The Lawrence Hall of Science

UNIVERSITY OF CALIFORNIA, BERKELEY

RESEARCH BRIEF

Articulating a Transformative Approach for Designing Tasks that Measure Young Learners' Developing Proficiencies in Integrated Science and Literacy

Alison K. Billman, Daisy Rutstein, and Christopher J. Harris

Abstract

As early elementary classrooms shift to implementing Next Generation Science Standards (NGSS) instruction, high-quality assessments are essential for providing teachers with information about where students are in the process of developing proficiency in science. In this paper, we introduce an approach for designing NGSS-aligned assessments that measure young learners' science progress while also attending to the scientific language and literacy practices that are integral parts of the NGSS Performance Expectations. Grounded in the tenets of evidence-centered design (ECD), this approach provides guidance for attending to the typical developmental characteristics of young learners with considerations of their emerging language and literacy development explicitly incorporated into the process. We describe the design process, provide an example task explicitly designed for first-grade students, and consider implications and future research.

Introduction

The National Research Council's (NRC) *Framework for K-12 Science Education* (2012) and the NGSS (2013) emphasize that all students, beginning in the earliest grades, must have opportunities to learn science. Importantly, even the youngest students should have sustained and coherent science experiences in which they are supported in applying what they know to make sense of the natural world. This vision holds promise for engaging a broad diversity of young students in learning science if teachers have the tools to examine, reflect on, and improve their science instruction. As early-elementary classrooms shift to NGSS instruction, high-quality classroom assessments are needed to provide critical information to help teachers determine what their students know and can do relative to the NGSS Performance Expectations—information that is critical for making informed instructional decisions (Ruiz-Primo & Furtak, 2007).

Designing assessments that align with the NGSS requires an approach that embraces a new way of thinking about what it means to demonstrate proficiency in science. The NGSS are grounded in the idea that proficiency in science means the ability to use science ideas to engage in real-world problem-solving, reasoning from evidence, and explaining natural phenomena. Rather than defining science proficiency as a discrete set of science content, the three dimensions of the NGSS combine Science and Engineering Practices (engaging in science) with Disciplinary Core Ideas (science content) and Crosscutting Concepts (unifying principles) to articulate Performance Expectations. These Performance Expectations provide examples of science knowledge-in-use; that is, what students should know and be able to do at the end of a grade level or grade band. Designing assessments that reflect this view of science proficiency presents a new challenge for developing assessment tasks that not only capture proficiency in the three dimensions of the NGSS, but that also support teachers in making instructional decisions (Furtak, 2017; Gorin & Mislevy, 2013; Harris, et al., 2018; NRC, 2014; Pellegrino, 2013).

Designing assessments to measure young students' science proficiency needs to account for who they are as novice learners and knowers. Assessment designers face unique challenges when designing any type of assessment for capturing what young students know and can do. These students are at the beginning of their school careers and are novices in all subjects as well as in the practices associated with those subjects. For example, in the case of science, as students are learning scientific language and using scientific language to learn, they are also learning more generally about language (how it is structured and how it is encoded in text) and are beginning to develop reading and writing skills. Additionally, other challenges arise because young students are still growing physically and socially in ways that influence how they are able to demonstrate what they know. When expressly tackling the need for NGSS-aligned assessments for use in early-grade classrooms, designers face these questions:

- *How do we measure what young students know and can do when these same students are just developing the ability to identify and express what they know?*
- *How do we measure proficiency with the three dimensions of the NGSS when knowledge-in-use is fundamentally facilitated and captured through language at a time when young students are just acquiring and developing proficiency with language?*

Background and Rationale

Robust NGSS-aligned assessments for young students need to be responsive to the NRC recommendations outlined in *Developing Assessments for the Next Generation Science Standards* (NRC, 2014) and account for the three dimensions of science as well as science-specific language and literacy practices associated with the discipline. As recommended in the report, design approaches should follow the argumentative reasoning of evidence-centered design (ECD) and should support NGSS teaching and learning. In accordance with the tenets of ECD (Mislevy & Haertel, 2006), the Next Generation Science Assessments for Young Scientists (NGSA-YS) is a systematic process for assessment design and development. It builds on the Next Generation Science Assessment (NGSA) process developed by Harris and colleagues (Harris, et al., 2019), which has been used to develop NGSS-aligned tasks for upper-elementary and secondary classrooms. The NGSA-YS extends the NGSA to early-elementary classrooms. It expands the NGSA by explicitly attending to science language and literacy to create NGSS-aligned tasks that assess relevant scientific language and literacy and by explicitly attending to the domains of child development to ensure that tasks are developmentally appropriate for early-elementary classrooms. The following three principles provide the theoretical grounding of the NGSA-YS design approach:

Principle 1: Language is a tool for both constructing and expressing science understanding. Sociolinguistic theories of language development identify language as the prototypical resource for making meaning (Halliday, 1993; Wells, 1994). Language is simultaneously a system for organizing specialized knowledge (e.g., science) and a resource for understanding and participating (through reading, writing, listening, speaking, and thinking critically) in a specialized community. In conjunction with this definition of language, literacy is defined as the functional use of language to navigate social contexts. In this case, scientific literacy is the functional use of the disciplinary language of science to navigate the social context of a science community situated within a classroom. In this respect, science and literacy are interwoven and synergistic rather than separate (Hart & Lee, 2003; Pearson et al., 2010; Stoddart et al., 2002), and science-specific disciplinary language is necessary for discussing science facts, hypotheses, arguments, and claims (Snow, 2008). As students investigate to understand phenomena, they must learn scientific ways of talking and writing to communicate findings and provide explanations supported by evidence (Bailey et al., 2007; Pearson & Billman, 2016).

Principle 2: Assessments aligned with the NGSS must attend to the practices of the discipline. The NGSS call for an integrated, three-dimensional approach to science instruction. In response, robust assessments must also be three dimensional and measure students' facility with Science and Engineering Practices in the context of Disciplinary Core Ideas and Crosscutting Concepts in an integrated, three-dimensional approach rather than knowledge of disciplinary ideas alone. Designing assessments that are aligned with the

NGSS is challenging because of the complex domain definitions of these standards (DeBarger, et al., 2016; Gorin & Mislevy, 2013; NRC, 2014). A three-dimensional approach raises additional challenges when designing assessments for early-elementary students due to students' degree of maturation across physical, cognitive, social and emotional, and language domains of development. For example, many early-elementary students are just beginning to develop the physical dexterity, cognitive endurance, attention span, and language and literacy skills required to participate in assessment conversations and/or tasks about what they know and can do (Shepard, 1994).

Principle 3: Science and literacy are interwoven and synergistic, so NGSS-aligned assessments must address both. As previously noted, science-specific disciplinary language is essential for discussing science facts, hypotheses, arguments, and claims. As assessment developers work at the intersection of Disciplinary Core Ideas, Science and Engineering Practices, and Crosscutting Concepts, they must attend to the science-specific language and literacy-related practices that are used as tools for constructing understanding and expressing science ideas. This sets a high bar for assessment in general and even more so for designing assessments for early-elementary students who are at the beginning stages of acquiring facility with language and literacy skills. For example, item formats that require independent reading and writing may interfere with these students' ability to provide valid evidence of what they know and can do because most of them are not yet proficient readers and writers. Differences in students' facility with language could be misinterpreted as differences in knowledge (Hobbs et al., 2012; Samarapungavan et al., 2008). An efficacious approach needs to explicitly identify which aspects of language and literacy are required to exhibit science proficiency (construct-relevant) and which are not required (construct-irrelevant). Attending to both construct-relevant science language and literacy and construct-irrelevant language skills and abilities results in tasks that are more likely to provide reliable evidence of student understanding while at the same time reducing the potential roadblocks (e.g., decoding skills) to students communicating their understanding.

The NGSA-YS Development Process

Building on the foundation of the NGSA, the NGSA-YS design process consists of six iterative steps across four distinct phases of development: (1) **domain analysis**, which calls for unpacking of the three dimensions of the NGSS and the integral language and literacy elements inherent in the Performance Expectations in order to identify and understand the assessable components; (2) **domain modeling**, which involves constructing learning performances and specifying the evidence for demonstrating them; (3) articulating a **conceptual assessment framework**, which involves specifying task templates that describe the characteristic and variable features to be designed into tasks; and (4) **assessment tasks development**, which entails using the task templates and design artifacts from prior steps to create the assessment tasks, rubrics, and administration guides. In the following subsections, we elaborate on each of these phases and the steps within them. (See Figure 1. NGSA-YS Diagram on page 4.)

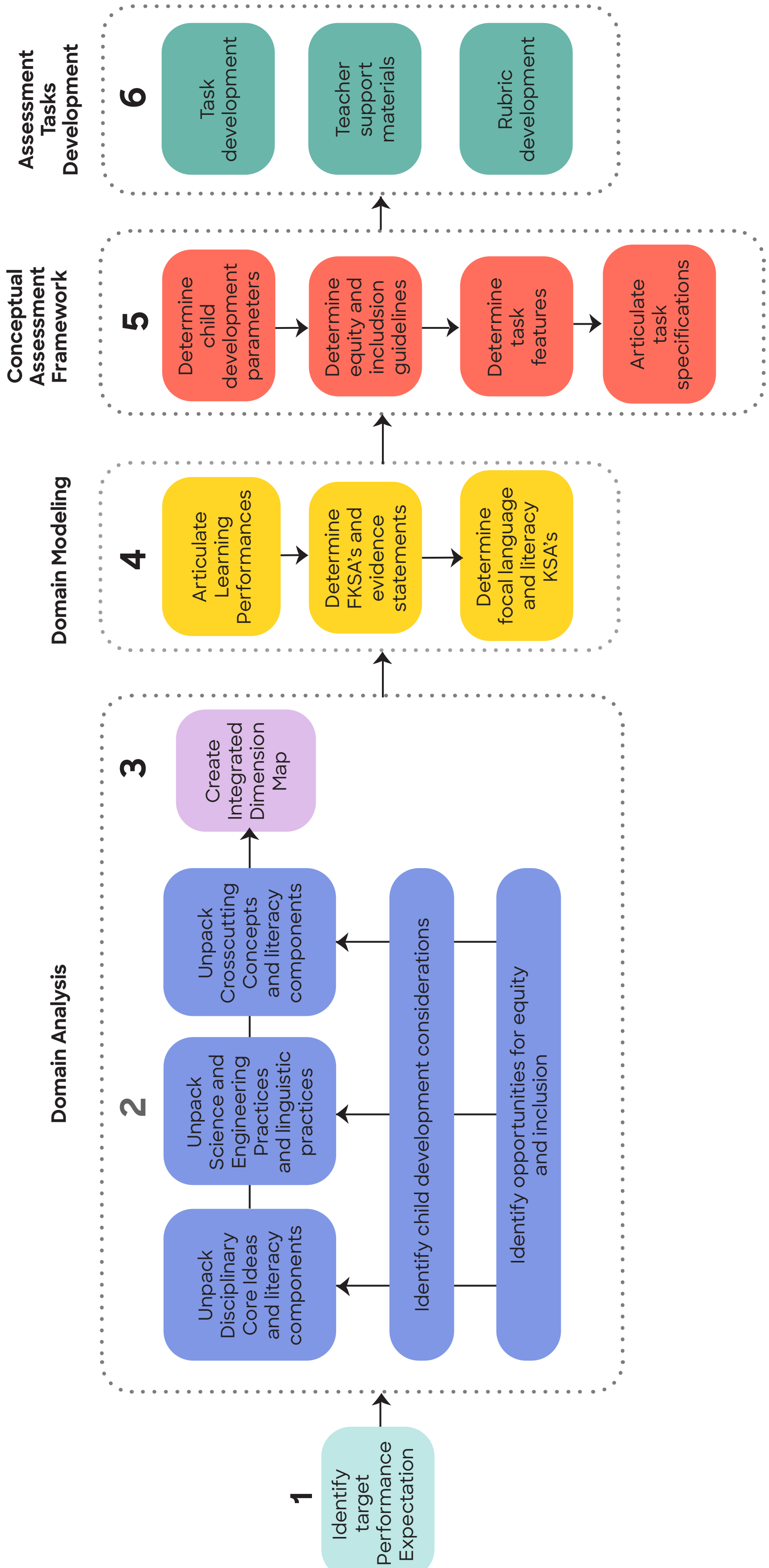
DOMAIN ANALYSIS

Developing an efficacious assessment requires a comprehensive understanding of the knowledge and learning goals that are inherent in the target Performance Expectation (identified in Step 1 of the NGSA-YS process). In other words, assessment developers need to know what they are measuring. To ensure that assessment developers have a deep understanding of the target performance, the NGSA-YS process involves a purposeful analysis (unpacking) of the three dimensions of the NGSS that comprise the Performance Expectation. The primary goal of the analysis or unpacking of the dimensions (Step 2 of the NGSA-YS process) is to identify the key aspects of each of the three dimensions and define the boundaries of proficiency required of young students.

Unpacking the dimensions. The unpacking of each dimension follows a similar pattern of thinking; however, the categories of analysis vary slightly based on the characteristic features of each dimension per the descriptions provided in the NGSS. For example, while the analysis of the Disciplinary Core Ideas identifies which core ideas students *must include* in a demonstration of proficiency, analysis of the Science and Engineering Practice identifies what students *must do* with the core ideas to demonstrate proficiency with the practice, and analysis of the Crosscutting Concept identifies the mental tools or perspectives students *must use* in a demonstration of proficiency (e.g., explaining observations in terms of patterns). Analysis of each dimension includes:

(continued on page 5)

Figure 1. NGSA-YS Diagram



Note: Adapted from the Next Generation Science Assessment Design Process (Harris, Krajcik, Pellegrino, & DeBarger, 2019).

- identifying the key aspects of the dimension and defining the expectations for students at the grade level.
- the expected prerequisite knowledge (the knowledge and skills that students should have learned prior to developing proficiency with the target Performance Expectation—it is expected that first graders would bring understandings and skills learned in kindergarten).
- the aspects of the dimension that are beyond the scope of the Performance Expectation.
- the research-based potential challenges or misconceptions students might have related to each dimension.

Unique to the NGSA-YS domain analysis, or unpacking, are the additional categories of analysis that identify the essential scientific-language and literacy aspects inherent in each of the three dimensions, as well as a careful identification of key elements of child development that should be considered when designing developmentally appropriate assessments for young students. We also identify and incorporate equity and inclusion considerations to meet the call for accessible assessments (Alozie, et al., 2018).

As noted in the NGSS documents, knowing and engaging in science is language intensive and requires that students learn and use science-specific language. In that respect, unpacking analysis identifies which science-specific language and literacy structures are inherent for each dimension in the Performance Expectation. For example, students who understand the Disciplinary Core Idea that many kinds of animal parents and offspring engage in behaviors to survive would be expected to use language structures that show cause and effect (e.g., *This offspring survives because* ___.), as well as specialized science words such as *survive*. When the Science and Engineering Practice involves reading grade-appropriate texts to obtain information, the analysis indicates that it is important to define the characteristics of grade-appropriate texts for target students (e.g., readability level) and also determine how much science information needs to be included in the text in order to accomplish the goals of the Performance Expectation.

During the unpacking, we also ask questions to determine which additional characteristics of students' age-related development (e.g., language, visual perceptual, cognitive, social and emotional, and physical) need to be considered in order to design developmentally appropriate assessment tasks. For example, domain analysis using the NGSA-YS design process identifies grade-level expectations related to language and literacy skills that may influence young students' ability to demonstrate proficiency. We begin to ask and answer questions such as: *In the case of a beginning reader, should the text be read aloud? In the case of an emerging bilingual student with limited English proficiency, should the administrator prompt the student to show what they know by acting out the information they obtained about the pattern of animal behavior?* We also ask questions to determine what additional characteristics of students' age-related development (e.g., social and emotional, cognitive, physical) need to be considered in order to design developmentally appropriate assessment tasks. In Tables 1–3 (on pages 6–9), we provide excerpts from the unpacking analyses of the first grade Performance Expectation 1-LS1-2: *Read tests and use media to determine patterns in behavior of parents and offspring that help offspring survive*.

Constructing the Integrated Dimension Map (IDM). Information obtained through the unpacking of the dimensions is brought together in an IDM during Step 3 of the NGSA-YS process (Harris, et al., 2019). This visual representation illustrates the relationships between the various aspects of the three dimensions of the NGSS called out in the Performance Expectation and indicates instances when science-specific language and literacy are in play. We begin by illustrating the logical flow between the key aspects of the Disciplinary Core Idea that were identified during the unpacking of that dimension. Key aspects of the Science and Engineering Practice and the Crosscutting Concept are then added to show what students are expected to do with the core idea(s) and what Crosscutting Concept lens they will use as a tool to construct and communicate information (or explanations) related to the core idea(s). In this way, the IDM helps articulate the ways in which the three dimensions work together to define proficiency with the Performance Expectation, and it lays the groundwork for specifying the knowledge, skills, and abilities that will be measured in the assessment tasks. (See Figure 2. Integrated Dimension Map on page 10.)

DOMAIN MODELING

The results of the Domain Analysis phase provide essential information for defining the breadth and depth of the knowledge, skills, and abilities that will be measured by the assessment tasks. Once designers understand what is encompassed in the Performance

(continued on page 11)

Table 1. Aspects from Unpacking a Disciplinary Core Idea related to LS1.B: *Growth and Development of Organisms*

Aspects of a Disciplinary Core Idea	• Adult plants and animals can have young. • In many kinds of animals, parents and the offspring themselves engage in behaviors that help the offspring to survive.
Elaborating the meaning of key subideas	• Some animal parents help their offspring survive by doing things to provide food, shelter, and other forms of protection. • Some young offspring do things that help them to survive, such as signaling their parents, making themselves hidden from predators, getting food.
Defining expectations for understanding	• Students should learn that (1) animal and plant parents produce offspring, (2) parents and offspring do things (behaviors) to survive, (3) survival behaviors include getting food and water and defending or finding shelter to avoid being eaten.
Defining expectations for scientific language and literacy proficiency*	• Students should be able to know and use these terms orally: *survive*, *offspring*, *parents*, *because/so.* • Students should be able to orally use causal phrasing to explain how offspring survive. For example: *Parents fed the offspring, so the offspring survived. A sea turtle can survive because it uses camouflage to not be found.* • Students should be able to visually discriminate between adults and offspring in images and media. • Students should be able to visually recognize and describe behaviors of adults and offspring that contribute to survival in images and media.
Identifying assessment boundaries	• Students do not need to know which animals do which behaviors. • Students do not need to know the growth and development time lines of organisms. • Students do not need to know about the life cycles of organisms.
Prerequisite knowledge	• Basic knowledge of what it means to be living or nonliving. • Knowledge that living things need water, air, and resources from the land (food) in order to survive.
Student challenges	• Students might not realize that a young animal is the same type of animal as its parent because of observable differences (e.g., color or type of feather). • Students might not believe that offspring can survive on their own.
Relevant phenomena	• Young birds chirp to signal to their parents that they need food. • Young penguins group together and make loud noises to defend themselves from predators.
Child Development Considerations	
Language*	• Students may be familiar with everyday ways of naming or describing phenomena (e.g., *baby* instead of *offspring* or *staying alive* instead of *surviving*).
Visual perceptual*	• Students are just learning to interpret images. Images used as sources of evidence should not be so complex that the main idea of the image is masked or obscured.

*Categories specific to the NGSA-YS approach

(Table 1 continued on next page)

Table 1. Aspects from Unpacking a Disciplinary Core Idea related to LS1.B: *Growth and Development of Organisms* (cont.)

Cognitive*	• Students' conceptual networks and representations are not as elaborate because of their relative inexperience with words and concepts. This increases the cognitive load they experience when interacting with new information.
Social and emotional*	• Students may be upset by images that show animals eating other animals.
Equity and Inclusion Considerations	
Cultural/Experiential	• Students' familiarity with different animals and plants may vary depending on community locale (i.e., urban or rural) or region.

*Categories specific to the NGSA-YS approach

Table 2. Aspects from Unpacking the Science and Engineering Practice of *Obtaining, Evaluating, and Communicating Information*

Aspects of the Science and Engineering Practice	• Read grade-appropriate texts and/or view media to obtain scientific information to determine patterns in the natural world. • Use information obtained from texts and/or media to communicate science information.
Intersections with other practices	• Obtaining information may involve analyzing and interpreting data.
Evidence required to demonstrate the practice	• Answer a scientific question or support a claim with information obtained from various texts and other media. This may require combining information from various sources. • Describe how specific images (e.g., a diagram showing how a machine works) support a scientific or engineering idea.
Defining expectations for scientific language and literacy (within the grade band)*	• Students should be able to construct meaning from grade-appropriate texts or media (including images) to obtain scientific information. • Students should be able to use scientific information from text or media (including images) to determine patterns in and/or evidence about the natural and designed world(s).
Identifying assessment boundaries (for the target grade band)	• Students do not need to choose the sources they will use as information sources. • Students do not need to know how to use text features to navigate a text. • Students do not need to decode the text independently. (It may be read aloud to them.)
Prerequisite knowledge	• Knowledge that science information can be obtained from text, media, and visual representations. • Knowledge that particular genres of text and media are resources for science information.
Student challenges	• Students may not know the difference between fiction and nonfiction genres of texts. • Students may have difficulty orally expressing some domain-specific words. • Students may have trouble interpreting linguistic information or visual images.

*Categories specific to the NGSA-YS approach

(Table 2 continued on next page)

Table 2. Aspects from Unpacking the Science and Engineering Practice of *Obtaining, Evaluating, and Communicating Information* (cont.)

Child Development Considerations	
Language*	• Students are just developing their vocabularies. • Students are just developing facilities with language to orally express what they know. • Students are just developing facilities with interpreting and producing written text.
Visual perceptual*	• Students are just developing visual discrimination—the ability to visually discriminate between objects and forms. • Students are just developing figure-ground acuity—the ability to perceive an individual object or form within a conglomerate or collection of objects or forms. • Students are just developing visual closure—the ability to fill in the missing details in an incomplete shape. This ability is important for discriminating where words start and stop and for interpreting images.
Cognitive*	• Students' conceptual networks and representations are not as elaborate because of students' relative inexperience with words and concepts. This increases the cognitive load they experience when interacting with new information. • Tasks that require too great a cognitive load can interfere with students' ability to demonstrate what they, in fact, do know. • Students' cognitive endurance (ability to sustain focus) is just developing and is measurably shorter the younger the age of the student.
Social and emotional*	• Students may be upset by certain types of images or actual aspects of the natural world (e.g., animals eating other animals).
Physical*	• Students' fine motor skills are just developing, constraining their ability to provide written responses.
Equity and Inclusion Considerations	
Equity/Inclusion	• Students may not have the language resources to communicate the information that they obtain from texts and media.

*Categories specific to the NGSA-YS approach

Table 3. Aspects from Unpacking the Crosscutting Concept of *Patterns*

Key aspects of the Crosscutting Concept	• Patterns in the natural world can be observed, used to describe phenomena, and used as evidence.
Intersections with other dimensions	• None identified.
Evidence required to demonstrate application	• Identify and describe patterns based on similarities and differences. • Identify and describe repeated occurrences of events, structures, or relationships. • Use observations of patterns as evidence to predict additional occurrences of an identified pattern.

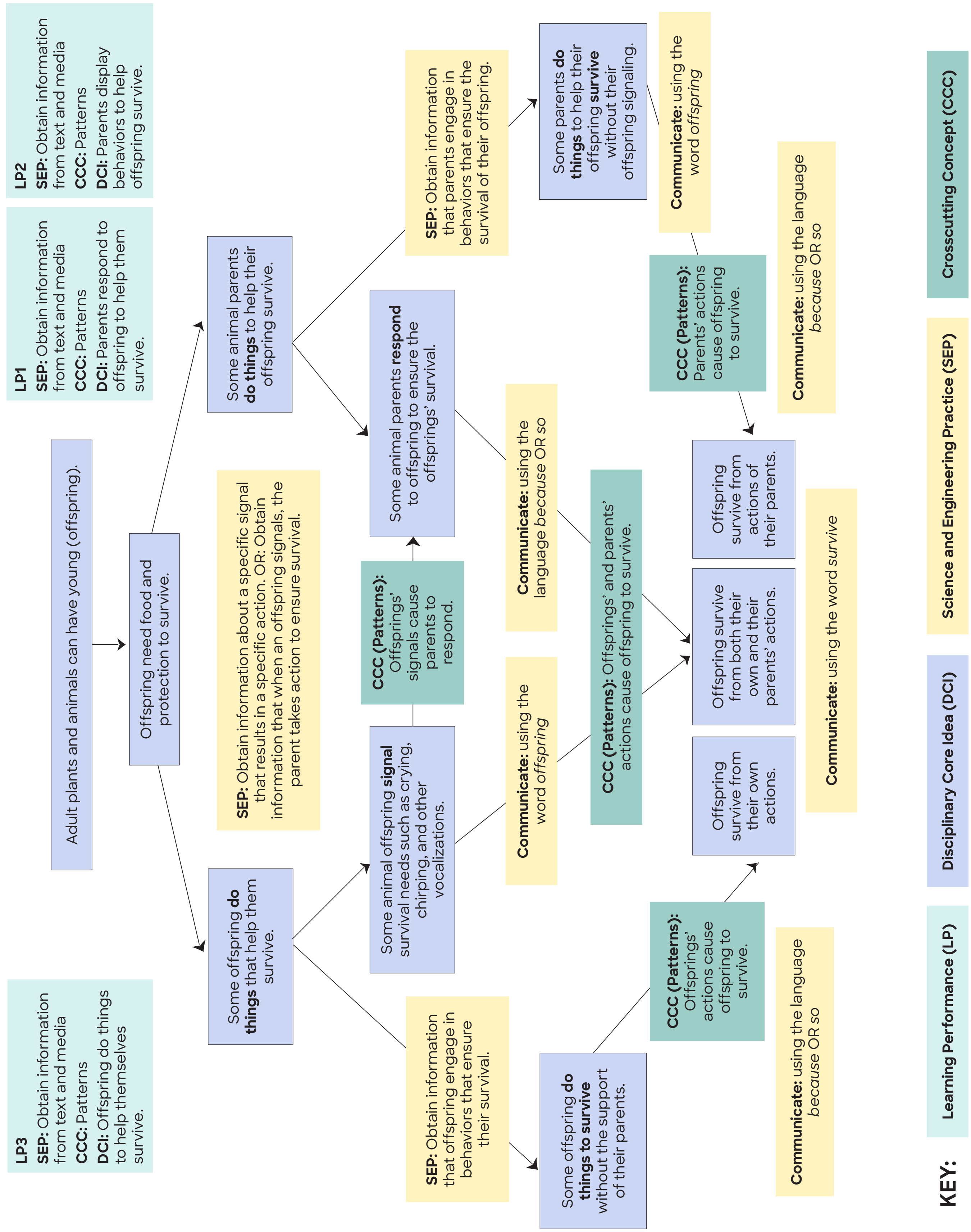
(Table 3 continued on next page)

Table 3. Aspects from Unpacking the Crosscutting Concept of *Patterns* (cont.)

Defining expectations for scientific language and literacy (within the grade band)*	· Students should know the word *pattern*. · Students should know and be able to use these terms and phrases: *observe*, *something that happens in the same way again*, *similar/same/alike*, *different*.
Identifying assessment boundaries (for the target grade band)	· Students do not need to explain why a pattern exists. · Students do not need to characterize the strength or direction of a pattern.
Prerequisite knowledge	· Knowledge of and ability to distinguish between similarities and differences. · Ability to group things into categories based on similarities and differences. · Developing knowledge that a pattern is something that repeats and is predictable. · Knowledge that things with surface-level differences (e.g., color of fur) can share common features (e.g., the same body covering—fur, number of legs).
Student challenges	· Students may have difficulties identifying the relevant information in their observations that together establish a pattern.
Child Development Considerations	
Language*	· Students are just developing language and may not have the vocabulary to describe or explain unfamiliar patterns.
Visual perceptual*	· Students are just developing visual discrimination—the ability to visually discriminate between objects and forms. This could interfere with perceiving the information that provides evidence of a pattern. · Students are just developing figure-ground acuity—the ability to perceive an individual object or form within a conglomerate or collection of objects or forms. This could interfere with perceiving the information that provides evidence of a pattern.
Cognitive*	· Pattern identification requires holding multiple features (e.g., similarities/differences, repeated occurrences) of a pattern in working memory. The more complex the pattern, the greater the cognitive load. · Tasks that require too great a cognitive load can interfere with students' ability to demonstrate what they, in fact, do know.
Social and emotional*	· None identified.
Physical*	· None identified.
Equity and Inclusion Considerations	
Cultural/Experiential	· Students begin to recognize and react to patterns in their own lives at an early age (e.g., peek-a-boo, nighttime and daytime). They are just learning to classify, record, and use patterns intentionally.
Equity/Inclusion	· Students may have difficulties recognizing patterns and may not be able to sort items by size, shape or color.

*Categories specific to the NGSA-YS approach

Figure 2. Integrated Dimension Map



Expectation through that analysis, the next step in the process is Domain Modeling—using the IDM and unpacking documents to articulate sets of intermediary performances that we call learning performances. Learning performances are knowledge-in-use statements that incorporate aspects of Disciplinary Core Ideas, Science and Engineering Practices, and Crosscutting Concepts that students need to be able to integrate as they progress toward achieving the Performance Expectation.

Articulating learning performances. We first use the IDM to determine the number of learning performances required to demonstrate mastery of the Performance Expectation. If proficiency can be demonstrated with one learning performance, the Performance Expectation can be measured with one task. If multiple learning performances are indicated, then a set of tasks will be required to fully assess the Performance Expectation. Table 4 (below) lists the learning performances identified for the NGSS Performance Expectation 1-LS1-2.

When articulating a learning performance, we describe the integrated proficiencies—that is, the proficiencies from each dimension that are used in tandem—that are required to demonstrate the learning performance. Articulating the learning performance includes identifying the *observable* evidence a student needs to provide in their response in order to show proficiency.

We organize this in a format called an assessment argument that includes the **student model** (the expected aspects of student proficiency), the **evidence model** (the evidence that students need to demonstrate while completing the task), and the **task model** (what tasks need to look like in order to ensure that students have an opportunity to provide the desired evidence). Taken together, these models create a design pattern that describes the task features required to elicit evidence of student proficiency with the learning performance.

The NGSA-YS design process includes several language and literacy considerations when articulating the student model, evidence model, and task model. Specifically, the student model not only includes the focal knowledge, skills, and attributes (FKSA's) for each dimension, but also for construct-relevant science language. An example FKSA reads: *Students are able to describe a pattern of how offspring help themselves survive, using causal language*. The evidence model would then specify that a high-level response would include the student's use of language that logically connects a cause with an effect, using language structures such as *because* or *so* and science language such as *offspring* and *survive*. Additional considerations for the task model include attending to developmental characteristics that need to be accounted for (e.g., construct-irrelevant language

Table 4. An NGSS Performance Expectation and Three Related Learning Performances

NGSS Performance Expectation 1-LS1-2	Read texts and use media to determine patterns in behavior of parents and offspring that help offspring survive.
Learning Performances (LP) for 1-LS1-2	• **LP 1:** Obtain information from text AND media to communicate a pattern in how parents help offspring when offspring signals for help. • **LP 2:** Obtain information from text AND media to communicate a pattern in how parents help offspring survive (when the offspring does not provide signals to the parents). • **LP 3:** Obtain information from text AND media to communicate a pattern in how offspring help themselves survive.

and literacy skills) to ensure that the tasks are developmentally appropriate. For example, a task may specify the grade-appropriate length (e.g., number of words) of a text that students read, or it may specify the types and quality of images to use in order to account for students' developing visual perception. Table 5 (on pages 13–14) illustrates the documentation for one of the three learning performances identified for 1-LS1-2.

CONCEPTUAL ASSESSMENT FRAMEWORK

Articulating task templates. Once we arrive at a set of learning performances that represent a Performance Expectation, the next step is to use the design patterns to develop task templates—detailed design tools that guide the creation of families of tasks (Mislevy & Riconscente, 2006). Task templates articulate four categories of information: (1) how the task is presented to the student, (2) what content is covered in the task, (3) options for how students can respond, and (4) initial sets of scoring rules. (See Table 6 on page 15 for an illustration of the task template for Learning Performance 3.) Task templates make design decisions explicit and set the boundaries for what should be included or not included in tasks. Importantly, they serve as the blueprints for task designers by providing shared information for creating tasks that are aligned to the learning performance. Throughout the work of creating the tasks, the templates serve as a common reference point for checking alignment of the task to the design goals.

During this step, the NGSA-YS process again brings attention to language and literacy and to developmental appropriateness of the task for young learners. For example, in an NGSA-YS exemplar task in which students are required to demonstrate proficiency with obtaining information from multiple sources, the task template specifies what degree of linguistic challenge (i.e., readability level, length) is appropriate for an informational text, what ratio of print-to-visual representations to include on a page of text, and what other types of media should be used as information sources. At this point, decisions are also made about the order in which the information sources are presented and at which points in the task students are prompted to identify patterns of behavior. Administration design decisions specify the degree of scaffolding at critical points during the task, including step-by-step modifications to prompts that may be used to provide additional support to students who may need it.

ASSESSMENT TASKS DEVELOPMENT

The final phase of the NGSA-YS design process involves operationalizing the task templates to develop tasks, administration materials, and scoring rubrics. In general, tasks consist of a set of items or individual questions tied together by a common, developmentally appropriate stimulus or scenario (e.g., penguin offspring survival behaviors). While each task's stimulus connects broadly to the focal Performance Expectation, the multiple items embedded within the task ensure that there are opportunities for students to demonstrate proficiency across the dimensions of the focal Performance Expectation and the science-specific language and literacy components that are called out in the learning performance task template. At this point, a complete set of task materials are developed. These include:

- student-facing stimulus materials (e.g., books, videos, cards, sets of manipulatives) and student response sheets to capture students' written responses,
- administration protocols with prompts for each item and response-capture forms for administrators to document non-written responses,
- rubrics to assist administrators in scoring student responses and then using the results of their scoring to inform instruction.

Administration protocols and scoring rubrics target which aspects of student responses should be observed, what aspects of language and literacy should be scored, and how scores relate to proficiency with the Performance Expectation.

The Penguin Task. Recall that the task template for a learning performance is the design tool that guides the creation of tasks and specifies the essential components of a task required to gather evidence of proficiency. NGSS 1-LS1-2 Learning Performance 3 (Table 4 on page 11) specifies that students obtain information from text and media to communicate a pattern in how offspring help themselves survive. The Penguin Task (which includes a full set of student and administrator materials) was designed using the task template developed for Learning Performance 3 (Table 6 on page 15). The task features two groups of penguin offspring and their responses to a common predator—a petrel. Students obtain information about one group of offspring from a video and information about the other group from a book. Each source shows one group of baby penguins as they are being

(continued on page 14)

Table 5. Knowledge-in-Use Design Pattern for a Learning Performance

Learning Performance	• **LP 3:** Obtain information from text AND media to communicate a pattern in how offspring help themselves survive.
Focal Knowledge, Skills, and Abilities (FKSA) (student model)	• Knowledge that food and protection are needed to survive. • Knowledge that animal offspring may engage in different behaviors—some that help them survive. • Students are able to identify information from text that shows how offspring help themselves survive. • Students are able to identify information from a media source that shows how offspring help themselves survive. • Students are able to identify a pattern of how offspring help themselves survive. • Students are able to describe a pattern of how offspring help themselves survive.
Evidence required to demonstrate proficiency (literacy is integrated into the statements) (evidence model)	• A statement that uses scientific language (*offspring*, *survive*) and language structures (cause and effect) to describe a pattern that relates offsprings' actions to their own survival.
Characteristic Task Features (task model)	
Integrated dimensions	• Sources of information show at least two examples of a specific type of animal offspring engaging in a behavior that helps the offspring survive. • Sources of information are texts or media. • Students are prompted to identify or describe a pattern in how offspring help themselves survive. • Students are prompted to make a prediction based on the pattern.
Literacy*	• Sources of information and prompts use scientific language (*offspring*, *survive*) and language structures as appropriate. • Sources of information provide evidence that when taken together, illustrate a pattern. • Sources of information do NOT explicitly describe the pattern that students are supposed to figure out.
Child Development-Related Features	
Language*	• Sources of information use grade-level appropriate vocabulary and language based on the recommended metrics in the Common Core State Standards for English Language Arts. • Text-based sources of information include images as well as words. • Directions and prompts use grade-appropriate language. • Tasks include follow-up prompts that reduce vocabulary load by using everyday language. If used, assessors document the use. • Students' ability to decode words is not being assessed. Tasks include the option of having the administrator read texts aloud.

*Categories specific to the NGSA-YS approach

(Table 5 continued on next page)

Table 5. Knowledge-in-Use Design Pattern for a Learning Performance (cont.)

Visual perceptual*	• Images in text or media should not be so complex that the evidence that students are expected to obtain from the image is masked or obscured. • Images only support identifying one pattern.
Cognitive*	• The number of sources of information is limited to two. • The length (e.g., time/words/pages) of each source of information follows these recommendations: 30–60 seconds of media, 60–100 words, 6–8 pages. • Scenarios provided as evidence of a pattern should only support identifying one pattern. • Scenarios should not include information that contradicts the pattern.
Social and emotional*	• Videos may show an animal surviving an attempt by a predator. • Images and videos do not show animals eating other animals.
Physical*	• Response options should not rely on students writing their answers to questions.
Equity and Includsion-Related Features	
Culturall/Experiential	• Choosing an animal that is familiar is not required but may support students in demonstrating proficiency. • Assessment resources (e.g., texts and media) do not reinforce biases.
Equity	• If students are not yet able to articulate/express their understanding, they should be provided with an option to respond receptively (e.g., point to a selection among a set of options to indicate understanding; act out the pattern to indicate). The alternate response should be documented. (For example, the assessor would document: *Student selects/demonstrates a pattern but did not use scientific language to orally describe the pattern.*)
Variable Task Features	
• the type of animal offspring and the type of offspring behaviors that are illustrating the target pattern • the type of information source presented to the student • the amount of contextual information provided to situate the task of figuring out and describing the pattern • task scaffolding features used to help students engage with the task	

*Categories specific to the NGSA-YS approach

approached by a petrel. The book (authored by the NGSA-YS team) is written in the form of a scientist's notebook with entries that document the scientist's observations of one group of baby penguins over several days, including the pengins' reactions when a petrel comes near. The task is designed as a one-on-one conversation between an administrator and a student. After interacting with each source of information, students are prompted to discuss the information they obtained—specifically, information about how the penguin offspring survive when there is a petrel. Next, after completing both interactions, students are prompted to generalize their observations by predicting how a different group of penguin offspring will survive when a petrel comes near. The average time for administering the task during a trial with students was 11 minutes. (See Figure 3. Penguin Task Student Print Materials on page 15.)

Implications and Future Research

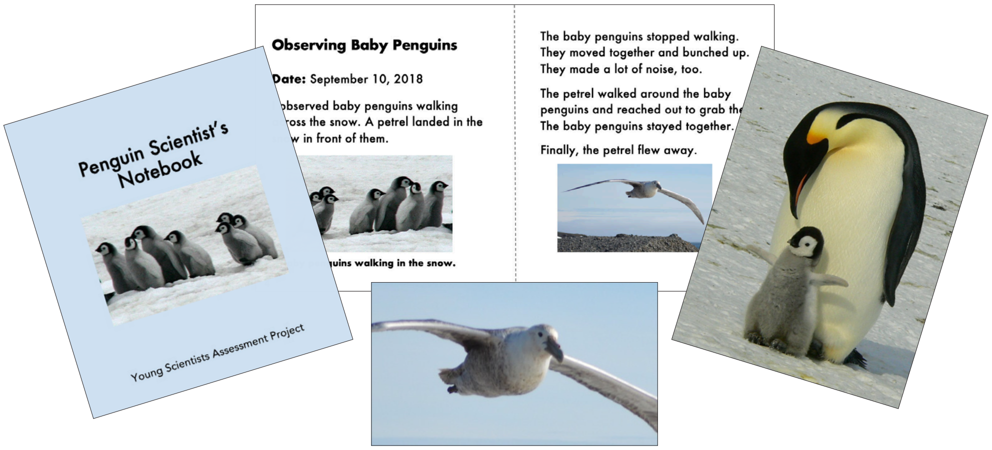
Evidence of promise. Development of the NGSA-YS included gathering initial evidence of the validity of the approach. The project team used the NGSA-YS

(continued on page 16)

Table 6. Task Template for 1-LS1-2 Learning Performance 3

How the task is presented	• Task prompts students to obtain information from two different sources (e.g., read a book and watch a video).
Content included in the task	• The content in both sources provides information about the same survival behavior pattern. • The behavior is demonstrated by the offspring of one species of animal. • The offspring are surviving on their own. • The behavior is demonstrated by two unique groups of offspring. • The words *offspring* and *survive* are included in the information sources and in the task administration script.
Options for student responses	• Students are prompted to orally describe survival behaviors of the young offspring presented in each source of information. • After obtaining information from both sources, students are prompted to provide a generalization statement about how that species of offspring can survive on their own. • Follow-up prompts are written using common everyday words (e.g., *baby* instead of *offspring*). • Follow-up prompts are used if students are not able to respond to the original prompt.
Initial sets of scoring rules	• Score for accurate identification of survival behavior. • Score for accurate identification of survival behavior pattern. • Score for students’ receptive understanding of scientific language and language structures (i.e., gauged by the need for follow-up prompts). • Score for students’ expressive use of targeted scientific language and language structures.

Figure 3. Penguin Task Student Print Materials



process to construct and try out two assessment prototype tasks that align with the NGSS Performance Expectations for first grade life science. The two tasks measure young learners' performances integrating the three dimensions of the NGSS—Disciplinary Core Ideas, Science and Engineering Practices, and Crosscutting Concepts—as well as their literacy performance within life science scenarios. To collect initial evidence of the NGSA-YS approach, the team convened a panel of experts in young students' science learning, literacy learning, and assessment to review the design process and one of the assessment tasks that was produced via the process. Cognitive interviews with first-grade students provided evidence that both tasks were able to elicit science and literacy performances. Taken together, evidence from the expert review and the cognitive interviews indicate early promise of the NGSA-YS approach.

With this evidence in hand, we hold that the NGSA-YS design process makes an important advance in NGSS assessment design for use with early-elementary students and stands to make an important contribution to a program of research focused on the design of NGSS-aligned science assessments. First, it is responsive to the NRC recommendations (NRC, 2014) for addressing the challenges in developing three-dimensional NGSS assessments, and it is grounded in the methodology of evidence-centered design (Mislevy & Haertel, 2006). Second, it extends the NGSA design process developed by Harris and colleagues (Harris et al., 2019) by paying particular attention to the emphasis on the use of scientific language and literacy required by the NGSS. Third, the NGSA-YS addresses the unique requirements for assessments for use with early-elementary students by intentionally attending to typical language and literacy profiles as well as the unique developmental characteristics of young learners. Finally, the NGSA-YS design process includes consideration for equity and inclusion with the goal of reducing bias and increasing inclusiveness. In these ways, the NGSA-YS offers a comprehensive structure that may prove beneficial to those concerned with the design and use of integrated science and literacy assessments in early-elementary classrooms.

Future research. While evidence from the expert review and the cognitive interviews indicate promise of the NGSA-YS approach, future research is needed to further validate the approach. Important to note is that the early-stage development of the NGSA-YS was focused on one grade level (first grade) and one domain of science (life science). Additional applications of the NGSA-YS are required to determine if it works equally well when designing NGSS-aligned assessment tasks for other early-elementary students (kindergarten, second, and third grades) and when applying the approach to additional science domains.

Concluding Remarks

Most existing science assessments do not capture the three-dimensional nature of the Performance Expectations in the NGSS. Nor do they attend to the language and literacy aspects of science that are called for in the NGSS and are vital in the elementary grades. In addition, few if any science assessments exist that account for child development in relation to how students are asked to demonstrate what they know and can do. The NGSA-YS approach provides a principled process for creating developmentally appropriate assessment tasks that integrate science and literacy in ways that fit the needs of early-elementary classrooms, thereby increasing the likelihood that they will be usable and instructionally beneficial to teachers and students. Through our ongoing efforts in research and development, we are further exploring how the approach can be used to develop robust classroom-based assessment tasks and resources that can help teachers better monitor students' developing proficiencies in science and literacy, which are foundational to future learning and school performance.

REFERENCES

Alozie, N., Haugabook Pennock, P., Madden, K., Zaidi, S., Harris, C. J., & Krajcik, J. S. (2018, March). *Designing and developing NGSS-aligned formative assessment tasks to promote equity*. Paper presented at the annual conference of National Association for Research in Science Teaching, Atlanta, GA.

Bailey, A. L., Butler, F. A., Stevens, R., & Lord, C. (2007). Further specifying the language demands of school. In A. L Bailey (Ed.) *The language demands of school: Putting academic English to the test* (pp. 103-156). New Haven, CT: Yale University Press.

DeBarger, A. H., Penuel, W. R., Harris, C. J., & Kennedy, C. A. (2016). Building an assessment argument to design and use next generation science assessments in efficacy studies of curriculum interventions. *American Journal of Evaluation*, *37*(2), 174–192.

Furtak, E. M. (2017). Confronting dilemmas posed by three-dimensional classroom assessment: Introduction to a virtual issue of Science Education. *Science Education*, *101*(5), 854–867.

Gorin, J. S., & Mislevy, R. J. (2013, September). *Inherent measurement challenges in the Next Generation Science Standards for both formative and summative assessment*. Paper presented at the Invitational Research Symposium on Science Assessment Washington, DC.

Halliday, M. A. K. (1993). Towards a language-based theory of learning. *Linguistics and Education*, *5*, 93–116.

Harris, C. J., Krajcik, J. S., Pellegrino, J. W., & DeBarger, A. H. (2019). Designing knowledge-in-use assessments to promote deeper learning. *Educational Measurement: Issues and Practice*, *38*(2), 53–67.

Harris, C. J., Krajcik, J. S., Pellegrino, J. W., McElhaney, K. W., Pennock, P., & Gane, B. D. (2018, March). *Designing classroom-based assessments for supporting three-dimensional teaching and learning.* Paper presented at the NARST 2018 Annual International Conference, Atlanta, GA.

Hart, J. E., & Lee, O. (2003). Teacher professional development to improve the science and literacy achievement of English language learners. *Bilingual Research Journal*, *27*(3), 475–501.

Hobbs, M. E., Williams, R. A., & Sherwood, E. A. (2012). Collaborating with teacher researchers to study what young children know and can do in science. *Voices of Practitioners, 7*.

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

Mislevy, R. J., & Riconscente, M. M. (2006). Evidence-centered assessment design: Layers, concepts, and terminology. In S. Downing & T. Haladyna (Eds.), *Handbook of test development* (pp. 61- 90). Mahwah, NJ: Lawrence Erlbaum.

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

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

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

Pearson, P. D., & Billman, A. K. (2016). Reading to learn science: A right that extends to every reader-expert or novice. In Z. Babaci-Wilhite (Ed.), *Human rights in language and STEM education*. Boston: Sense.

Pearson, P. D., Moje, E., & Greenleaf, C. (2010). Literacy and science: Each in the service of the other. *Science*, *328*(5977), 459–463.

Pellegrino, J. W. (2013). Proficiency in science: Assessment challenges and opportunities. *Science*, *340*(6130), 320–323.

Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, *44*(1), 57–84.

Samarapungavan, A., Mantzicopoulos, P., & Patrick, H. (2008). Learning science through inquiry in kindergarten. *Science Education*, *92*(5), 868–908.

Shepard, L. A. (1994). The challenges of assessing young children appropriately. *The Phi Delta Kappan*, *76*(3), 206–212.

Snow, C. E., & Van Hemel, S. B. (2008). *Early childhood assessment: Why, what and how*. Report of the Committee on Developmental Outcomes and Assessment for Young Children for the National Research Council. Washington, DC: The National Academies Press.

Stoddart, T., Pinal, A., Latzke, M., & Canaday, D. (2002). Integrating inquiry science and language development for English language learners. *Journal of Research in Science Teaching*, *39*(8), 664–687.

Wells, G. (1994). The complementary contributions of Halliday and Vygotsky to a "language-based theory of learning." *Linguistics and Education*, *6*, 41–90.

AUTHORS

Alison K. Billman (corresponding author) is Director of Early Elementary Curriculum with the Learning Design Group at The Lawrence Hall of Science at the University of California, Berkeley (alison_billman@berkeley.edu).

Daisy Rutstein is Principal Researcher at SRI International and specializes in the application of the evidence-centered design process (daisy.rutstein@sri.com).

Christopher Harris is Senior Director of Science and Engineering Education Research at WestEd (christopher.harris@wested.org).

CONTRIBUTING MEMBERS OF THE NGSA-YS TEAM

We gratefully acknowledge the partnership and contributions of our colleagues in this research and development work: **Lauren Brodsky**, Assessment Lead, and **Meredith W. Moran**, Curriculum and Professional Learning Specialist, at The Lawrence Hall of Science; **Nonye Alozie**, Senior Education Researcher at SRI International; and **Gary Weiser**, Science Education Research Associate at WestEd.

Next Generation Assessments FOR Young Scientists

SPECIAL THANKS

We extend a special thank you to the teachers and students who so graciously allowed us to try out our assessment tasks with them. With their help, we were able to gain greater insight into how tasks can be designed for young students and made usable for teachers.

SUGGESTED CITATION

Billman, A. K., Rutstein, D., and Harris, C. J. (2021). Articulating a transformative approach for designing tasks that measure young learners' developing proficiencies in integrated science and literacy. Berkeley, CA: The Lawrence Hall of Science, University of California, Berkeley.

The Next Generation Science Assessments for Young Scientists project was supported with funding from the National Science Foundation (NSF grant numbers 1853923, 1853927, and 1853951) and was based in part on the prior work of the Next Generation Science Assessment project (NSF grant numbers 1316874, 1316903, 1316908, and 1903103). Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily represent the views of their institutions or the National Science Foundation.

IN PARTNERSHIP WITH

The Next Generation Science Assessments for Young Scientists project was a collaborative effort between The Lawrence Hall of Science, University of California, Berkeley; SRI International; and WestEd.

The Lawrence Hall of Science

SRI Education™
A DIVISION OF SRI INTERNATIONAL

WestEd
WestEd.org

Our Mission: Discover. Engage. Innovate.

To inspire and foster learning of science and mathematics for all, especially those who have limited access to science.

lawrencehallofscience.org