task engaging, whether the task and questions posed are accessible to all students (e.g., English Language Learners, students with disabilities) and whether the assessment is successful at surfacing distinct student conceptions the teacher is able to use to provide feedback and adjust instruction.

As an example of validity evidence specific to the more summative use of monitoring student learning, consider the use case of two different forms of an assessment targeting the modeling energy flows LP, with each form comprised of a combination of a unique PBIC along with a collection of conceptually oriented MC items. Every student in the school district taking a grade 9 course in the physical sciences will take the assessment once a few months into the school year, and then again near the end of the school year. On each occasion, the scores from the assessment will be used assign students grades, and the growth in scores across occasions is intended to be used by district staff to make comparisons across classrooms and schools. For the assessments to be valid for these uses, considerable scrutiny needs to be placed on the psychometric properties of the assessments as measures of a student's location on the modeling energy flows construct. Critical sources of evidence would include

- the alignment of the PBIC and MC items with the different levels of the modeling energy LP;
- the distribution of item difficulty and whether variability in item difficulty can be explained by intentional design features of the assessment;
- the intercorrelation of assessment items and whether this can be accounted for by a single dimension of student ability or whether multiple dimensions are needed;
- the reliability of assessment scores and the distinctions among individuals students that they support;
- the comparability of scores from two different assessment forms; and
- whether a common scale could be created to depict growth across the two assessment forms.

Importantly, any single assessment item (or even groups of items) that might contribute to the validity of one particular use, may not necessarily contribute to the validity of another. For example, when administered in tandem with a PBIC, a single MC item may contribute supporting information about specific student conceptions that help to increase the generalizability of the score inferences from the assessment. But if the same MC item is used in isolation as a concept question to spur discussion at the start of class, it may not support valid inferences relative to an LP for modeling energy if it has not been connected to either a motivating phenomenon or an intent to focus on diagramatic or explanatory modeling.

A full discussion of the concept of building and testing interpretive arguments for assessment use is outside the scope of this chapter, but see Shepard, 1993; Kane, 2006; Chapter 3 of NRC, 2014; Pellegrino, DiBello & Goldman, 2016. In the specific context of an LP for energy, past empirical research has focused on using student response data and psychometric models not so much to validate a specific use of scores to make inferences about individual students, but to validate the developmental theory underlying the delineation of levels (Neumann et al., 2013; Hermann-Abell & DeBoer, 2017; Park & Liu, 2016). Such work is also relevant in the particular context we have described here, and we think it has the potential to be even more informative and defensible to the extent data collection is embedded within a known curricular sequence, something that was a focal point of our project. For more on issues related to the use of psychometric modeling to validate a learning progression hypothesis, see Briggs, 2012.

Conclusion

In this chapter we have used the context of our work building an assessment system for a research project with a school district to illustrate the way that a learning progression and a system of curricular embedded assessments can be used to both provide feedback about student understanding and to monitor evidence of student learning in the context of NGSS performance expectations. The particular learning progression that we introduce here on modeling energy flows has some important defining characteristics. The first characteristic is that it is an embodiment of a sociocognitive learning theory. It draws from the research literature to speculate about a path students are likely to traverse as they are exposed to instruction about energy within and across a disciplinary sequence. The second characteristic is that the learning progression can be mapped to both the content standards that represent the coin of the realm in most school districts, as well as to the curricular units to which it best applies. In this particular example, it was important to appreciate that the learning progression could be seen as an elaboration of a given NGSS performance expectation that includes energy as a cross-cutting concept and modeling as a focal practice. The elaboration comes in the specification of levels that fall below and above the statement of what it entails for students to meet the performance expectation. The third characteristic is that the learning progression contains information that teachers can use "move" students from one level to the next. These are three characteristics (embodiment of a learning theory, aligned with content standards and curriculum, and providing instructionally relevant feedback) that should generalize to any learning progression if it is under consideration for use in classroom settings.

Taken together, a learning progression with embedded assessments has the potential to comprise a powerful framework for professional development, and we presented the formative assessment design cycle that takes places within teacher PLC meetings as the location where this framework is realized. In these design cycles teachers have the opportunity to revise or tailor the learning progression to the specifics of their curricular units, and use or develop assessment tasks to support these units so long as the tasks are designed with an eye toward making the distinctions in student thinking hypothesized by the learning progression. Student-work focus sessions provide teachers with opportunities to make these distinctions visible and to take a critical look at the quality of their available assessment tasks. All of this is intended to give teachers greater ownership over the assessment of their students.

The assessment system that we introduce as part of the infrastructure of the learning progression contains three types of assessment tasks: performance-based tasks and labs, phenomenon-based item clusters, and conceptually oriented multiple choice items. Performance-based tasks can be written with a generality that cuts across disciplinary boundaries and may be the most authentic to the ideal the NGSS might have for students as scientists in training. However they can be very time-consuming to administer, and without the right supports, may be unlikely to elicit useful distinctions in student thinking. A rich use for performance-based tasks is as a basis for multi-day scaffolded projects or labs. The PBICs attempt to mimic these scaffolded lab activities, but over a more constrained domain and in a much more limited amount of time. Finally, conceptually oriented MC items remain an important tool because they are efficient to administer while still having the potential to provide insights about student misconceptions. We argue that assessments for both formative and summative purposes can be supported within a single assessment system when the assessments are motivated by a common learning progression hypothesis. It is the learning progression that, in principle, can help maintain the coherence of the assessments for these different purposes. However, the validity of any learning progression and the assessments that are motivated by the progression are always a subject for ongoing investigation and improvement.

References

Alonzo, A. C. (2011). Learning progressions that support formative assessment practices. *Measurement: Interdisciplinary Research & Perspective*, 9(2-3), 124-129.

Alonzo, A. C. & Gotwals, A. W. (2012). Learning progressions in science: current challenges and future directions. Rotterdam, The Netherlands: Sense Publishers.

Board on Science Education. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas.* Washington D.C.: National Academies Press.

Briggs, D. C. (2012). Making progress in the modeling of learning progressions. In A. Alonzo & A. Gotwals (Eds.) *Learning Progressions In Science* (pp. 293-316). Sense Publishers.

Briggs, D., Alonzo, A., Schwab, C. & Wilson, M. (2006) Diagnostic assessment with ordered multiple-choice items. *Educational Assessment*, 11(1), 33-64.

Chen, R. F., Eisenkraft, A., Fortus, D., Krajcik, J., Neumann, K., Nordine, J. & Scheff, A (2014). Teaching and Learning of Energy in K-12 Education. Springer: DOI 10.1007/978-3-319-05017-1.

Coffey, J. E., Hammer, D., Levin, D. M., & Grant, T. (2011). The Missing Disciplinary Substance of Formative Assessment. *Journal of Research in Science Teaching*, *48*(10), 1109–1136.

Corcoran, T., Mosher, F. A., & Rogat, A. (2009). Learning Progressions in Science: An Evidence-Based Approach to Reform. Philadelphia, PA: Consortium for Policy Research in Education.

Duschl, R., Maeng, S., & Sezen, A. (2011). Learning progressions and teaching sequences: a review and analysis. *Studies in Science Education*, *47*(2), 123–182.

Duit, R. (2014). Teaching and learning the physics energy concept. In Teaching and Learning of Energy in K–12 Education (pp. 67-85). Springer International Publishing.

Feynman, R. P., R. B. Leighton, and M. L. Sands. (1989). *The Feynman lectures on physics.* Vol. 1. Redwood City, CA: Addison-Wesley.

Furtak, E. M. (2012). Linking a learning progression for natural selection to teachers' enactment of formative assessment. *Journal of Research in Science Teaching*, *49*(9), 1181–1210.

Furtak, E. M., & Heredia, S. (2014). Exploring the Influence of Learning Progressions in Two Teacher Communities. *Journal of Research in Science Teaching*, *51*(8), 982–1020.

Furtak, E. M., Morrison, D. L., & Kroog, H. (2014). Investigating the Link Between Learning Progressions and Classroom Assessment. *Science Education, 98*(4), 640-673.

Furtak, E.M., Binder, T., & Henson, K. (2018). *Designing from outer space: Designing from Outer Space: Tensions in the Development of a Task to Assess a Crosscutting Concept.* In Kay, J. and Luckin, R. (Eds.), Rethinking Learning in the Digital Age: Making the Learning Sciences Count, 13th International Conference of the Learning Sciences (ICLS) 2018, Volume 3. London, UK: International Society of the Learning Sciences, p. 528-535.

Gröschner, A., Seidel, T., & Katharina, A. P. (2014). Facilitating collaborative teacher learning : the role of " mindfulness " in video-based teacher professional development programs, 1–18.

Herrmann-Abell, C. F., & DeBoer, G. E. (2018). Investigating a learning progression for energy ideas from upper elementary through high school. Journal of Research in Science Teaching, 55(1), 68-93.

Jin, H., & Anderson, C. W. (2012). A learning progression for energy in socio-ecological systems. Journal of Research in Science Teaching, 49(9), 1149-1180.

Jin, H., Zhan, L., & Anderson, C. W. (2013). Developing a fine-grained learning progression framework for carbon-transforming processes. International Journal of Science Education, 35(10), 1663-1697.

Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From Evidence to Action: A Seamless Process in Formative Assessment? *Educational Measurement: Issues and Practice*, *28*(3), 24–31.

Kane, M. T. (2006). *Validation* in R. Brennan (Ed.) Educational Measurement, 4th Edition, Westport, CT: Praeger.

Kang, H., Thompson, J., & Windschitl, M. (2014). Creating Opportunities for Students to Show What They Know: The Role of Scaffolding in Assessment Tasks. *Science Education*, *98*(4), 674–704. https://doi.org/10.1002/sce.21123

Lacy, S., Tobin, R. G., Wiser, M., & Crissman, S. (2014). Looking through the energy lens: a proposed learning progression for energy in grades 3–5. In Teaching and Learning of Energy in K–12 Education (pp. 241-265). Springer International Publishing.

Lee, H. S., & Liu, O. L. (2010). Assessing learning progression of energy concepts across middle school grades: The knowledge integration perspective. Science Education, 94(4), 665-688.

Liu, O. L., Ryoo, K., Linn, M. C., Sato, E., & Svihla, V. (2015). Measuring knowledge integration learning of energy topics: A two-year longitudinal study. International Journal of Science Education, 37(7), 1044-1066.

Liu, X., & Collard, S. (2005). Using the Rasch model to validate stages of understanding the energy concept. Journal of Applied Measurement, 6(2), 224–241.

Liu, X., & McKeough, A. (2005). Developmental growth in students' concept of energy: Analysis of selected items from the TIMSS database. Journal of Research in Science Teaching, 42(5), 493-517.

Loughran, J., Mulhall, P., & Berry, A. (2004). In Search of Pedagogical Content Knowledge in Science: Developing Ways of Articulating and Documenting Professional Practice. *Journal of Research in Science Teaching*, *41*(4), 370–391.

McLaughlin, M. W., & Talbert, J. E. (2001). *Professional communities and the work of high school teaching*. Chicago, IL: University of Chicago Press.

National Research Council. (2001). *Classroom Assessment and the National Science Education Standards*. Washington, D.C.: National Academy Press.

National Research Council. (2014). *Developing Assessments for the Next Generation Science Standards.* Committee on Developing Assessments of Science Proficiency in K-12. Board on Testing and Assessment and Board on Science Education, J.W. Pellegrino, M.R. Wilson, and A.S. Beatty, *Editors.* Division of Behavioral and Social Sciences and Education. Washington, D.C.: The National Academies Press.

Neumann, K., Viering, T., Boone, W. J., & Fischer, H. E. (2013). Towards a learning progression of energy. Journal of Research in Science Teaching, 50(2), 162-188.

Nordine, J. (2016). Teaching Energy Across the Sciences. NSTA Press: Arlington, VA.

Nordine, J., Krajcik, J., & Fortus, D. (2011). Transforming energy instruction in middle school to support integrated understanding and future learning. Science Education, 95(4), 670-699.

Opitz, S. T., Harms, U., Neumann, K., Kowalzik, K., & Frank, A. (2015). Students' energy concepts at the transition between primary and secondary school. Research in Science Education, 45(5), 691-715.

Park, M., & Liu, X. (2016). Assessing understanding of the energy concept in different science disciplines. Science Education, 100(3), 483-516.

Pellegrino, J. W., DiBello, L. V. & Goldman, S. R. (2016) A framework for conceptualizing and evaluating the validity of instructionally relevant assessments, *Educational Psychologist*, 51:1, 59-81, DOI: 10.1080/00461520.2016.1145550

Penuel, W. R., & Shepard, L. A. (2016). Assessment and Teaching. In D. H. Gitomer & C. Bell (Eds.), *Handbook of Research on Teaching* (Fifth, pp. 787–850). Washington D.C.: American Educational Research Association.

Pierson, A. E., Clark, D. B., & Sherard, M. K. (2017). Learning progressions in context: Tensions and insights from a semester-long middle school modeling curriculum. *Science Education*, 101, 1061–1088.

Popkin, G. (2009). Physics first battles for acceptance. APS News. Retrieved June 28, 2018 from https://www.aps.org/publications/apsnews/200907/physicsfirst.cfm.

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D.,& Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632-654.

Shepard, L. (1993). Evaluating test validity. *Review of Educational Research*, 19, 405-450.

Shepard, L., Penuel, B., & Pellegrino, J. (2018a). Using Learning and Motivation Theories to Coherently Link Formative Assessment, Grading Practices, and Large-Scale Assessment. *Educational Measurement: Issues and Practice*, 37(1), 21-34.

Shepard, L., Penuel, B., & Pellegrino, J. (2018b). Classroom Assessment Principles to Support Learning and Avoid the Harms of Testing. *Educational Measurement: Issues and Practice*, 37(1), 52-57.

Wilson, M. (2018). Making measurement important for education: the crucial role of classroom assessment. *Educational Measurement: Issues and Practice*, 37(1), 5–20.

Table 1. NGSS Progressions of Energy Concepts Across Grade Bands

	K-2	3-5	6-8	9-12
Energy as a DCI				
PS3.A Definitions of Energy	N/A	Moving objects contain energy. The faster the object moves, the more energy it has. Energy can be moved from place to place by moving objects, or through sound, light, or electrical currents. Energy can be converted from one form to another form.	Kinetic energy can be distinguished from the various forms of potential energy. Energy changes to and from each type can be tracked through physical or chemical interactions. The relationship between the temperature and the total energy of a system depends on the types, states, and amounts of matter.	The total energy within a system is conserved. Energy transfer within and between systems can be described and predicted in terms of energy associated with the motion or configuration of particles (objects). Systems move toward stable states.
PS3.B Conservation of Energy and energy transfer	Sunlight warms Earth's surface.			
PS3.C Relationship between energy and forces	Bigger pushes and pulls cause bigger changes in an object's motion or shape.	When objects collide, contact forces transfer energy so as to change the objects' motions	When two objects interact, each one exerts a force on the other, and these forces can transfer energy between them.	Fields contain energy that depends on the arrangement of the objects in the field.
PS3.D Energy in chemical processes and everyday life	Sunlight warms Earth's surface.	Energy can be "produced," "used," or "released" by converting stored energy. Plants capture energy from sunlight, which can later be used as fuel or food.	Sunlight is captured by plants and used in a reaction to produce sugar molecules, which can be reversed by burning those molecules to release energy.	Photosynthesis is the primary biological means of capturing radiation from the sun; energy cannot be destroyed, it can be converted to less useful forms.
Energy as a CC				
Energy and Matter	observe objects may break into smaller pieces, be put together into larger pieces, or change shapes.	matter is made of particles and energy can be transferred in various ways and between objects.	matter is conserved because atoms are conserved in physical and chemical processes. They	the total amount of energy and matter in closed systems is conserved. They can describe changes of

		Students observe the conservation of matter by tracking matter flows and cycles before and after processes and recognizing the total weight of substances does not change.	also learn within a natural or designed system, the transfer of energy drives the motion and/or cycling of matter. Energy may take different forms (e.g. energy in fields, thermal energy, energy of motion). The transfer of energy can be tracked as energy flows through a designed or natural system.	energy and matter in a system in terms of energy and matter flows into, out of, and within that system. They also learn that energy cannot be created or destroyed. It only moves between one place and another place, between objects and/or fields, or between systems. Energy drives the cycling of matter within and between systems. In nuclear processes, atoms are not conserved, but the total number of protons plus neutrons is conserved.

Figure 1. A Learning Progression for Modeling Energy Flows in Systems

Level	Description	Key Indicators
5: Modeling energy flows for generalizing and predicting	• Students are able to ***generalize their model*** to unknown or multiple phenomena, and can ***explain limitations of applying the model*** to a new phenomenon.	○ Multiple phenomena or ○ Predictable changes within given phenomenon ○ Limitations to generalization in scope or intensity
4: Modeling energy flows, including the mechanisms by which energy is stored or changed, to account for changes in a phenomenon	• Students ***develop a model*** that illustrates a mechanism that can explain or predict the phenomenon, AND ***use*** the model to make predictions about how changing one part of the model would influence energy flows elsewhere in the system. • Students can explain how the total energy of the system constrains the magnitude of change possible. • Students can ***describe limitations of the model*** in explaining or predicting the phenomenon	○ Multiple scales within model (zoom-out or zoom-in) ○ Kinetic energy represented as molecular motion ○ Potential energy represented as stored in fields within a system ○ Radiation represented as particles or waves ○ Describes how the total energy constrains the system in some way. Either by requiring energy into the system, loss due to degradation/dissipation, or limits to the amount of change possible.

3: Modeling energy flows to account for changes in a phenomenon	• Students ***use*** or ***develop*** a model that relates changes in the phenomenon directly to changes in energy through transfers/transformations by identifying specific indicators. • Students begin to show evidence that their model is accounting for conservation and dissipation. • Model includes energy flows into, within, and out of the system.	○ Explicitly relate changes in energy to changes in phenomenon. ○ Transfers into and/or out of the system.
2: Modeling energy flows to illustrate the pattern of energy flow	• Students ***use*** or ***develop*** a model to illustrate a relationship or pattern between the increase in one form of energy and the decrease in another form, or the transfer of energy from one location or object to another. • Students identify the most relevant components and relationships in the model *and* distinguish between the system and surroundings • Model focuses on energy flows within the system only.	○ Transfers start in an object and end in another object. ○ Transformations start in one form and end in another form. ○ System and surroundings clearly identified and justifiable.
1: Using models as literal representations	• Students ***use*** or ***develop*** a model that shows, through drawings or labels, the components involved in a phenomenon, and some (but not necessarily all) relevant energy forms, transfers, or transformations.	○ Literal components ○ Energy forms labeled ○ Energy transfers may not start or end in an object ○ No clear source for transformations (Energy created or destroyed)

Phenomenon: *An observable event or state that can be explained or predicted through scientific investigation.*
System and Surrounding: *The system includes the part of the universe under investigation and the surroundings include everything outside of the investigation. A system is an organized group of related objects or components that form a whole. Systems can consist, for example, of organisms, machines, fundamental particles, galaxies, ideas, and numbers.*
Indicator: *How changes in energy are manifested. These are the observable differences in a phenomenon (e.g., when an object speeds up or slows down), that let students know there is a transfer or transformation of energy.*
Mechanism: *The entities and activities that produce the changes in energy flow (e.g., changes in kinetic energy can best be explained through particle motion).*

Table 2. Map of Performance Expectations and Unique DCIs Relevant to Assessments of Modeling Energy Flows LP in High School

Course	Performance Expectation	Associated DCIs and Unique CCs
Physics (grade 9)	HS-PS3-2 Energy Develop and use models to illustrate that energy at the macroscopic scale can be accounted for as a combination of energy associated with the motions of particles (objects) and energy associated with the relative positions of particles (objects).	PS3.A Definitions of Energy
Physics (grade 9)	HS-ESS2-3 Earth's Systems Develop a model based on evidence of Earth's interior to describe the cycling of matter by thermal convection.	ESS2.A Earth Materials and Systems ESS2.B Plate Tectonics and Large-Scale Interactions PS4.A Wave Properties
Chemistry (grade 10)	HS-PS1-4 Matter and its Interactions Develop a model to illustrate that the release or absorption of energy from a chemical reaction system depends upon the changes in total bond energy.	PS1.A Structure & Properties of Matter PS1.B Chemical Reactions
Chemistry (grade 10)	HS-PS3-4 Energy Plan and conduct an investigation to provide evidence that the transfer of thermal energy when two components of different temperature are combine within a closed system results in a more uniform energy distribution among the components in the system (second law of thermodynamics)	PS3.B Conservation of Energy and Energy Transfer PS3.D Energy and Chemical Processes CC: Systems and System Models
Biology (grade 11)	HS-LS1-7 From Molecules to Organisms: Structures and Processes Use a model to illustrate that cellular respiration is a chemical process whereby the bonds of food molecules and oxygen molecules are broken and the bonds in new compounds are formed resulting in a net transfer of energy.	LS1.C Organization for Matter and Energy Flow in Organisms
Biology (grade 11)	HS-LS2-5 Ecosystems: Interactions, Energy, and Dynamics Develop a model to illustrate the role of photosynthesis and cellular respiration in the cycling of carbon among the biosphere, atmosphere, hydrosphere, and geosphere.	LS2.B: Cycles of Matter and Energy Transfer in Ecosystems PS3.D: Energy in Chemical Processes CC: Systems and System Models