Measuring the Power of Learning.®

Research Report
ETS RR–17-43

Practice-Based Measures of Elementary Science Teachers' Content Knowledge for Teaching: Initial Item Development and Validity Evidence

Jamie N. Mikeska

Geoffrey Phelps

Andrew J. Croft

September 2017

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

Practice-Based Measures of Elementary Science Teachers' Content Knowledge for Teaching: Initial Item Development and Validity Evidence

Jamie N. Mikeska, Geoffrey Phelps, & Andrew J. Croft

Educational Testing Service, Princeton, NJ

This report describes efforts by a group of science teachers, teacher educators, researchers, and content specialists to conceptualize, develop, and pilot practice-based assessment items designed to measure elementary science teachers' content knowledge for teaching (CKT). The report documents the framework used to specify the content-specific teaching practices and instructional tools that are critical to the work that elementary science teachers engage in with students, curriculum, and instruction. Drawing on this framework, the report details the development process for practice-based assessment items designed to measure CKT elementary science in three content areas: (a) structure and properties of matter, (b) ecosystems, and (c) Earth's place in the universe. These practice-based assessment items address the various content challenges elementary science teachers face in their work and were designed to be used as a foundation for building large-scale assessments of elementary science teachers' CKT science. This report presents initial validity evidence examining the practice-based CKT science item characteristics using results from online administrations of these items to 250 upper elementary science teachers. Findings reveal that the majority of these new assessment items capture variability in elementary science teachers' performance and that a large proportion of the items differentiate moderately well, supporting a beginning proof-of-concept for the conceptualization and design of these practice-based CKT elementary science assessment items.

Keywords Science; assessment; content knowledge for teaching; specialized content knowledge; pedagogical content knowledge; subject matter knowledge; elementary teachers

doi:10.1002/ets2.12168

Enhancing the capacity of K–12 teachers is paramount to address the new vision for science education being promoted by the K–12 Framework for Science Education (National Research Council, 2012) and the Next Generation Science Standards (NGSS; NGSS Lead States, 2013). This new vision highlights the importance of science instruction that reflects the interconnected nature of science within the real world by engaging students in learning that simultaneously addresses disciplinary core ideas, scientific and engineering practices, and crosscutting concepts. To enact high-quality science instruction, science teachers must have access to the full range of content knowledge necessary to engage successfully in critical science teaching practices (Kloser, 2014; National Research Council, 2007; Windschitl, Thompson, Braaten, & Stroupe, 2012), such as eliciting and using students' ideas about science, engaging students in scientific investigations, and connecting science concepts to phenomena and applications.

As conceptualized, content knowledge for teaching (CKT) includes multiple forms of knowledge, such as an in-depth knowledge of the subject matter at particular grade bands and specialized knowledge "tailored to the work that teachers do with curriculum, instruction, and students" (Ball, Hill, & Bass, 2005, p. 16). However, the field currently has few measures that can be used to determine whether science teachers have the relevant CKT needed to effectively teach particular topics at various grade levels (Minner, Martinez, & Freeman, 2012; National Research Council, 2013; Wilson, 2016). This has resulted in national calls to develop assessments of CKT in science (National Research Council, 2013; Wilson, 2016). This report describes our efforts to develop and investigate the statistical properties, including item difficulty and item discrimination, of a set of practice-based CKT elementary science assessment items. In addition, this report examines how teachers' performances on these practice-based CKT science assessment items vary across topics. Finally, the actual items that were developed, along with detailed explanations of how they function to assess CKT science, are provided.

Corresponding author: J. N. Mikeska, E-mail: jmikeska@ets.org

This provides an important resource for researchers and assessment developers in understanding both the potential and the challenges involved in assessing CKT science.

In this report, we begin by situating this development and research work in current efforts to conceptualize and measure science teachers' knowledge. We then share the framework we used and describe the processes we employed to develop these practice-based items designed to measure elementary science teachers' CKT. Next, we describe the study's sample and the online administration procedures for piloting these items. In the results section, we report item- and scale-level statistics from the online administration of these practice-based CKT science items with 250 upper elementary science teachers. We end by suggesting next steps to extend the scope of this exploratory research.

Conceptualizing Practice-Based Measures of Content Knowledge for Teaching Science

Practice-based views of teacher learning and development have long dominated discussions about the structure and components of science teachers' professional knowledge base, resulting in a number of different conceptualizations of the CKT[1] that science teachers apply in their classrooms (Gess-Newsome, 2015; Magnusson, Krajcik, & Borko, 1999; Park & Oliver, 2008; Park & Suh, 2015). By *practice based*, we refer to the ways in which science teachers deploy and use their subject-specific knowledge as they engage in the work of teaching science. The work of teaching science can occur both within and outside the school building as teachers interact with students during instruction, engage in planning for and reflect on instructional episodes, and communicate and collaborate with parents and colleagues. To positively impact students' science learning, research has suggested that science teachers are called on to use their subject matter knowledge in a wide range of teaching practices across varied contexts (Kloser, 2014; National Research Council, 2007; Windschitl et al., 2012). Perhaps two of the most critical practices for supporting students' learning in science involve the ability to (a) attend to and use students' ideas and experiences as the basis for learning and (b) evaluate and select instructional strategies and resources for classroom use—both of which place substantial demands on teachers' subject matter knowledge.

Studies have suggested that science teachers who are adept at recognizing, understanding, and leveraging students' ideas are better able to support students in building their understanding of key scientific concepts (Avraamidou & Zembal-Saul, 2005; Herrenkohl & Guerra, 1998; Hogan, Natasi, & Pressley, 2000; Ruiz-Primo & Furtak, 2006; Zembal-Saul, Krajcik, & Blumenfeld, 2002). Science teachers are constantly confronted with students' ideas about scientific phenomenon, including alternative conceptions and misconceptions, and must draw on their understanding of the subject matter to interpret students' ideas and probe for understanding (Coffey, Hammer, Levin, & Grant, 2011; Forbes, Sabel, & Biggers, 2015; Levin, 2013). Research has suggested that this aspect of the work of teaching science is neither intuitive nor straightforward (Nilsson & van Driel, 2010; Otero & Nathan, 2008). Novice science teachers tend to be unaware of students' ideas; when they do start to notice students' ideas, they tend to focus on more basic features like the concreteness or accuracy of these ideas (Schneider & Plasman, 2011). As science teachers gain experience, they are more likely to begin to consider how to build on students' prior ideas during instruction, how to link students' ideas to scientific ideas, and why students might find certain ideas conceptually difficult to learn—all significant content challenges that arise in the course of teaching in this area (Meyer, 2004; Schneider & Plasman, 2011).

To develop students' ideas about key scientific concepts, research has also pointed to the importance of determining which instructional strategies and resources would be most beneficial for addressing specific student outcomes; this selection usually occurs through the critical analysis and use of curriculum materials (Davis, 2006; Davis & Smithey, 2009). In particular, research has suggested the importance of ensuring that science teachers can fluently discern which instructional resources to use to target specific learning goals, to provide insight into student thinking, and to address specific student challenges (Carlsen, 1991; Schneider & Plasman, 2011). One of the most important aspects of this practice for a science teacher is being able to use one's subject matter knowledge to determine whether and how specific instructional strategies and resources can be employed to engage students productively with scientific phenomena, although this can be challenging, especially for novice teachers (Carlsen, 1991; Davis, Petish, & Smithey, 2006; Schneider & Plasman, 2011). It is important that science teachers understand the different ways in which scientific phenomena can be represented, for example, through a wide array of models, examples, or simulations, and be able to use these various representations to challenge student ideas (Clermont, Borko, & Krajcik, 1994; Schwarz et al., 2009). Finally, science teachers are called on to use instructional resources to create coherent content story lines, which has been linked to increased student learning (Roth et al., 2011).

In our work, we leverage these research findings to propose a framework, which is shown in Table 1, for designing measures of elementary science teachers' CKT. This framework is organized around the most important content-intensive practices, or tasks of teaching, that elementary science teachers engage in both inside and outside the classroom to support their students' learning. It highlights the various instructional tools and resources, such as scientific models, explanations, and students' ideas, that science teachers interact with on a daily basis and foregrounds the application of content knowledge as they interact with these tools.

The theory of CKT (Ball, Thames, & Phelps, 2008; Hill, Ball, & Schilling, 2008; Hill, Dean, & Goffney, 2007; Hill, Schilling, & Ball, 2004) undergirds this framework. This theory is one of knowledge use and asserts that the professional knowledge base for teaching goes beyond merely knowing the subject matter and includes forms of professional knowledge that are unique to the content challenges that teachers face in their daily work. For example, researchers have suggested that specialized content knowledge includes forms of content knowledge only used in teaching (Hill et al., 2008; Hill et al., 2007; Hill et al., 2004; Kersting, 2008; Krauss, Baumert, & Blum, 2008; Phelps, Weren, Croft, & Gitomer, 2014), whereas pedagogical content knowledge targets the knowledge demands of teaching that occur at the intersection of the discipline, students, and instruction (Shulman, 1986, 1987). These aspects of the professional knowledge base for science teaching include such practice-based tasks as determining whether a student's unconventional scientific explanation is conceptually valid, recognizing common student misconceptions, and knowing which content representations or instructional activities are most likely to support students' science learning.

As this CKT tasks of teaching framework suggests (see Table 1), defining teachers' professional knowledge base requires close attention to the ways in which teachers engage in the content-specific aspects of their work and the knowledge that they draw on when doing so. Each of these tasks of teaching requires science teachers to leverage their conventional subject matter knowledge, defined as their knowledge of core disciplinary facts and concepts (Schwab, 1978), that is the main focus of both K–12 instruction and college courses and is used in common across a wide variety of settings or professions. This conventional subject matter knowledge has been the focus of most traditional content assessments whereby teachers are called on to do the work of the K–12 student curriculum that they are expected to teach (Minner et al., 2012; National Research Council, 2013)—as denoted in the last task of teaching in this framework. Yet, science teaching often requires teachers to use unique types of professional content knowledge that go beyond doing the work of the student-level curriculum.

The other seven tasks of teaching shown in this framework identify instructional practices that science teachers engage in on a daily basis using their CKT. However, few practice-based assessments are designed to measure the professional knowledge that science teachers are called on to use as they engage in these tasks of teaching. The practice-based CKT assessments that measure the knowledge used in teaching science focus mainly on science teachers' pedagogical content knowledge (Lee, Brown, Luft, & Roehrig, 2007) and involve teachers in tasks that require substantial time to administer and score, such as analyzing video-recorded episodes of practice (Roth et al., 2011), participating in think-aloud or cognitive interviews (Henze & van Driel, 2015; Park & Suh, 2015), conducting classroom observations (Park & Oliver, 2008; Park & Suh, 2015), or documenting their pedagogical content knowledge for particular science topics using graphic organizers (Bertram & Loughran, 2012; Loughran, Mulhall, & Berry, 2004). To date, only a few notable efforts have developed large-scale, practice-based assessment items to measure the professional knowledge specialized to science teaching (Sadler, Sonnert, Coyle, Cook-Smith, & Miller, 2013; Smith & Taylor, 2010).

We think that this tasks of teaching framework offers a more focused conceptualization of the measurable aspects of the CKT science construct, which addresses some of the concerns in the field about the need for increased clarity regarding the nature of CKT science and what is being measured (Abell, 2008; Gess-Newsome, 1999; van Driel, Verloop, & de Vos, 1998). In our work, we focus on the conceptual challenges that elementary science teachers encounter in the work of teaching science and on developing practice-based assessment items to measure the CKT that is embedded within these recurrent tasks of teaching. Because it is in the context of teaching practice where this CKT science is used, our assessment design focused on developing practice-based items to measure the core elements of elementary science teachers' professional knowledge in ways that are useable in their work.

It is important to emphasize that the goal of this project was not to develop a full-scale assessment of CKT for each of these three science topics. Instead, our goal was to determine if these kinds of assessment questions could be developed according to a task of teaching framework, if the number of questions within each topic (e.g., matter, ecosystems, Earth's place in the universe) is sufficient for forming a reliable scale measure, and whether there is variation across items in

Table 1 Content Knowledge for Teaching Tasks of Teaching

Task of teaching	Description	Examples in science instruction
1. Anticipating student challenges, misconceptions, partial misconceptions, alternate conceptions, strengths, interests, capabilities, and background knowledge	This task of teaching describes the work of integrating knowledge about content to be learned and about students and anticipating how students are likely to interact with the content and content practices. This work is key in planning instruction, both in advance and in the moment as adjustments are made. These anticipations can become the basis for selecting appropriate explanations, examples, or tasks as instructional strategies.	• Anticipating student challenges in reasoning about and doing science due to the interplay of content demands and students' understanding • Anticipating the impact of limited English language proficiency on students' comprehension of science concepts and text and on their ability to present scientific ideas, make scientific arguments, and give explanations • Anticipating likely misconceptions, partial conceptions, and alternate conceptions about particular science content and practices • Anticipating student interest and motivation around particular science content and practices • Anticipating how students' background knowledge can interact with new science content and practices
2. Evaluating student ideas evident in work, talk, actions, and interactions	This task of teaching describes the work of making sense of things that students do, say, and produce. It can involve deciding if an idea is valid and/or if an explanation is adequate in a particular context, identifying evidence of partial or alternate understandings, or analyzing errors. Evaluating includes characterizing, analyzing, describing, and judging student work.	• Evaluating student work, talk, or actions to identify conceptions in science, including incorrect or partial conceptions • Evaluating student explanations or arguments for use of appropriate scientific practices • Evaluating nonstandard responses for evidence of scientific understanding • Evaluating student models, explanations, arguments from evidence, and investigations for evidence of understanding science concepts • Evaluating or comparing multiple approaches to a scientific investigation, model, or explanation • Evaluating discussion among groups of students for evidence of understanding science concepts and practices
3. Explaining concepts, procedures, representations, models, examples, definitions, and hypotheses	This task of teaching describes the crafting and using of appropriate explanations at any point in time when explanations are called for, including in setting purposes for instruction, planning, giving feedback, or in responding to students in the moment. Explaining also includes activities that show, such as modeling.	• Explaining scientific concepts • Explaining scientific procedures • Explaining scientific representations • Explaining scientific models • Explaining scientific examples • Explaining scientific definitions • Explaining scientific hypotheses • Explaining why a particular scientific definition, model, or representation serves a particular purpose • Interpreting a particular representation in multiple ways to further scientific understanding • Explaining why a scientific practice, process, or procedure is efficient or appropriate

Table 1 Continued

Task of teaching	Description	Examples in science instruction
4. Creating and adapting resources for instruction (examples, models, representations, explanations, definitions, hypotheses, procedures)	This task of teaching describes the work of inventing new instructional tools or adapting available ones to fit particular purposes. While curricular resources can be a source of material for adaptation, student-generated materials can also be adapted into opportunities for learning. The creation or adaptation of instructional tools is thought of as having an instructional goal, examples of which are in the subject-specific bulleted list. Examples are selected for a reason — to introduce new material, to showcase certain attributes of content, or to challenge students' current conceptions.	• Creating and adapting examples to introduce a concept or illustrate an idea or to demonstrate a scientific practice, process, or procedure • Creating and adapting examples that support particular scientific practices, processes, or procedures or to address particular student questions, misconceptions, or challenges with content • Creating and adapting representations or models to support students' scientific understanding • Creating and adapting representations or models that support multiple interpretations • Creating and adapting multiple representations or models to support students' scientific understanding across representations • Creating and adapting definitions to fit instructional purposes • Adapting student-generated hypotheses to support instructional purposes • Creating and adapting procedures for working with scientific content
5. Evaluating and selecting resources for instruction (examples, models, representations, explanations, definitions, hypotheses, procedures)	This task of teaching describes the work of evaluating instructional resources and selecting those most appropriate for a particular instructional purpose. This work can occur through advance planning or during instruction. Teachers may evaluate and select from curricular materials or from potential resources generated during the course of instruction, including work generated by students.	• Evaluating and selecting examples to introduce a concept or illustrate an idea or to demonstrate a scientific practice, process, or procedure • Evaluating and selecting examples that support particular scientific practices, processes, or procedures or to address particular student questions, misconceptions, or challenges with content • Evaluating and selecting representations or models to support students' scientific understanding • Evaluating and selecting representations or models that support multiple interpretations • Evaluating and selecting definitions to fit instructional purposes • Evaluating and selecting explanations of scientific concepts for their potential to support scientific learning • Evaluating and selecting procedures for working with scientific content • Evaluating and selecting student-generated hypotheses to support instructional purposes • Evaluating and selecting resources for their potential to support scientific learning

Table 1 Continued

Task of teaching	Description	Examples in science instruction
6. Developing questions, activities, tasks, and problems to elicit student thinking	This task of teaching includes the work of developing and using questions, tasks, and problems to engage students in developing content understanding. The development of questions, tasks, and problems is used to make student understanding visible and to build deeper and more accurate understanding.	• Creating questions, tasks, or problems with the potential to elicit student scientific thinking, discussions, justifications, or explanations • Creating or adapting questions, tasks, or problems that support particular scientific practices, processes, or procedures • Creating or adapting questions, tasks, or problems that demonstrate desired scientific characteristics • Creating or adapting investigations, issues, or applications that address the same scientific topic or concept or that vary in difficulty or complexity • Creating or adapting questions, activities, or tasks to elicit evidence that students have a particular scientific understanding or skill
7. Evaluating and selecting student tasks (questions, tasks, problems) to elicit student thinking	This task of teaching includes the work of evaluating and selecting things for students to work on or interact with. The evaluation and selection of student tasks is thought of as having an implicit instructional goal and relies on anticipations about how students will understand and interact with the tasks.	• Evaluating and selecting questions, tasks, or problems to elicit student scientific thinking, justifications, or explanations • Evaluating and selecting questions, tasks, or problems that support particular scientific practices, processes, or procedures • Evaluating and selecting questions, tasks, or problems that demonstrate desired scientific characteristics • Evaluating and selecting investigations, issues, or applications that address the same scientific topic or concept or that vary in difficulty or complexity • Evaluating and selecting questions, activities, or tasks to elicit evidence that students have a particular scientific understanding or skill
8. Doing the work of the student curriculum	This task of teaching includes doing the student-level tasks that make up the intended curriculum. Doing the student-level work is not uniquely demanded by teaching but is an essential part of the content work necessary to do in the course of teaching, especially in preparation for assigning student tasks. This work is often incorporated into or prerequisite for doing other tasks, such as anticipating student difficulties or selecting problems or tasks.	• Doing the work that will be demanded of the students as part of the intended curriculum

Note. The first two columns of this chart — tasks of teaching and description — have been reprinted with permission from “Developing Content Knowledge for Teaching Assessments for the Measures of Effective Teaching Study” by G. Phelps, B. Weren, A. Croft, & D. Gitomer, 2014, ETS Research Report No. RR-14-33, Princeton, NJ: ETS, 2014 their difficulty and relation to the other items that were developed. We intend these initial items to serve as models for developing comprehensive assessments of science teachers' CKT for particular science topics and grade levels.

Methods and Data Sources

Developing Practice-Based Content Knowledge for Teaching Elementary Science Questions

Assessment Design

The conceptualization and development for the practice-based CKT elementary science assessment questions occurred during an 18-month period and involved a group of educational researchers, cognitive scientists, teacher educators, test development specialists, and former K–12 science teachers. Using a process that was closely modeled after the item development work on the Measures of Effective Teaching Project (Phelps et al., 2014), our team began by first identifying the ways in which science teachers use their content knowledge in the work that they do, which resulted in the CKT tasks of teaching framework in Table 1. The major phase of assessment design began in January 2013 and continued through mid-2014. Because the NGSS had yet to be released at the beginning of the project, the development team initially consulted the National Science Education Standards (National Research Council, 1996) and the Framework for K–12 Science Education (National Research Council, 2012) to identify the key concepts that upper elementary students were expected to learn within three different science content topics: matter, ecosystems, and Earth's place in the universe. The development team targeted concepts at Grades 4, 5, and 6, following the distinction for upper elementary grades adopted by the local school district the team was working with at the time. These content topics were selected because they each address one of the three major subject matter areas in science, and they serve as critical topics in the standards within these content areas at these grade levels (National Research Council, 1996). In addition, including multiple topics provided a way to investigate the topic-specific nature of elementary teachers' CKT science.

To begin, the development team generated a list of scenarios in which upper elementary teachers face content challenges when teaching science in these topic areas. These scenarios directly mapped onto the first seven content-intensive practices described in the CKT tasks of teaching framework. For example, when teaching about the transfer of matter among plants, animals, decomposers, and the environment, elementary teachers need to anticipate a number of common student misconceptions (see Table 1, Task of Teaching 1), such as the idea that plants obtain their food from the soil or the idea that changes in one population in a food web only affect populations that are directly connected to them in the food web. In addition, the development team also leveraged current research findings, especially the literature on learning progressions, and team members' collective knowledge about the cognitive work that elementary teachers engage in when teaching these particular topics in order to generate ideas for these practice-based CKT science questions. For example, when teaching about ecosystems at the upper elementary level, it is common practice to use food webs to model the interdependence among producers, consumers, and decomposers in an ecosystem. To do so, teachers need to make decisions about what types of food webs to use (see Table 1, Task of Teaching 5) and face challenges in evaluating students' conceptions about these representations (see Table 1, Task of Teaching 2). After the NGSS were released, we identified which performance expectation best mapped onto each of the newly developed practice-based CKT science questions.

Our development efforts resulted in a set of questions designed to assess aspects of upper elementary science teachers' CKT science across three science domains: (a) physical sciences, including matter and its interactions, with a focus on weight, volume, and density (11 questions); (b) life sciences, including ecosystems interactions, energy, and dynamics, with a focus on food webs (14 questions); and (3) Earth and space sciences, including Earth's place in the universe, with a focus on patterns created by motions of the Moon, Earth, and Sun (14 questions). Each question set was not designed to assess the full range of CKT required to teach each of these topics at the upper elementary grade level. Instead, we identified a subset of key ideas and concepts in each of these selected content areas to focus initial development work and set out to determine if these types of practice-based CKT questions could be developed in a new content area using a task of teaching framework, leveraging work that had been completed in other disciplines (Phelps et al., 2014). Because this development work was exploratory in nature, the team decided to use a variety of different question formats, including single-selection multiple-choice questions, constructed-response questions, table questions, and other types of selected-response questions (e.g., matching, select all that apply), that had been used previously to develop practice-based CKT questions in elementary English language arts and mathematics.[2]

Table 2 Next Generation Science Standards (NGSS) Addressed by the Practice-Based Content Knowledge for Teaching Questions

Topic	NGSS performance expectations addressed	Disciplinary core ideas addressed	Crosscutting concepts addressed	Scientific practices addressed
Earth's place in the universe	(5-ESS1–2) Represent data in graphical displays to reveal patterns of daily changes in length and direction of shadows, day and night, and the seasonal appearance of some stars in the night sky	(ESS1.A) The Universe and its Stars; (ESS1.B) Earth and the Solar System	Patterns; scale, proportion, and quantity	Analyzing and interpreting data; engaging in argument from evidence
Ecosystems	(5-LS2–1) Develop a model to describe the movement of matter among plants, animals, decomposers, and the environment	(LS2.A) Interdependent Relationships in Ecosystems (LS2.B) Cycles of Matter and Energy Transfer in Ecosystems	Systems and system models	Developing and using models
Matter	(5-PS1–3) Make observations and measurements to identify given materials based on their properties (5-PS1–3) Develop a model to describe that matter is made of particles too small to be seen	(PS1.A) Structure and Properties of Matter	Scale, proportion, and quantity	Developing and using models; planning and carrying out investigations; using mathematics and computational thinking

In the pages that follow, we provide examples of four different question formats designed to assess elementary science teachers' CKT. As illustrated by these practice-based CKT questions, the emphasis in the CKT design theory is on the application of content knowledge as it is used by teachers to address problems of practice that they encounter in their daily work. Each practice-based CKT science question addresses one of the first seven tasks of teaching from the CKT tasks of teaching assessment framework and one of the performance expectations from the NGSS. Table 2 provides a listing of the NGSS performance expectations, disciplinary core ideas, crosscutting concepts, and scientific practices addressed within these practice-based CKT questions.

Figure 1 shows an example of a question we developed to assess elementary teachers' CKT about the properties of matter. In this question, the teacher is called on to evaluate and select a set of instructional resources. Specifically, the task requires a teacher to select which of a pair of objects would be most useful in identifying students who confuse the relationship between weight and volume. A teacher would need to evaluate each set of objects given in the answer choices to determine whether the set provides an opportunity to reveal students' naive conceptions about the relationship between weight and volume. Teachers need to go beyond determining which object in each set weighs more or has the greater volume and use their knowledge of these two properties of matter to evaluate the objects as instructional resources. In particular, teachers need to recognize and draw on their knowledge of the common misconception that students have when they use the visual cue of size as a proxy for weight. With this knowledge, teachers need to determine that students with this misconception would most likely select an object that has a larger volume as the one that also weighs more, regardless of its composition. Thus teachers have to determine which pair of objects could be used to identify students who would *incorrectly* predict that the object with the larger volume is also the object that weighs more. The only option in which this outcome is a possibility is Option B, where an object that takes up more volume than the other weighs less because it is made of wood as opposed to aluminum.[3] Because this question requires a respondent to select one of the four options provided, this question counts as one item in the final scoring model.

At the beginning of a lesson on the relationship between the weight and volume of common solid objects, Ms. Zhang gives her students some prediction problems to check their current understanding. Before handling and weighing a pair of solid objects, the students are to predict (just by looking at them) which of the two objects weighs more. Of the following pairs of objects, which would be most useful in determining the students who confuse the relationship between weight and volume?	A. An aluminum sphere with a 7 cm diameter and a wood sphere with a 4 cm diameter B. An aluminum square pyramid with 4 cm sides and a 4 cm height and a wood square pyramid with 4 cm sides and 7 cm height C. A wood cube with 4 cm sides and a wood cube with 7 cm sides D. All of the above are equally useful in determining the students who confuse the relationship between weight and volume.

Figure 1 Example of a single-selection multiple-choice question.

The students in Ms. Colt's fifth-grade class have been conducting multiple controlled experiments to learn about plant growth. Through their experiments, they found that plants do not continue to grow without sunlight, water, or air. Ms. Colt explained the experiments provide evidence that plants use sunlight, water, and air to make their own food. Some of the students wondered about how fertilizer helps plants grow. To examine the effects of fertilizer on plant growth, the students have just conducted another investigation to compare plants grown with and without fertilizer. The students observed that the plants grown with fertilizer looked very healthy with green leaves and thick, strong stems while the plants grown without fertilizer had leaves that were turning yellow and thin, weak stems. Since the fertilizer helped the plants grow better, some students have concluded that fertilizer is also food for plants. During the next few lessons, Ms. Colt will have her students further investigate plant growth, and she plans on explaining why fertilizer does help plants grow but is not considered food for plants. Write a brief explanation that the teacher could give as part of her next lesson to explain why fertilizer does help plants grow but is not considered food for plants. The explanation should address the underlying science and be accessible to fifth-grade students.

Figure 2 Example of a constructed-response question.

Figure 2 shows an example of a constructed-response question we developed to assess elementary teachers' CKT about ecosystems. This question requires the teacher to craft an appropriate explanation to develop students' conceptual understanding. In particular, the task is to write a brief explanation that a fifth-grade teacher could give as part of his or her next lesson to explain why fertilizer helps plants grow but is not food for plants. The explanation must address the relevant scientific concepts and use language that is appropriate for the specified grade level without compromising the underlying scientific concepts. To be successful, a teacher would need to understand the difference between food for plants (energy-providing substances that plants produce) and fertilizer (minerals that do not provide energy for plant growth and development) and use this knowledge, as well as knowledge about what is accessible to students at this grade level, to generate a coherent and useful explanation. Because these questions required additional effort and time to complete compared to a single-select multiple-choice question, each constructed-response question counted as two items in the final scoring model.

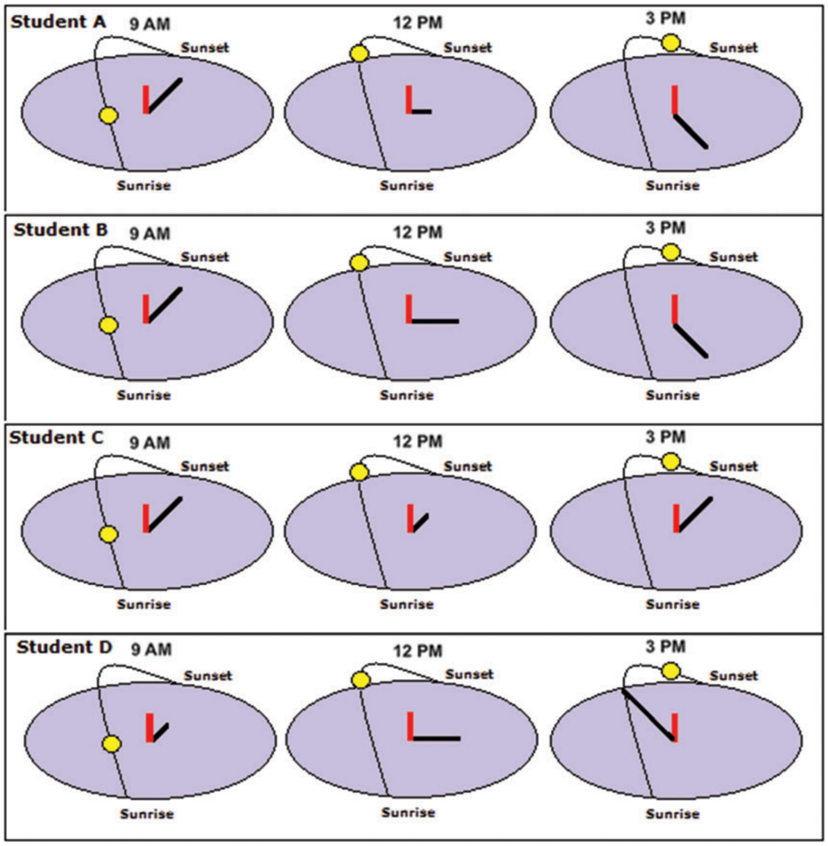
Figure 3 provides an example of a table question we used for assessing elementary teachers' CKT about Earth's place in the universe. In table questions, the teacher is provided with one set of directions or criteria and asked to select one response for each row within the table. In this table question, the teacher is engaged in evaluating the accuracy of students' ideas as evidenced in their written work. The task is to evaluate which, if any, of the student responses provides evidence that the student understands each concept indicated. To do this, the teacher must interpret and evaluate each student response individually to determine if it shows that the student understands each of the concepts given. Not only does the teacher need to understand when a shadow is created and how shadows change but he or she must also determine what to look for in the diagrams to provide evidence of understanding for each concept. Because each table question requires teachers to separately consider and mark a response for each row, we count each row as a separate item when scoring teachers' responses. Thus a table question like this one would be counted as four separate items in the scoring model.

Figure 4 provides an example of another type of selected-response question, a matching question, about CKT Earth's place in the universe. In this question, the teacher must match the feedback prompts to the appropriate student. Teachers must apply their knowledge of the different Moon phases to evaluate each student's written diagram. First, the teacher must determine whether the student's diagram correctly shows the alignment of Earth, the Moon, and the Sun during the

For the last few weeks, students in Mr. Yamamoto's class have been investigating the relationship between the Sun's position in the sky and the length and position of shadows on the ground. In order to check for student understanding near the end of the lesson series, Mr. Yamamoto asks his students to complete an exit ticket. In this short assessment, Mr. Yamamoto has his students use Sun path diagrams to draw the relative lengths and positions of the shadows that would be cast by the Sun at 9 A.M., 12 P.M., and 3 P.M.

In the Sun path diagrams below, the red line represents a meter stick, and the black line represents the shadow cast by the Sun. In addition, each diagram has a line representing the Sun's apparent path from sunrise to sunset and a small yellow circle representing the Sun. The yellow circle (the Sun) is at a precise place along the line to show where the Sun is located in the sky at a specific time.

The Sun path diagrams below contain the drawings of four students.



In the chart below, select the student responses, if any, that provide evidence that the student understands the concept indicated. For each concept, select all that apply. Evaluate each concept separately.

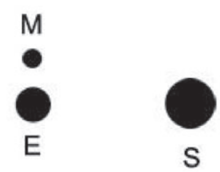
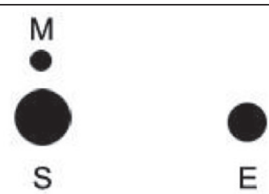
Concepts	Student Responses				
	A	B	C	D	None of the responses
1) Shadows' lengths change as the Sun's position changes.					
2) Shadows are longer when the Sun is positioned lower in the sky.					
3) The shadows cast by the Sun are in the opposite direction from the Sun's position in the sky.					
4) Patterns in the daily changes to the length and direction of shadows are caused by Earth's rotation.					

Figure 3 Example of a table question.

Mr. Brenner's class has been investigating the basic lunar phases. To check for understanding, he asks his students to complete the following exit ticket at the end of class.

> Draw a diagram that shows the positions of Earth (E), the Moon (M), and the Sun (S) during the first quarter Moon phase.

The diagrams from four of Mr. Brenner's students are below.

Student	**First Quarter Moon Phase**
Jana	
Tia	
Lucius	
Marcus	

After seeing the students' diagrams, Mr. Brenner gives each student a feedback prompt to address what is shown in the student's response. Match each feedback prompt with the student for whom it is most likely to be helpful. Each student should be matched to one of the feedback prompts.

Feedback Prompt	**Student**
"Why can we not see the Moon during the new Moon phase?"	
"Why do we see the whole Moon disc lit up during the full Moon phase?"	
"Now can you diagram the positions of Earth, the Moon, and the Sun during the third quarter Moon phase?"	
"What object in the sky does the Moon revolve around?"	

Figure 4 Example of a matching question.

first quarter Moon phase and, if not, determine what specific error(s) the student made in his or her diagram. Second, and most importantly, teachers also have to determine which feedback prompt is most helpful for the teacher to use for each student to move his or her thinking forward. They come to this determination by making the connection between the focus of the feedback prompt and the students' current understanding (or lack of understanding) revealed in the diagram that each student produced. Like table questions, each "match" counts as a separate item in the scoring model. This question, therefore, includes four possible matches and would count as four items in the scoring model.

As noted in the preceding examples, each practice-based CKT question included one or multiple items and therefore counted as one or multiple points in the final scoring model. In general, questions that required additional time or effort to complete, such as constructed-response questions, were given more weight in the final scoring model. Experienced elementary teachers, science teacher education researchers, and science assessment specialists reviewed the questions for technical merit and content accuracy prior to the online administration. Our research team also created a task design rationale to specify how knowledgeable respondents are expected to reason through each question. Appendix A includes an example of a task design rationale for the practice-based CKT table question in Figure 3. Because a main goal of this work is to better understand and illustrate the key characteristics of the practice-based items designed to measure CKT science, we have included all of the items in Appendices B, C, and D that were developed across these three science topics and used in the online assessment administration.

Question Piloting

Prior to administering these practice-based CKT science questions in the online survey, our research team conducted a small-scale pilot to gather initial validity evidence to ensure that the questions were accurately keyed and required respondents to deploy the kind of professional knowledge intended (Kane, 2006). We used this evidence to identify any questions that did not function as intended and to revise those questions prior to use with a large sample of elementary teachers. Our research team piloted these newly developed practice-based CKT science questions with a sample of 16 experienced upper elementary science teachers (four fourth grade, eight fifth grade, and four sixth grade). In the piloting, we engaged teachers in cognitive interviews using a concurrent think-aloud methodology (Ericsson & Simon, 1984) to elicit teachers' knowledge and reasoning as they responded to a subset of these new assessment questions. This methodology allowed us to learn about the decision processes that the teachers made and the knowledge that the teachers used as they worked through each question. In addition, we asked follow-up probing questions to determine if the teachers thought that the questions were clear, comprehensible, and connected to their teaching practice and, if not, the reasons for any difficulties they encountered. For each question, our goal was to have a clear sense of what task the teachers believed they were being asked to do in the assessment question, whether this was the same or different from what was intended and delineated in the task design rationale, why they selected and eliminated particular answers, and what knowledge and reasoning they used in this process.

We analyzed the cognitive interview data for patterns in the knowledge and reasoning the teachers used when responding to these questions. Using the task design rationale, which specifies the knowledge that each question is intended to measure and the ways in which teachers are expected to reason through each one, we conducted a detailed comparison of how teachers' answer choices and their justifications of these selections conformed with the intended function of each question as hypothesized in the task design rationale (Gitomer, Phelps, Weren, Howell, & Croft, 2014; Howell, Phelps, Croft, Kirui, & Gitomer, 2013). The conformity test involved coding teachers' think-aloud responses for evidence that they were situating their reasoning in the specific task of teaching and relying on the appropriate combination of CKT to select their answers.

Teachers' think-aloud comments could be coded either as "conforming to the task design rationale" or as "not conforming to the task design rationale," depending on the knowledge and skills evident in their responses. For example, for a teacher's think-aloud response on the Confusing Weight and Volume question (Figure 1) to conform to the task design rationale, a teacher would need to use his or her CKT to understand aspects of this instructional situation and apply his or her subject matter knowledge in this task of teaching. In particular, a teacher would need to (a) determine which object in each pair has a greater volume or weighs more by leveraging his or her knowledge about how the weight of an object is a function of the material it is made of and its volume, (b) understand what it means for an example to be useful in this context (in this case, usefulness refers to the ability to reveal, not obscure, student misconceptions), (c) recognize that students tend to use visual cues as a proxy for weight, and (d) evaluate each set of

Table 3 Comparison of Teachers' Think-Aloud Responses to Task Design Rationales

	Assessment percent			
	Matter ($n = 53$) %	Ecosystems ($n = 56$) %	Earth's place ($n = 53$) %	Total ($n = 162$) %
Response conformed to TDR				
Correct answer	43	34	62	46
Incorrect answer	40	48	26	38
Response did not conform to TDR				
Correct answer	8	14	9	10
Incorrect answer	9	4	2	5

Note. The n reported here refers to the number of individual item-level teacher think-aloud responses. Some columns do not add to 100% due to rounding. TDR = task design rationale.

objects provided in the answer choices to determine which pair of objects could be used to identify a student who would *incorrectly* predict that the object with the largest volume is also the one that weighs more, which would provide a means through which a teacher could determine that the student is confusing the relationship between weight and volume.

If the teacher's think-aloud comments provided evidence of these key knowledge and skills, then we coded the teacher's response as conforming to the task design rationale. However, if the teacher's think-aloud comments revealed critical errors in his or her reasoning in relationship to one or more of these key knowledge and skills, then we coded the teacher's response as not conforming to the task design rationale. For example, for the Confusing Weight and Volume question in Figure 1, one teacher's response showed that she did not clearly understand what it meant for an instructional resource—in this case, the pair of solid objects—to be most useful. This teacher's more general sense of most useful as meaning "best example ... most productive" did not tie back to the usefulness of the instructional resources for identifying students who held a particular misconception. Alternatively, a teacher who did answer this question correctly recognized that there "are a couple of variables ... size and material it is made out of" and that "most useful would be the model that would most clearly help you see which students have an understanding of the difference between weight and volume ... one that most clearly isolates the specific understanding and not other things."

In general, pilot findings revealed strong evidence that the questions functioned as designed. Teachers' cognitive interview responses provided confirmatory evidence showing that the majority of the practice-based CKT science questions required teachers to apply their content knowledge to particular tasks of teaching in ways that were hypothesized in the task design rationale. Table 3 highlights the results from the cognitive interviews and identifies the percentage of teachers' responses that did or did not conform to the task design rationale. Overall, 85% of teachers' responses conformed to the task design rationale, and this pattern of the majority of teachers' responses conforming to the task design rationale remained consistent for questions within each of the three topics. Conforming to the task design rationale meant that teachers with correct responses leveraged the relevant CKT needed to answer each question, whereas teachers with incorrect responses revealed critical errors in their CKT. The primary reasons that teachers struggled to answer some of these practice-based CKT questions correctly were difficulty with understanding the work of teaching they had to engage in, challenges with understanding key aspects of the teaching scenario, and lack of relevant content understanding.

On the basis of the pilot study results, we made revisions to questions to improve clarity and to account for potential sources of construct-irrelevant variance (Kane, 2006; Messick, 1989). For example, for the Confusing Weight and Volume question in Figure 1, originally the question had one object composed of wood and the other object composed of foam. However, one teacher eliminated the correct answer because she explained how it would be unrealistic to think that students would predict that the foam object weighed more, even if students generally used visual size as a proxy for weight, due to students' previous experience handling foam. As a result, we decided to select two materials that would result in solid objects closer in weight so that a teacher with the relevant CKT would not eliminate the correct answer for this reason. Prior to the online survey administration of these practice-based CKT questions, we also decided to remove seven questions from the larger pool due to item design flaws (questions that did not require CKT to answer correctly) or variants (questions with two similar versions). Table 4 provides information about the number

Table 4 Number and Types of Practice-Based Content Knowledge for Teaching Questions and Items Administered by Topic

	Questions (items)		
	Matter	Ecosystems	Earth's place
Multiple choice	7 (7)	8 (8)	3 (3)
Constructed response	1 (2)	1 (2)	2 (3)
Table	3 (10)	2 (10)	3 (13)
Other selected response	0	0	3 (12)
Total	11 (19)	11 (20)	11 (31)

Note. Each response to a selected-response question is considered an item. Therefore a single-selection multiple-choice question corresponds to a single item, but other selected-response questions with different formats may correspond to several items (e.g., one item for each row in a table question, as seen in Figure 3). Three of the four constructed-response questions were considered as two items because they were scored separately for accuracy and accessibility.

and types of practice-based CKT questions and items used in the online survey assessments across the three topic areas.

Sample

For this study, we recruited upper elementary teachers in Grades 4–6 who varied in science teaching experience, grade levels taught, and teaching contexts. Our recruitment strategy included two parts. First, we wanted to ensure that our sample of teachers included a proportion of elementary science teachers who would have high levels of the CKT we were trying to assess. To do so, we used our professional connections with science education researchers, professional development facilitators, and local school districts to recruit teachers who had taught these topics for numerous years and/or participated in rigorous science professional development on one or more of these topics. Second, to meet our targeted recruitment sample size and to increase the diversity of our sample, we recruited more broadly using a national database of elementary science teachers.

Our recruitment efforts yielded 516 interested upper elementary teachers who had taught science in one or more of these topic areas at one or more of these grade levels. Due to limited project resources, we had to select a subset of these teachers to participate in the study. We stratified the interested sample based on the extent of their science teaching experience and professional development within each of the three topics. Then we selected participants randomly from each stratum to ensure that the final sample included teachers who had varying degrees of science teaching experience and professional development across the three science topics. Overall, we had a 63% completed response rate (250 of the 394 invited teachers completed one or more CKT question sets).

Table 5 provides demographic information about the 250 participating teachers in this study by topic. Table 6 provides detailed information about how much experience these participants had teaching at the K–8 grade level overall, teaching science at the K–8 grade level, and teaching these particular topics at the K–8 grade level. The majority of the participating teachers were White women who were currently teaching at public noncharter schools. Approximately half of these teachers had an elementary education undergraduate major, while approximately two-thirds of the sample had earned a master's degree or higher. Most of the participating teachers were currently teaching science at Grades 4, 5, or 6, although some of the teachers had moved on to other teaching positions. In general, these teachers varied from more limited (3 years or fewer) to quite extensive (more than 15 years) science teaching experience within these topic areas. Overall, the sample teacher characteristics across the three topic areas were quite consistent.

Administration Process

In this study, we administered all assessment questions and surveys within an online delivery system. Each assessment form included one set of subject matter knowledge questions,[4] one set of practice-based CKT questions, and two perceptions surveys (one following each set of assessment questions). The perceptions surveys asked teachers to evaluate the extent to which they found the questions to be challenging, important to the work teachers do, and related to their teacher

Table 5 Teacher Background Characteristics

	Assessment percent		
Demographic variable	Matter ($n = 183^a$) %	Ecosystems ($n = 205^a$) %	Earth's place ($n = 203$) %
Gender			
Female	84.2	82.4	83.7
Ethnicity[b]			
Asian	3.8	2.9	3.0
Black	5.5	4.4	4.9
Hispanic	5.5	4.9	3.9
White	86.3	87.3	87.2
Best language			
English	98.9	99.0	99.0
Undergraduate major			
Elementary education	54.3	54.1	55.7
Secondary education	3.8	3.9	3.9
Science education	4.9	5.9	5.4
Biology/life science	10.4	11.7	11.3
Physics/chemistry	2.2	2.9	3.0
Engineering	2.2	2.0	2.0
Earth/space science	0.5	0.5	0.5
Teacher preparation program			
Undergraduate	29.5	28.8	31.0
5-year	7.7	7.8	7.9
Master's degree	50.8	50.7	49.8
Alternate route	10.4	10.7	9.9
Did not attend	1.6	2.0	1.5
Educational level			
Earned bachelor's degree	31.7	33.7	34.0
Earned master's degree or higher	68.3	66.3	66.0
Certification status			
Yes	97.3	97.1	97.0
Teaching status			
Grades 1–3	1.1	2.4	2.5
Grades 4–6	90.7	89.3	89.2
Grades 7–8	1.6	1.0	1.5
Instructional coach	6.0	5.4	5.4
Other	0.5	1.0	0.5
School locale			
Urban	29.5	28.3	27.1
Suburban	53.0	55.6	56.7
Rural	17.5	16.1	16.3
School type			
Public (noncharter)	92.3	90.2	91.1
Public charter	2.7	4.4	3.4
Private	4.9	5.4	5.4
Current teaching assignment			
Elementary subject specialist	48.6	49.8	49.8
Middle school teacher	14.8	13.7	14.3
Self-contained classroom teacher	28.4	28.3	28.6
Instructional coach	6.0	5.4	5.4
Other	2.2	2.9	2.0

[a]Only teachers who completed the demographic survey are included in this table. For the matter and ecosystems topics, one teacher was missing demographic data. [b]Respondents could identify as more than one race so percentages do not add to 100.

preparation and professional development experiences. In addition, each teacher completed two additional surveys: (a) the Science Teaching Efficacy Belief instrument (Riggs & Enochs, 1990) and (b) a demographics survey about their professional, teaching, and educational background. Teachers could choose to complete one, two, or all three assessment forms, depending on their interest and availability. Teachers could complete the online survey from any remote location and were specifically directed to answer the questions on their own with no support from outside resources. To reduce cognitive burden, we administered each assessment form during one 60-minute time period. Teachers received a link to

Table 6 Number of Years K–8 Teaching Experience for Teacher Participants by Topic

	Number of years teaching experience, *n* (%)				
	<1	1–3	4–6	7–15	>15
Matter					
$n = 183$[a]					
Any subject	14 (8)	18 (10)	16 (9)	75 (41)	60 (33)
Science	17 (9)	22 (12)	26 (14)	74 (40)	44 (24)
Matter	24 (13)	40 (22)	44 (24)	59 (32)	16 (9)
Ecosystems[a]					
$n = 205$[a]					
Any subject	22 (11)	19 (9)	13 (6)	79 (39)	72 (35)
Science	23 (11)	16 (8)	26 (13)	85 (41)	55 (27)
Ecosystems	16 (8)	33 (16)	52 (25)	77 (38)	27 (13)
Earth's place					
$n = 203$					
Any subject	22 (11)	17 (8)	13 (6)	81 (40)	70 (34)
Science	24 (12)	15 (7)	25 (12)	85 (42)	54 (27)
Earth's place	30 (15)	50 (25)	38 (19)	62 (31)	23 (11)

Note. Teachers reported years of experience teaching two to three subtopics within a specific topic (e.g., ecosystems). Because these responses could vary, we used the value for the subtopic the teacher had the most experience teaching to represent experience teaching that topic. The *n* reported here is the number of teachers who reported experience teaching for the specified number of years in each category.

[a]One additional teacher who completed the assessment did not complete the demographic survey.

the next assessment form within 3 days of completing the current assessment form. The testing window remained open for approximately 8 weeks, from late 2014 through early 2015.

Procedures for Excluding Teachers' Responses From the Scored Sample

Prior to the final analysis, we reviewed teachers' responses to ensure that they met the criteria for a "good faith effort" and could be used within the analyses. Because this setting was not high stakes for these teachers, we sought to identify criteria that could be used to determine teachers who likely failed to intellectually engage with these questions and answer them to the best of their ability. In general, we aimed to remove participants whose scores might reflect other, potentially problematic aspects, such as lack of motivation or effort, in their responses, which would mean their scores represent something other than their CKT. On the basis of the analysis, we decided to remove a subset of teachers' responses from the data set due to a variety of issues that would compromise their scores, similar to the process used to exclude cases from the scored sample for the Measures of Effective Teaching study (Phelps et al., 2014). Given that this work is exploratory in nature, we sought to maximize the opportunity to analyze data that suggested a good faith effort occurred.

In particular, we identified and applied three criteria for excluding teachers' responses from the analysis. We applied the three criteria in a stepwise progression, moving from the least to the most inferential exclusion criterion in terms of suggesting a lack of good faith effort when responding to the practice-based CKT items. Although a teacher's response could meet one or more of the following criteria, Table 7 indicates the first criterion for which a teacher's response was initially excluded.

First, if a teacher failed to complete 75% or more of the assessment items in a particular set, then we removed that teacher's responses for that item set from the analyses due to insufficient response completion. Second, a subset of teachers' responses suffered from latency issues. Timing data revealed that some teachers entered their answers to a set of assessment items too quickly to have given adequate consideration to the answer choices.[5] We removed teachers' responses who completed any item set in less than the minimum latency period, which is the time it would have taken merely to read the question and click on a response without fully considering the options. Finally, a small number of teachers' responses scored at or below chance,[6] suggesting that they did not give sufficient consideration to the answer choices when responding online and might have been guessing to select their answers. Although it is possible that teachers scoring at or below chance could be systematically wrong in their responses, one of the patterns we noted was that a subset of the teachers'

Table 7 Exclusion Codes by Topic

	Topic, n (%)		
	Matter[a]	Ecosystems[b]	Earth's place[c]
Exclusion category			
Missing data >25%	9 (4)	1 (<1)	5 (2)
Insufficient latency	3 (1)	2 (<1)	4 (2)
At or below chance level	14 (7)	10 (4)	6 (3)
Final scored sample	184 (87)	206 (94)	203 (93)

[a] $n = 210$. [b] $n = 219$. [c] $n = 218$.

responses met multiple criteria for exclusion. For example, 13 teachers' responses to these item sets (4 Earth's place in the universe, 1 ecosystems, 8 matter) met all three exclusion categories, while six teachers' responses met both the latency and the "at or below chance" exclusion criteria. Our goal was to include participants' results when we had good reason to believe that their scores reflected their CKT in these areas. Therefore we decided that these three reasons for excluding cases from the analysis—in combination or alone—indicated that teachers might not have exhibited a good faith effort when responding to that set of practice-based CKT items.

Overall, the percentage of responses removed ranged from 6% to 13% in each topic area. The matter topic area had the greatest number of teachers who failed to complete 75% or more of the items and who scored at or below chance, which might be due to the fact that this item set was administered last, and teachers may have been less motivated or more fatigued at this point. After exclusion criteria were applied, the final sample included 250 upper elementary teachers who completed the following assessments: 184 matter CKT, 206 ecosystems CKT, and 203 Earth's place CKT.

Data Analysis

Selected-Response Scoring

As mentioned earlier, the selected-response questions included both traditional single-select multiple-choice questions and more innovative question formats (e.g., table, matching, rank order, or select all that apply). Each single-select multiple-choice question was scored as either correct or incorrect and counted as one item in the final scoring model. However, because the other types of selected-response questions required the test taker to make multiple selections (e.g., by selecting answers for four different rows in one table question), we explored scoring these questions as multiple items in the scoring model. First, we had to determine whether it made sense to score each question as a single item, so that all selections would need to be correct to receive credit, or whether each selection the teacher made in these questions could be scored as a separate item.

To score these innovative questions as multiple items in the scoring model, we had to ensure that there was not a strong relationship between teachers' selections on the various individual items within each of these selected-response questions (e.g., the relationship between teachers' selections on two different rows of a table question). To do so, we examined the intra-item correlations for all of these selected-response questions. Findings revealed mainly weak or negative intra-item correlations among selections for these item types: 97% of intra-item correlations for table questions were less than .50; 75% of correlations for matching or rank-order items were less than .50; and 90% of select-all-that-apply items had intra-item correlations less than .50. On the basis of these findings, we decided to treat each selection within these questions as separate items in the scoring model. This decision resulted in each of these types of selected-response questions counting as between three and six items in the final scoring model, depending on the number of items within each question. Finally, prior to calculating overall teacher scores for each topic, we removed any items with negative biserials from the final scoring model, because this could indicate that those items were not functioning similarly to the other practice-based CKT items in the item set.

Constructed-Response Scoring

Before the constructed-response questions could be incorporated into the scoring model, all of the constructed responses had to be human scored. Across the three sets of practice-based CKT questions, four constructed-response questions were included (see Table 4). We used a similar process to develop and apply a scoring rubric for each constructed-response question. First, we reviewed a subset of the teachers' responses to each question and identified key features of the participants' written responses. Then, we developed an analytic rubric to note the absence and presence of these key features and to evaluate the quality of the features included in each response. For example, for the Solar System Poster question (EAR_0029a; Appendix D), teachers had to identify three misconceptions that this poster may cause or reinforce for upper elementary students. Each idea nominated in a teacher's response would be evaluated as (a) complete and correct, (b) incomplete and/or incorrect, or (c) absent. For an idea to be complete and correct, the response had to explain the particular student misconception that the representation of the solar system on the poster might cause or reinforce, and this misconception must be clearly linked to what could be discerned by looking at the poster. An idea was scored as incomplete and/or incorrect if the response only indicated what was wrong with the poster but did not explain how that relates to a specific student misconception that would be caused or reinforced (e.g., nominated a category but did not explain how or what about that aspect would be confusing/problematic) or if the response described a misconception that did not relate to that representation of the solar system. Overall, we found 10 different misconceptions nominated across the teachers' responses for this question and scored each of their responses for the presence or absence of each misconception as well as for the quality of the description. Next, we identified the nature of the key ideas that should be included in each response to provide a complete and accurate answer to the given prompt.

For each constructed-response question, we developed a holistic rubric on a 3-point scale ranging from 0 to 2. For example, a teacher would receive a score of 2 — the highest score possible — for the Solar System Poster question if his or her response identified and clearly explained three aspects of the poster that did not accurately represent the components of the solar system, how they are arranged, or how they interact and if these three aspects related to student misconceptions that the poster might cause or reinforce. In addition, we identified two teachers' responses as exemplars for each scoring level in the holistic rubric. Appendices E – H show the holistic scoring rubrics used to score each of the four constructed-response questions and examples of teachers' responses at each scoring level.

For the three constructed-response questions that required teachers to construct an accurate and accessible scientific explanation, we developed two holistic rubrics — one evaluating the conceptual accuracy of the explanation and another evaluating the accessibility of the explanation. The accessibility criteria referred to the extent to which the explanation was clear, comprehensible, and easily understandable by students at a particular grade level. Accessible explanations used language that students at that grade level are likely to understand; used comprehensible analogies or examples to explain key concepts and ideas; and exhibited strong coherence and connectivity across different parts of the explanation by using a logical progression of ideas. In our scoring, we found that teachers could provide an explanation that accurately and completely addressed the main science conceptual ideas but the explanation could vary in terms of how clear, comprehensible, and easily understood it would be by students at the specified grade level. Because both features of providing explanations are critical for the work of teaching elementary science, and these features could vary within the explanations provided, we scored these aspects of teachers' explanations separately.

After creating the scoring rubrics, two raters independently scored approximately 15% of the teachers' responses to each constructed-response question. These raters met to reconcile any scoring discrepancies and to clarify or revise any aspects of the scoring rubric. If raters failed to achieve an 80% reliability level after the initial round of scoring for each question, they completed another round of independent scoring and reconciling to develop a shared understanding of how to apply the scoring rubric consistently. Multiple rounds of double-scoring were needed for some of the constructed-response questions. Once raters achieved an interrater reliability greater than 80%, one of the raters scored the remaining teachers' responses independently. During the final independent scoring, any responses that were deemed challenging to score received a second rating, and these double scored responses were reconciled as needed.

Analyzing Item Properties and Item Set Scores

The main analysis focused on examining the extent to which these practice-based CKT items capture variability in elementary science teachers' performance and differentiate among lower and higher performing teachers. In addition, the analysis explored how teachers' performances on these items varied within and across topics. To begin, we conducted classical test and item analyses to examine teachers' performance on the practice-based CKT items. We generated scores for each teacher per topic on each set of practice-based CKT items the teacher completed. Each teacher could receive a maximum of three scores, depending on which item sets the teacher completed. We examined item difficulty (proportion of teachers answering each item correctly) and item discrimination (biserial or polyserial correlations depending on item type, between-item scores and scale scores within topics) for all of the practice-based CKT items by topic. In addition, we examined the score distributions and reliability (Cronbach's alpha) for each set of practice-based CKT items. Finally, we used correlations and scatterplots to examine relationships in teachers' scores across topics.

Results

Item Difficulty

Selected-Response Items

On the basis of teachers' responses to the selected-response items, we calculated the total percentage of teachers who answered each item correctly. To understand how difficult these selected-response items were for this sample of teacher participants, we classified items into one of three categories, depending on the percentage of teachers who answered each item correctly. Each item was categorized as either (a) quite difficult because less than 25% of the teachers responded correctly to that item, (b) not very difficult because more than 75% of the teachers answered the item correctly, or (c) somewhere in the middle of the difficulty range because between 25% and 75% of the teachers answered the item correctly. Table 8 shows a detailed picture of teachers' performances on these items per topic and provides a sense of how difficult these items were for this sample of elementary science teachers.

Across all three topics, findings show that the majority of the items fell somewhere in the middle of the difficulty range because between 25% and 75% of the teachers answered those items correctly. In the matter topic, 12 of the 15 items were answered correctly by 25–75% of the teachers, while only 3 of the 15 items were answered correctly by 75% or more of the teachers. For the ecosystems topic, approximately half of the items were not very difficult, while the other half of the items were moderately difficult. Item difficulty for the Earth's place in the universe item set fell in the middle of the other two topics. Approximately two-thirds of the items in this topic were answered correctly by 25–75% of the teachers, while approximately one-third of the items were on the easier end of the difficulty scale. Overall, there is a fairly good distribution in terms of item difficulty within all three practice-based CKT item sets, although there were no items that less than 25% of the teachers answered correctly.

In addition, when comparing item difficulty across topics, these item-level scores suggest that, on average, the practice-based CKT matter items tended to be more difficult than the items within the other two topics. For example, in the matter topic, 20% of the items (3 of the 15 items) were answered correctly by 75% or more of the teachers. However, for the ecosystem topic, a little less than 50% of the items (8 of the 17 items) were answered correctly by 75% or more of the teachers. A similar pattern exists for the other topic, although the difference is not quite as pronounced. Appendix I provides the percentage correct values for all of the practice-based CKT items administered in the online assessments.

Constructed-Response Items

Table 9 provides the results from the scoring across the four constructed-response questions in these item sets. One constructed-response question required teachers to identify possible student misconceptions that could be caused or reinforced by a poster representation of the solar system (EAR_0029a, Solar System Poster); teachers' responses to this question were scored on a 2-point scale. The other three constructed-response questions required teachers to provide an accurate and accessible explanation about key scientific concepts or phenomena. For example, one question required teachers to provide an explanation for upper elementary students about the relationship between mass and weight. Another question required an explanation of why fertilizer helps plants grow but is not considered food for plants. These three constructed-response questions had two scoring aspects reported in this table: one for the scientific accuracy of the explanation and one for the accessibility of the explanation. Although the scoring for each aspect was completed on a 2-point scale, these numerical scores were translated into a 1-point score (0, .5, or 1 point) so that each of these three constructed-response questions (e.g., MAT_0009, Weight and Mass Explanation) counted between 0 and 2 points in the final scoring model.

Overall, findings reveal that teachers' responses on these constructed-response questions show variability across the score range, although for a few items, the distribution is shifted toward the higher or lower end of the scale. For the most

Table 8 Item Difficulty for Selected-Response Items

Percentage correct for items	Topic, *n* (%)		
	Matter[a]	Ecosystems[b]	Earth's place[c]
>75 (easier)	3 (20)	8 (47)	10 (36)
25–75	12 (80)	9 (53)	18 (64)
<25 (harder)	0 (0)	0 (0)	0 (0)

Note. The *n* reported here is the number of selected-response assessment items within each topic that were used to create the final scale scores. Two of the selected-response items in the matter topic were excluded, one due to negative biserial correlation and one due to inaccurate programming of answer options. One selected-response item in the ecosystems topic was excluded due to negative biserial correlation. The percentages reported in the table are the percentages of selected-response items within each topic that met the item characteristics noted.
[a] $n = 15$. [b] $n = 17$. [c] $n = 28$.

Table 9 Scoring Results for Constructed-Response Items

Item ID	Total possible points	*n*	*M*	*SD*
MAT_0009_Accessibility	1	184	0.68	0.34
MAT_0009_Accuracy	1	184	0.65	0.41
ECO_0026_Accessibility	1	206	0.51	0.38
ECO_0026_Accuracy	1	206	0.35	0.31
EAR_0029a	2	203	1.46	0.50
EAR_0038_Accessibility	1	203	0.50	0.30
EAR_0038_Accuracy	1	203	0.56	0.39

part, the standard deviations for these constructed-response items suggest that teachers' performances were distributed similarly across the full range of the scale, except for the Solar System Poster (EAR_0029a) item and the Plant Food Explanation accuracy item (ECO_0026_Accuracy). Almost all of the teachers who responded to the Solar System Poster item received a score of 1 or 2, suggesting that this item was on the easier end of the difficulty scale. In contrast, 90% of the teachers received a score in the bottom two levels of the rubric for the scientific accuracy of the Plant Food Explanation item, suggesting that these teachers struggled to incorporate correct ideas about the role of fertilizer in plant growth and development into their explanations. Taken together, these findings suggest that we were able to develop selected- and constructed-response practice-based CKT items within these three topic areas that captured sufficient variation in teachers' performance across the difficulty scale.

Item Discrimination

To understand how well these newly designed practice-based CKT items discriminate among lower and higher performing teachers in this sample, we first calculated an overall scale score for each teacher per topic. We then used correlations to examine the relationship between teachers' responses on these items and their performance on the entire item set within each topic. Table 10 shows the results of this analysis to provide a sense of how well these items discriminated among lower and higher performing teachers within each topic area.

The correlations shown compare how well the teachers did when answering one item relative to how well they did answering all the items within a particular item set. The key idea to remember is that these correlations indicate the extent to which the teachers who performed well on the whole item set are also the ones who performed well on the item being measured (and vice versa). Another way to think about this is, were the higher performing teachers more likely to answer this item correctly? Were the lower performing teachers more likely to get this item wrong? For this analysis, we categorized items based on three possible ranges for these biserial or polyserial correlations: (a) greater than .50, suggesting a better fit between the item and the overall item set; (b) between .25 and .50; and (c) less than .25, suggesting a poor fit between the item and the overall item set.

 ETS Research Report No. RR-17-43.

Table 10 Item Discrimination

Biserial/polyserial correlation[a]	Topic, *n* (%) Matter[b]	Ecosystems[c]	Earth's place[d]
>.50 (better fit)	1 (6)	4 (21)	9 (29)
.25–.50	9 (53)	10 (53)	16 (52)
<.25 (poor fit)	7 (41)	5 (26)	6 (19)

Note. The *n* reported here is the number of total assessment items (selected response and constructed response) within each topic that were used to create the final scale scores. The percentages reported in the table are the percentages of items within each topic that met the item characteristics noted.
[a]Biserial correlations were calculated between each selected-response item and the overall scale score for that item set. Polyserial correlations were calculated between each constructed-response item and the overall scale score for that item set. [b]$n = 17$. [c]$n = 19$. [d]$n = 31$.

Findings show that within each topic, the majority of items have a moderate to strong relationship to the overall scale score, suggesting that a large proportion of the items differentiate moderately well. Across the three topics, 59–81% of items had biserials greater than .25. All of the constructed-response questions showed polyserial correlations with the relevant scale score between .25 and .50 (see Appendix I). However, a sizable number of items do not have a strong correlation with the overall scale. This finding is most pronounced for the matter item set, where 41% of the items had biserial correlations less than .25. There are two likely reasons for this large number of items that do not discriminate well. One reason could be that a particular item is measuring a type of knowledge that is different from what the other items on the scale are measuring. The item does not discriminate well because it is not a good fit with the other items. This would be particularly apparent for a scale measure with a large number of items of this kind. For example, in the matter topic, teachers can have CKT for addressing some of these properties (e.g., weight and volume) without an understanding of other ones (e.g., density). A second reason could be that the item itself is flawed. The item may not have a defensible correct answer, or there may be issues with how the item is understood by test-taking participants. To better understand these results, we are currently conducting cognitive interviews with teachers to learn more about the knowledge and reasoning teachers use when responding to these items.

Score Distribution and Reliability

Figures 5, 6, and 7 show the overall score distributions for each of these item sets. In general, these score distributions reveal a wide range of scores within each item set. The charts indicate that the distribution of the practice-based CKT matter items is slightly positively skewed. However, the distributions for the practice-based items in the other two topics are negatively skewed, with slightly more scores toward the higher end of the scoring range.

As shown in Table 11, in terms of the proportion correct, teachers' mean scores on these three item sets varied from a low of 60% for the matter topic to a high of 73% for the ecosystems topic. Across these three topics, the mean score was

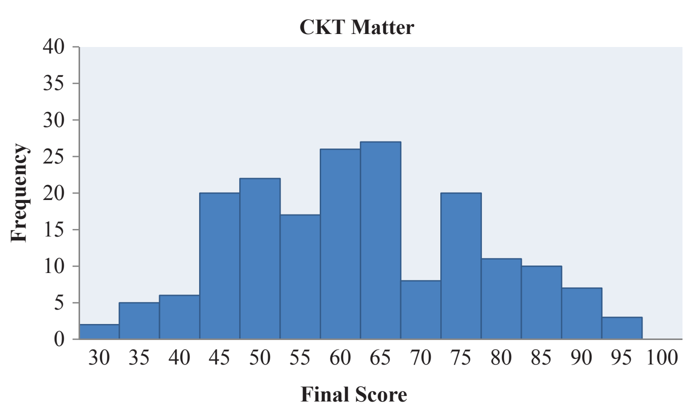


Figure 5 Score distribution for practice-based content knowledge for teaching matter item set (% correct).

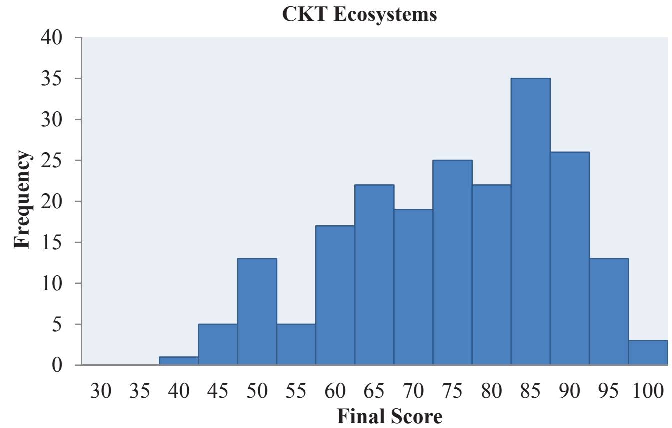


Figure 6 Score distribution for practice-based content knowledge for teaching ecosystems item set (% correct).

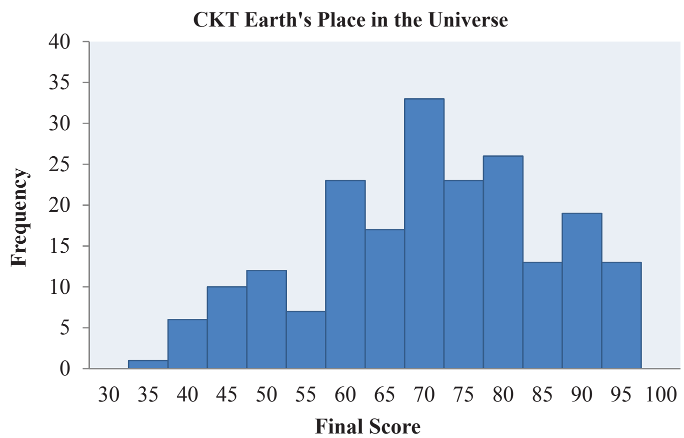


Figure 7 Score distribution for practice-based content knowledge for teaching Earth's place in the universe item set (% correct).

Table 11 Descriptive Statistics by Topic

	Matter	Ecosystems	Earth's place
n (participants)	184	206	203
Total score points[a]	17	19	32
M	0.60	0.73	0.69
SD	0.15	0.14	0.15
Minimum	0.29	0.37	0.34
Maximum	0.94	1.00	0.95
Scale reliability[b]	0.57	0.63	0.77

[a]The total score points reported here are based on the items included in the final scoring models, after removal of items with negative biserial or polyserial correlations. One item from the matter item set was removed from the final scoring model due to inaccurate programming of answer options into the online delivery system. [b]Scale reliability was calculated using Cronbach's alpha as a measure of internal consistency across the items within each item set.

lower for the practice-based CKT items in the matter topic, suggesting that these items were more challenging for teachers in this sample. In addition, the standard deviations suggests that for all three item sets, teachers' scores are not clustered too tightly around the mean. These findings suggest that overall, the practice-based CKT item sets captured a range of scores across the sample teacher population.

Table 11 reveals that only one of these item sets—the Earth's place in the universe item set—achieved scale reliability greater than .7. A likely reason for this finding is that the matter and ecosystems practice-based CKT item sets had a significantly reduced number of items compared to the Earth's place in the universe item set. If we were to increase the number of items on these two topics to equal the number of items on the Earth's place in the universe item set, the reliability for the other two measures would increase to greater than .7.[7] Finally, we also learned that the scale reliability did not change significantly when adding the constructed-response items to the overall scale. The scale reliability for each item set increased by .01 or less, except for the matter topic, which increased by .07 with the addition of two constructed-response items. This finding is probably due to the lower number of items on this scale, which would make it more sensitive to adding in two items that are constructed-response item types.

Relationship Between Teachers' Scores Across Topics

For each teacher, we created up to three scores, one for each set of practice-based CKT items the teacher completed, representing the percentage of items each teacher answered correctly within each topic. We used correlations to examine the relationship between teachers' performance on these three item sets and to determine the extent to which teachers' performance on these practice-based CKT science items varied across topics. In particular, we were interested in better understanding whether teachers who performed well on the practice-based CKT items in one topic area were also the ones who were most likely to score well on these items in the other topic areas. Because each teacher had the choice to complete one, two, or three items sets, the sample for this analysis varied by topic comparison. For this analysis, our sample included the following number of teachers when comparing their performances across these three topics: 195 teachers for ecosystem and Earth's place in the universe, 141 teachers for ecosystem and matter, and 144 teachers for matter and Earth's place in the universe.

Findings show weak to moderate correlations between teachers' performance on these practice-based CKT item sets across each topic comparison: ecosystem and Earth's place in the universe, $r = .390$, $p < .01$; ecosystem and matter, $r = .293$, $p < .01$; and matter and Earth's place in the universe, $r = .471$, $p < .01$. These correlations suggest that there is not strong consistency in how well teachers performed on these item sets across topics. In Figures 8, 9, and 10, we plot these relationships to illustrate the general trend in teachers' performance across topics showing much variability at different score levels within these plots. For example, findings show that 14 teachers scored 82% correct on the ecosystems item set and that these 14 teachers had scores on the Earth's place in the universe item set ranging from a low of 39% correct to a high of 95% correct. Likewise, the 11 teachers who scored 53% correct on the matter item set had a range of scores from 47% to 95% correct on the ecosystems item set.

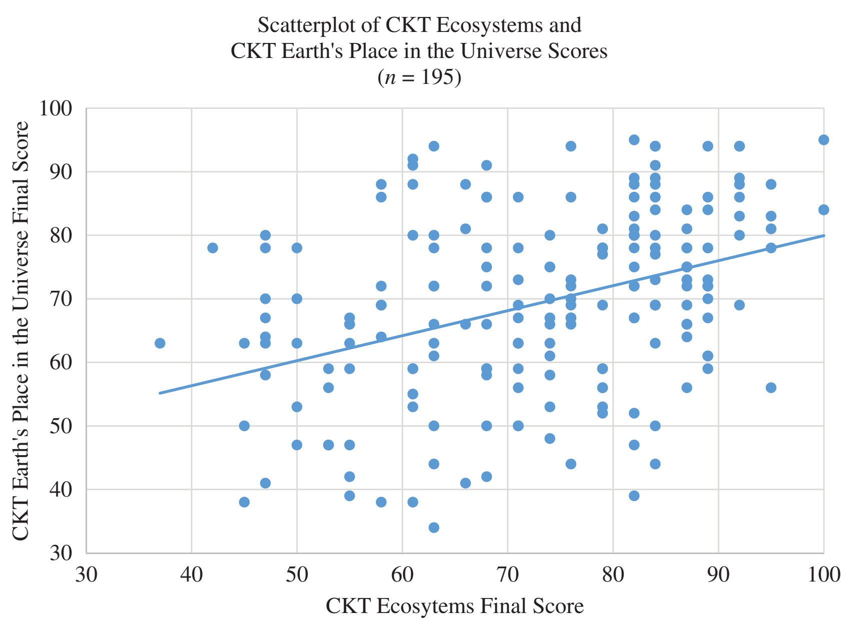


Figure 8 Scores for teachers who completed both the practice-based content knowledge for teaching ecosystem and Earth's place in the universe item sets.

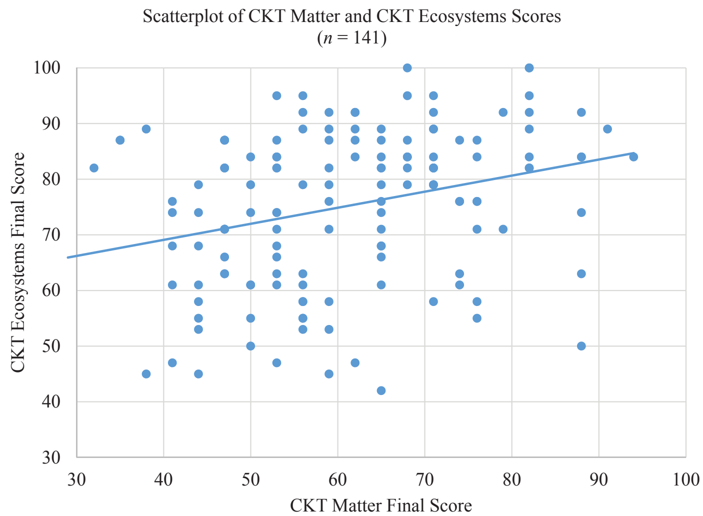


Figure 9 Scores for teachers who completed both the practice-based content knowledge for teaching ecosystem and matter item sets.

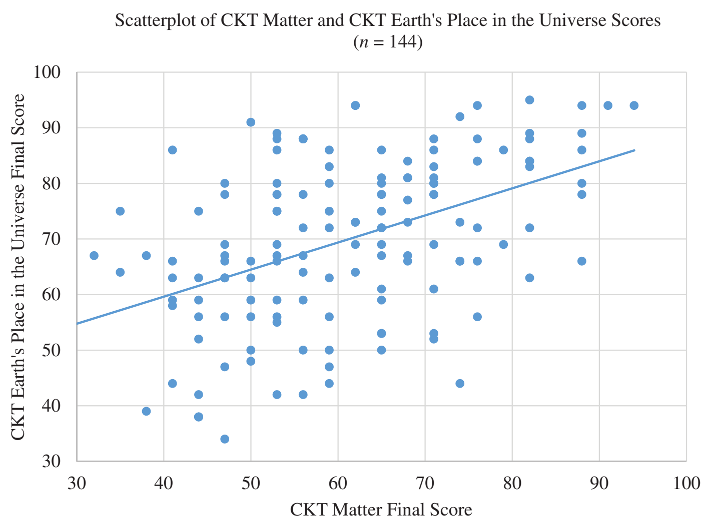


Figure 10 Scores for teachers who completed both the practice-based content knowledge for teaching matter and Earth's place in the universe item sets.

Summary

One of the pressing problems in the science education field right now is identifying the critical areas of knowledge that support science teachers in being effective practitioners. In science, in particular, there has been extensive research to describe and determine the components of this professional knowledge base. More recently, efforts in the field have turned toward developing assessments to measure the professional knowledge needed for teaching science. The development and research efforts reported here target this problem space and focus on how to assess the CKT that science teachers use as they engage in various tasks of teaching.

In particular, the research described was guided by two primary goals. First, we set out to develop an innovative type of assessment item that measured the types of science knowledge used primarily in elementary science teaching. Second, we aimed to gather evidence on how these items perform in terms of item difficulty and discrimination and to understand the extent to which teachers' performances on these item sets differ across topics.

Overall, findings suggest that we were able to develop new science assessment items that situate teachers in applying their subject matter knowledge to the content problems that they face. Results show that the majority of items are able to capture variability in teachers' performances. In addition, teachers' overall scores on these items sets reveal near-normal distributions across the majority of the scale within each topic. In terms of the biserial and polyserial correlations, which measure the relationship between teachers' item-level responses and their performance on the entire item set, within each topic, the majority of items show a moderate to strong relationship to the criterion measures. This finding suggests that a large proportion of the practice-based CKT items differentiate moderately well between lower and higher performing teachers within each topic.

Another important finding is that there was variation in teachers' performances across topics, suggesting that teachers can have adequate or strong CKT in one topic but have relatively weak CKT in another topic. The idea that CKT occurs at the topic level (e.g., matter or ecosystems), as opposed to the disciplinary level (e.g., biology or science), has been gaining momentum in the field. Recent models of teacher professional knowledge have attended to the topic-specific nature of this knowledge base, and assessment tools have been designed to gather, record, and measure the knowledge needed to teach particular topics within different grade levels (Gess-Newsome, 2015; Loughran, Berry, & Mulhall, 2012). This finding suggests that researchers and assessment developers need to pay attention to the topic-specific nature of science teachers' CKT. Future work in this area with larger teacher samples and more fully specified and developed tests could include the use of factor analysis to investigate whether practice-based CKT items can be used to assess measurably distinct domains of science knowledge for teaching.

As mentioned earlier, the field has made progress in developing measures of science teachers' subject matter knowledge, with limited attempts to develop practice-based CKT measures that can be used on a large scale. In this research project, we developed practice-based CKT assessment questions within three different science topics and addressed various tasks of teaching within these questions. Overall, the item-level statistics showed adequate item functioning for a majority of the new assessment items, supporting a beginning proof-of-concept for these practice-based CKT science items. However, this work is just the beginning of a larger trajectory of research that is important to understanding the critical areas of knowledge needed for effective science teaching and how to assess that knowledge.

Specifically, there are three key areas where we think there is a need for additional work. First, it will be important to better understand the kinds of reasoning and knowledge that science teachers leverage when responding to these items. The item-level statistics are just one part of the picture. Data on how participants reason through these practice-based CKT items would provide evidence that they are using the types of knowledge that are used in science teaching, which is an important part of the validity argument for CKT assessment (Kane, 2006). In addition, studies that compare how teachers respond to the practice-based CKT items and to conventional subject matter knowledge assessment items will help support the claim that practice-based CKT items measure types of knowledge that differ from the science knowledge assessed on the conventional assessments. Such evidence would provide important guidance to test developers in modeling CKT for assessment, for example, as they employ methods such as evidence-centered design (Mislevy & Riconscente, 2006) to construct tests that assess the range of knowledge used in teaching science.

Second, it will be important to understand how these items function with specific populations of interest. In this research, we used these practice-based CKT items with a sample of practicing upper elementary science teachers. However, it is unclear how these items would function with preservice elementary teachers or teachers with different backgrounds or teaching experiences. Third, this exploratory research could be extended by studying a variety of uses of these practice-based CKT items. For example, these practice-based CKT science measures can be used to support some key needs in the field, including (a) tracking teachers' learning and providing formative feedback to teachers; (b) serving as evaluation tools for professional development providers, teacher education programs, and school districts; and (c) supporting teacher selection efforts for licensure decisions. To take the next step toward developing valid assessments of CKT science, it will be important to better understand how these items and associated assessments function for different uses (e.g., formative uses in teacher preparation, licensure, teacher evaluation) and for different populations of interest (Kane, 2006).

Acknowledgments

This research was supported by a grant from the National Science Foundation (DGE-1445641). The opinions expressed herein are those of the authors and not of the funding agency. The authors would like to thank Joe Ciofalo, Barbara Weren, Aaron Rogat, and Tim Yanka for their support in conceptualizing and authoring the practice-based items designed to assess CKT elementary science teaching that serve as the focus for this research. The authors also extend their appreciation to Israel Solon, Adrian Flores, Laina Allen, and James Capps — all experienced science assessment developers — for providing their time and expertise to review and critique these items prior to administration and to score teachers' answers to the constructed-response items. In addition, the authors are grateful to Kathy Roth, Ed Smith, Deb Smith, and Tim Smith for providing critical feedback on these newly designed assessment items. Finally, the authors appreciate the commitment and time of the elementary teachers who volunteered to participate in this study and share their expertise with project researchers.

Notes

1 We recognize that many terms have been used to describe science teachers' knowledge base for teaching, including *pedagogical content knowledge* (PCK), *professional knowledge*, and *subject matter knowledge for teaching*. We use the term *content knowledge for teaching* to encompass the subject matter knowledge that science teachers are called on to use as they engage in the work of teaching science, which includes both conventional content knowledge and other specialized forms of knowledge used in teaching practice.

2 Each practice-based CKT question required teachers to make one or multiple selections, depending on the question format used. For example, each single selection multiple choice question required teachers to select the single best answer choice among four options provided. However, a table question includes multiple rows and required teachers to select the best answer choice for each row, resulting in multiple selections per table question. In this report, we refer to each of these individual selections as an item, which means that table questions correspond to multiple items, while a single-selection multiple choice question corresponds to only one item.

3 On the basis of the boundary statement provided in the NGSS, noting that "weight and mass are not distinguished at this grade level," we decided to use the term *weight* instead of the term *mass* in this question because *weight* is more likely to be a familiar term to upper elementary students.

4 In this report, we report on the efforts to conceptualize, develop, and pilot the practice-based CKT science questions. Results of teachers' performance on the straight subject matter knowledge questions, and how their performance on these two types of knowledge measures compare to one another, will be reported in a forthcoming manuscript.

5 The criterion of test speediness was based on establishing a minimal latency for completing each part of the assessment. The minimal latency was based on timing one of the researchers as he read each question and chose a response at random. In other words, this latency criterion was a proxy for time to read through the assessment with minimal time spent on choosing responses. Teachers' responses for an assessment were excluded if their time spent on the assessment did not meet this minimum threshold. The minimum latency for the Earth's place CKT section was 9 minutes; for the ecosystems CKT section, it was 7 minutes; and for the matter CKT section, it was 8 minutes.

6 Chance scores were calculated by dividing the number of total possible points by the total number of answer choices for selected-response items only. The following are chance scores for each set of practice-based CKT items: Earth's place CKT, 34%; ecosystems CKT, 39%; and matter CKT, 34%.

7 We used the Spearman–Brown prediction formula to determine what the reliability would be if we were to extend the number of items within the matter and ecosystems item sets to match the number of items in the Earth's place in the universe item set. The reliability for both item sets would increase to .79 for the matter item set and to .81 for the ecosystem item set.

References

Abell, S. K. (2008). Twenty years later: Does pedagogical content knowledge remain a useful idea? *International Journal of Science Education, 30*(1), 1405–1416. https://doi.org/10.1080/09500690802187041

Avraamidou, L., & Zembal-Saul, C. (2005). Giving priority to evidence in science teaching: A first-year elementary teacher's specialized practices and knowledge. *Journal of Research in Science Teaching*, *42*(9), 965–986. https://doi.org/10.1002/tea.20081

Ball, D. L., Hill, H. C., & Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator, 29*(1), 14–22, 43–46.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389–407. https://doi.org/10.1177/0022487108324554

Bertram, A., & Loughran, J. (2012). Science teachers' views on CoRes and Pap-eRs as a framework for articulating and developing pedagogical content knowledge. *Research in Science Education, 42*(6), 1027–1047. https://doi.org/10.1007/s11165-011-9227-4

Carlsen, W. S. (1991). Effects of new biology teachers' subject matter knowledge on curriculum planning. *Science Education, 75*(6), 631–647. https://doi.org/10.1002/sce.3730750605

Clermont, C. P., Borko, H., & Krajcik, J. S. (1994). Comparative study of the pedagogical content knowledge of experienced and novice chemical demonstrators. *Journal of Research in Science Teaching, 31*(4), 419–441. https://doi.org/10.1002/tea.3660310409

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. https://doi.org/10.1002/tea.20440

Davis, E. A. (2006). Preservice elementary teachers' critique of instructional materials for science. *Science Education, 90*(2), 348–375. https://doi.org/10.1002/sce.20110

Davis, E. A., Petish, D., & Smithey, J. (2006). Challenges new science teachers face. *Review of Educational Research, 76*(4), 607–651. https://doi.org/10.3102/00346543076004607

Davis, E. A., & Smithey, J. (2009). Beginning teachers moving toward effective elementary science teaching. *Science Education, 93*(4), 745–770. https://doi.org/10.1002/sce.20311

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Forbes, C. T., Sabel, J. L., & Biggers, M. (2015). Elementary teachers' use of formative assessment to support students' learning about interactions between the hydrosphere and geosphere. *Journal of Geoscience Education, 63*(3), 210–221. https://doi.org/10.5408/14-063.1

Gess-Newsome, J. (1999). Pedagogical content knowledge: An introduction and orientation. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 3–17). Dordrecht, Netherlands: Kluwer. https://doi.org/10.1007/0-306-47217-1_1

Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 28–42). New York, NY: Routledge. https://doi.org/10.4324/9781315735665

Gitomer, D. H., Phelps, G., Weren, B. H., Howell, H., & Croft, A. J. (2014). Evidence on the validity of content knowledge for teaching assessments. In T. Kane, K. Kerr, & R. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching Project* (pp. 493–528). San Francisco, CA: Jossey-Bass. https://doi.org/10.1002/9781119210856.ch15

Henze, I., & van Driel, J. H. (2015). Toward a more comprehensive way to capture PCK in its complexity. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 120–134). New York, NY: Routledge. https://doi.org/10.4324/9781315735665

Herrenkohl, L. R., & Guerra, M. R. (1998). Participant structures, scientific discourse, and student engagement in fourth grade. *Cognition and Instruction, 16*(4), 431–473. https://doi.org/10.1207/s1532690xci1604_3

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education, 39*(4), 372–400.

Hill, H. C., Dean, C., & Goffney, I. M. (2007). Assessing elemental and structural validity: Data from teachers, non-teachers, and mathematicians. *Measurement: Interdisciplinary Research and Perspectives, 5*(2–3), 81–92. https://doi.org/10.1080/15366360701486999

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematical knowledge for teaching. *The Elementary School Journal, 105*(1), 11–30. https://doi.org/10.1086/428763

Hogan, K., Natasi, B. K., & Pressley, M. (2000). Discourse patterns and collaborative scientific reasoning in peer and teacher-guided discussions. *Cognition and Instruction, 17*(4), 379–432. https://doi.org/10.1207/s1532690xci1704_2

Howell, H., Phelps, G., Croft, A., Kirui, D., & Gitomer, D. (2013). *Cognitive interviews as a tool for investigating the validity of content knowledge for teaching assessments* (ETS Research Report No. RR-13-19). Princeton, NJ: ETS. https://doi.org/10.1002/j.2333-8504.2013.tb02326.x

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: Praeger.

Kersting, N. B. (2008). Using video clips of mathematics classroom instruction as item prompts to measure teachers' knowledge of teaching mathematics. *Educational and Psychological Measurement, 68*(5), 845–861. https://doi.org/10.1177/0013164407313369

Kloser, M. (2014). Identifying a core set of science teaching practices: A delphi expert panel approach. *Journal of Research in Science Teaching, 51*(9), 1185–1218. https://doi.org/10.1002/tea.21171

Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *ZDM*, *40*(5), 873–892. https://doi.org/10.1007/s11858-008-0141-9

Lee, E., Brown, M. N., Luft, J. A., & Roehrig, G. H. (2007). Assessing beginning secondary teachers' PCK: Pilot year results. *School Science and Mathematics, 107*(2), 418–426. https://doi.org/10.1111/j.1949-8594.2007.tb17768.x

Levin, D. T. (2013). *Becoming a responsive science teacher: Focusing on student thinking in secondary science*. Arlington, VA: NSTA Press.

Loughran, J., Berry, A., & Mulhall, P. (2012). *Understanding and developing science teachers' pedagogical content knowledge* (2nd ed.). Rotterdam, Netherlands: Sense. https://doi.org/10.1007/978-94-6091-821-6

Loughran, J., Mulhall, P., & Berry, A. (2004). In search of pedagogical content knowledge in science: Developing ways of articulating and document professional practice. *Journal of Research in Science Teaching, 41*(4), 370–391. https://doi.org/10.1002/tea.20007

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 95–132). Dordrecht, Netherlands: Kluwer Academic. https://doi.org/10.1007/0-306-47217-1_4

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York, NY: Macmillan.

Meyer, H. (2004). Novice and expert teachers' conceptions of learners' prior knowledge. *Science Education, 88*(6), 970–983. https://doi.org/10.1002/sce.20006

Minner, D., Martinez, A., & Freeman, B. (2012). *Compendium of research instruments for STEM education: Part I. Teacher practices, PCK, and content knowledge*. Washington, DC: Community for Advancing Discovery Research in Education.

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: Erlbaum.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academies Press.

National Research Council. (2007). *Take science to school: Learning and teaching science in grades K–8* (R. A. Duschl, H. A. Schweingruber, & A. W. Shouse, Eds.). Washington, DC: National Academies Press. https://doi.org/10.1007/s11191-011-9353-4

National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press. https://doi.org/10.17226/13165

National Research Council. (2013). *Monitoring progress toward successful K–12 STEM education: A nation advancing?* Washington, DC: National Academies Press. https://doi.org/10.17226/13509

Next Generation Science Standards Lead States. (2013). *Next Generation Science Standards: For states, by states*. Washington, DC: National Academies Press. https://doi.org/10.17226/18290

Nilsson, P., & van Driel, J. (2010). Teaching together and learning together—Primary science student teachers' and their mentors' joint teaching and learning in the primary classroom. *Teaching and Teacher Education, 26*(6), 1309–1318. https://doi.org/10.1016/j.tate.2010.03.009

Otero, V. K., & Nathan, M. J. (2008). Preservice elementary teachers' views of their students' prior knowledge of science. *Journal of Research in Science Teaching, 45*(4), 497–523. https://doi.org/10.1002/tea.20229

Park, S., & Oliver, J. S. (2008). Revisiting the conceptualization of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education, 38*(3), 261–284. https://doi.org/10.1007/s11165-007-9049-6

Park, S., & Suh, J. K. (2015). From portraying to assessing PCK: Drivers, dilemmas, and directions for future research. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 104–119). New York, NY: Routledge. https://doi.org/10.4324/9781315735665

Phelps, G., Weren, B., Croft, A., & Gitomer, D. (2014). *Developing content knowledge for teaching assessments for the Measures of Effective Teaching study* (ETS Research Report No. RR-14-33). Princeton, NJ: ETS. https://doi.org/10.1002/ets2.12031

Riggs, I., & Enochs, L. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education, 74*(6), 625–637. https://doi.org/10.1002/sce.3730740605

Roth, K. J., Garnier, H. E., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. I. (2011). Videobased lesson analysis: Effective science PD for teacher and student learning. *Journal of Research in Science Teaching, 48*(2), 117–148. https://doi.org/10.1002/tea.20408

Ruiz-Primo, M. A., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment, 11*(3–4), 205–235. https://doi.org/10.1080/10627197.2006.9652991

Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal, 50*(5), 1020–1049. https://doi.org/10.3102/0002831213477680

Schneider, R. M., & Plasman, K. (2011). Science teacher learning progressions: A review of science teachers' pedagogical content knowledge development. *Review of Educational Research, 81*(4), 530–565. https://doi.org/10.3102/003465431142338

Schwab, J. (1978). Education and the structure of the disciplines. In I. Westbury & N. J. Wilkof (Eds.), *Science curriculum and liberal education* (pp. 229–272). Chicago, IL: University of Chicago Press.

Schwarz, C., Reiser, B., Davis, E., Kenyon, L., Acher, A., Fortus, D., … Krajcik, J. S. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching, 46*(1), 232–254. https://doi.org/10.1002/tea.20311

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14. https://doi.org/10.3102/0013189x015002004

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Smith, P. S., & Taylor, M. J. (2010, March). *New tools for investigating the relationship between teacher content knowledge and student learning*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Philadelphia, PA.

van Driel, J. H., Verloop, N., & de Vos, W. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching, 35*, 673–695. https://doi.org/10.1002/(sici)1098-2736(199808)35:6<673::aid-tea5>3.0.co;2-j

Wilson, S. M. (2016). *Measuring the quantity and quality of the K–12 STEM teacher pipeline. (SRI education white paper)*. Menlo Park, CA: SRI International.

Windschitl, M., Thompson, J., Braaten, M., & Stroupe, D. (2012). Proposing a core set of instructional practices and tools for teachers of science. *Science Education, 96*(5), 878–903. https://doi.org/10.1002/sce.21027

Zembal-Saul, C., Krajcik, J., & Blumenfeld, P. (2002). Elementary student teachers' science content representations. *Journal of Research in Science Teaching, 39*(6), 443–463. https://doi.org/10.1002/tea.10032

Appendix A

Task Design Rationale for Practice-Based Content Knowledge for Teaching Item

Earth's Place in the Universe Item 0036b: Sun Shadow

What Is This Assessment Task Asking?

The task here is to evaluate which, if any, of the student responses provides evidence that the student understands each concept indicated. To do this, you must interpret and evaluate each student response individually to determine if it shows that the student understands each of the concepts given. While comparing student answers may be helpful, it is not necessary for answering the question correctly.

What Information Is Important?

It is important to note that the teacher asks his students to use Sun path diagrams to show the approximate length and position of shadows cast for a meter stick at three different time points in the day (9 A.M., 12 P.M., and 3 P.M.). You need to recognize that each Sun path diagram has a line representing the Sun's apparent path from sunrise to sunset and a small yellow circle representing the Sun. The yellow circle (the Sun) is at a precise place along the line to show where the Sun is located in the sky at a specific time. In addition, in these diagrams the red line represents a meter stick, and the black line represents the shadow cast by the Sun. You need to evaluate each student's response to see if it shows evidence that the student understands each of the concepts given. You can select more than one letter for each concept. However, you might select "none of the responses" if none of the diagrams provide evidence that any of the students understands the given concept.

To accurately evaluate the students' responses, it is important that you understand the meaning of each concept and what to look for in the diagrams that would provide evidence of understanding. The first concept—"Shadows lengths change as the Sun's position changes"—means that the shadows have different lengths depending on the Sun's position in the sky. For this concept, the students do not have to show the correct lengths of the shadows. Instead, they need to show that the shadows vary in length as the Sun's position in the sky changes. The second concept—"Shadows are longer when the Sun is positioned lower in the sky"—means that longer shadows are cast when the light source is positioned at lower angles. Students who have diagrams that show the shadow lengths as shorter when the Sun is higher in the sky, and as longer when the Sun is lower in the sky, exhibit evidence of understanding this concept. The third concept—"Shadows cast by the Sun are in the opposite position from the Sun in the sky"—means that the light is being blocked by the object, causing a shadow to appear on the opposite side of the object from the light source. Students need to have all of the shadows appearing on the opposite side of the meter stick from the Sun in order to show evidence of understanding this concept. The fourth concept—"Patterns in the daily changes to the length and direction of shadows are caused by Earth's rotation"—is a statement that explains the reason for the patterns in the shadows' changing lengths and positions. However, it is not possible that any of the Sun path diagrams can provide evidence that a student understands the scientific explanation for this phenomenon.

What Is the Rationale for Selecting an Answer?

Concept 1

The drawings that students A, C, and D produced show that the shadows change lengths as the Sun's position changes. Students A and C correctly show that the shadows' lengths at both 9 A.M. and 3 P.M. are longer than at 12 P.M. Student C does not show the shadows' correct position at 12 P.M. or 3 P.M.; however, the shadows' lengths at those two times are different from one another. Student D also shows the shadows changing lengths as the Sun's position changes, but student D's shadows get longer throughout the day. Despite the incorrect lengths shown, student D's drawing suggests that student D understands that the shadows' lengths change when the Sun's position changes. Student B shows the shadows having the same lengths at all three time points, which indicates that student B does not understand that the shadows change lengths as the Sun's position changes. Student responses A, C, and D provide evidence that they understand concept 1: "Shadows lengths change as the Sun's position changes."

Concept 2

The drawings that students A and C produced show that the shadows are longer when the Sun is positioned lower in the sky. Both students show the shadow longer at 9 A.M. and at 3 P.M. than at 12 P.M., when the Sun is highest in the sky.

Although student C has the shadows at 12 p.m. and 3 p.m. positioned in the wrong direction, this student's diagrams still show the shadows as longer when the Sun is lower in the sky. Student B shows the shadows' lengths as exactly the same at all three time points, which indicates that student B does not understand that the shadows are longer when the Sun is positioned lower in the sky. Student D shows the shadow shorter at 9 a.m., when the Sun is low in the sky, than at 12 p.m., when the Sun is highest in the sky. This response suggests that student D does not understand that the shadows are longer when the Sun is positioned lower in the sky. Student responses A and C provide evidence that they understand concept 2: "Shadows are longer when the Sun is positioned lower in the sky."

Concept 3

The drawings that students A and B produced show that the shadows cast by the Sun are in the opposite direction from the Sun's position in the sky. These two students show the shadow at each time point as facing opposite the position of the Sun on the Sun path diagram (although student B does not show the correct shadow length at 12 p.m.). Student C shows all three shadows facing the same direction at 9 a.m., 12 p.m., and 3 p.m. However, the shadows at 12 p.m. and at 3 p.m. should be positioned in directions opposite the Sun. Student D does not show the shadow at 3 p.m. located in the opposite direction from the Sun's position in the sky. Student responses A and B provide evidence that they understand concept 4: "Shadows cast by the Sun are in the opposite position from the Sun in the sky."

Concept 4

None of the students demonstrates this concept because the Sun path diagrams cannot be used to represent this concept.

Summary of Key Knowledge, Skills, and Reasoning

This assessment task draws on the following:

- knowledge that a shadow is created when light is blocked by an object
- knowledge that shadows are created on the opposite side of the object from the light source
- knowledge that the shadow of an object will be longer as the angle between the light source and the object increases (the opposite occurs as the angle decreases)
- ability to interpret and evaluate students' written diagrams for scientific accuracy

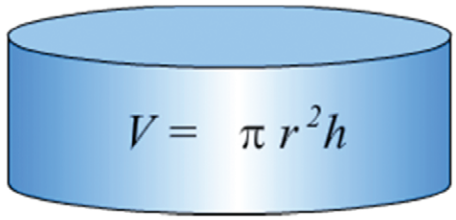
Appendix B

Matter Practice-Based Content Knowledge for Teaching Items and Item Key

Item ID	Item key
MAT_0001a	C
MAT_0001b	C
MAT_0003	B
MAT_0004_Row1	Understands both
MAT_0004_Row2	Understands method only
MAT_0004_Row3	Understands volume only
MAT_0004_Row4	Understands both
MAT_0005_Row1	Accurate
MAT_0005_Row2	Accurate
MAT_0005_Row3	Accurate
MAT_0006	A
MAT_0007	D
MAT_0009_Accessibility	see Appendix E
MAT_0009_Accuracy	see Appendix E
MAT_0010	B
MAT_0011a_Row1	Reason
MAT_0011a_Row2	Not a reason
MAT_0011a_Row3	Reason
MAT_0013	A

MAT_0001a

Mr. Fonte's class is investigating different ways to measure the volume of solid shapes. For one of the activities, Mr. Fonte discusses how to measure the volume of a cylinder using a mathematical formula. He wants to give his students a quick problem at the end of class because he is concerned that some of them are still not able to use exponents correctly to calculate the volume of a cylinder. Mr. Fonte gives his students the following formula for calculating the volume of cylinders.

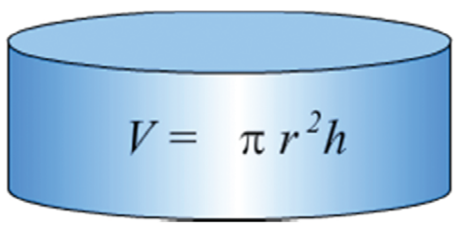


Mr. Fonte then needs to decide on the cylinder measurements to give his students for the problem. Of the following measurements, which, if any, would be <u>least</u> useful for Mr. Fonte in assessing whether his students are able to use exponents correctly to calculate the volume of a cylinder?

A. A cylinder with a 5 cm radius and a 10 cm height
B. A cylinder with a 3 cm radius and a 10 cm height
C. A cylinder with a 2 cm radius and a 10 cm height
D. All the cylinder measurements above are equally useful for Mr. Fonte in assessing whether his students are able to use exponents correctly to calculate the volume of a cylinder.

MAT_0001b

Mr. Fonte's class is investigating different ways to measure the volume of solid shapes. For one of the activities, Mr. Fonte discusses how to measure the volume of a cylinder using a mathematical formula. He wants to give his students a quick problem at the end of class because he is concerned that some of them are still not able to use exponents correctly to calculate the volume of a cylinder. Mr. Fonte gives his students the following formula for calculating the volume of cylinders.



Mr. Fonte then needs to decide on the cylinder measurements to give his students for the problem. Of the following measurements, which, if any, would be <u>least</u> useful for Mr. Fonte in assessing whether his students are able to use exponents correctly to calculate the volume of a cylinder?

A. A cylinder with a 5 cm radius and a 4 cm height
B. A cylinder with a 3 cm radius and a 6 cm height
C. A cylinder with a 2 cm radius and an 8 cm height
D. All the cylinder measurements above are equally useful for Mr. Fonte in assessing whether his students are able to use exponents to calculate the volume of a cylinder.

MAT_0003

At the beginning of a lesson on the relationship between the weight and volume of common solid objects, Ms. Zhang gives her students some prediction problems to check their current understanding.
Before handling and weighing a pair of solid objects, the students are to predict (just by looking at them) which of the two objects is heavier. Of the following pairs of objects, which would be most useful in determining the students who confuse the relationship between weight and volume? NOTE: Aluminum is relatively more dense than wood.
A. An aluminum sphere with a 7 cm diameter and a wood sphere with a 4 cm diameter B. An aluminum square pyramid with 4 cm sides and a 4 cm height and a wood square pyramid with 4 cm sides and 7 cm height C. A wood cube with 4 cm sides and a wood cube with 7 cm sides D. All of the above are equally useful in determining the students who confuse the relationship between weight and volume.

MAT_0004

Ms. Slocum is teaching a lesson on determining the volume of irregular-shaped solids, using the water displacement method with graduated cylinders. After leading a whole-class discussion and demonstrating the method, she asks students to get into small groups. The students discuss their current understanding of volume as a property of solid objects and how they would use the water displacement method to determine the volume of irregular-shaped solid objects. Ms. Slocum walks around the classroom, listens to portions of the students' conversations, and evaluates their understanding.
For each of the following student comments that Ms. Slocum hears, indicate whether it demonstrates that the student understands (1) volume as a property of solid objects and (2) how to determine the volume of irregular-shaped solids using the water displacement method with graduated cylinders. Place an X in the appropriate column for each student comment. (<u>Note</u>: Tara and Alex are discussing the beginning part of the water displacement method, while Micah and Dani are discussing the latter part of the water displacement method. You should assume that the part of the water displacement method that each student did not discuss is accurate.)

	Descriptions of Student Responses			
Student Responses	**Student Understands What Volume Is Only**	**Student Understands Water Displacement Method Only**	**Student Understands Both**	**Student Does Not Understand Either**
Tara (discussing the beginning part of the water displacement method): "Since bigger objects take up more space or volume than smaller objects, first you need to make sure that the object can fit inside the graduated cylinder and that there will be enough room for water to cover it completely. However, don't put the object inside the cylinder just yet. First, fill the cylinder with enough water to totally cover the object. Record the amount of water you put in the cylinder by looking at the lowest part of the curve and then carefully place the object in the cylinder. This will cause the water in the cylinder to rise."				
Alex (discussing the beginning part of the water displacement method): "An object's volume is based on the amount of matter in the object, which is what you'll be figuring out when you drop the object in the water. First, you'll need an empty glass graduate. Fill it with enough water so that the shape you're going to drop into it and measure will be drowned. Then, check how much water is in the glassy thing. It's really important to check the water level where the curve bends most. Then, drop the shape into the graduate but try not to smash it. The water will come up some, but it should not spill out all over the table."				
Micah (discussing the latter part of the water displacement method): "Carefully place the irregular object in the graduated cylinder. Record the volume of the water with the object fully covered by it. The measurement is taken by reading the number on the side of the cylinder where the bottom of the water curve falls. That number is the volume of the irregular solid object. The more space that the object takes up, the more the water level goes up. Objects with bigger volumes take up more space and make the water go higher."				
Dani (discussing the latter part of the water displacement method): "Put the thing you want to measure in the glass tube without busting it and find out how much water is in there. Be sure to take the measurement at the lowest point of the curvy part of the water. Then, to find out how much space the solid thing in the water takes up, take away the amount of the water in the glass tube without the thing placed in it from the amount of water in the tube with the thing placed in it. The amount of space taken up by the object in the water is the object's volume."				

MAT_0005

After learning about the properties of mass and volume of solids, Mr. Nichol's students have been trying to identify substances produced when other substances are mixed. The property of density as a useful property to consider was developed in a series of lessons to support the students in being able to identify these substances. At the end of one of the final lessons in the series, Mr. Nichol gives his students the following exit ticket prompt to check their understanding of the relationship among the mass, volume, and density of solid materials.

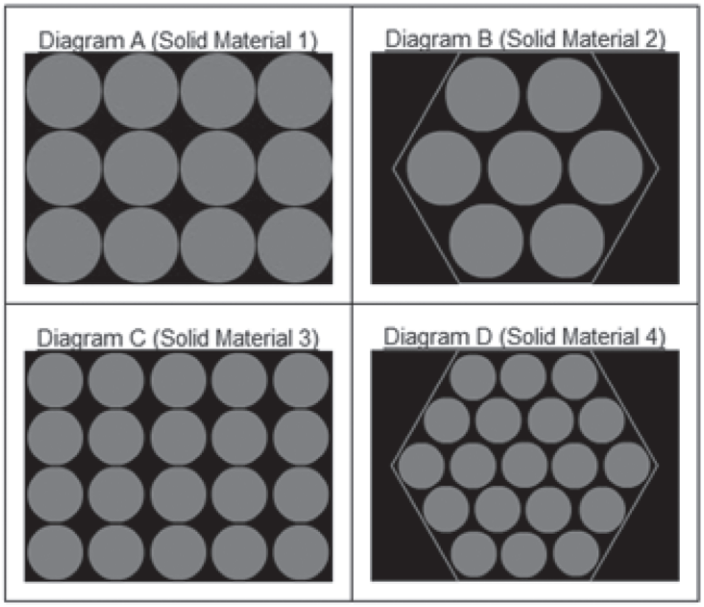
> Consider any solid object and a smaller piece of that object. Then use words, numbers, or both to describe the relationship among the mass, volume, and density of the solid object and the smaller piece of that object.

For each of the following responses, indicate whether the student uses words, numbers, or both to provide an accurate description of the relationship among the mass, volume, and density of solid materials. For each student response, place an X in the appropriate box.

Student Responses	**Accurate Description**	**Inaccurate Description**
Aaron: "First, I carefully measure the mass and volume of something like a plastic cylinder. Then, I find the density of that cylinder by dividing the mass by the volume. Next, I cut the cylinder in half. Again, I carefully measure the mass and volume of one of the halves of the cylinder and find its density by dividing the mass by the volume. Both the mass and the volume are now each half of the original whole cylinder, so the density of the half cylinder is the same as the density of the whole cylinder."		
Barbara: "I would use a whole chocolate bar and a chocolate bar broken into four equal pieces. If $M/V = D$ for the whole bar, then $M/4 \div V/4 = D$ for one of the smaller pieces. The density is the same in both cases."		
Carlos: "I start with a large block of wood of any size. I determine its mass and volume. I then divide the mass of the block by the volume and come up with the density for that block of wood. Now, I cut the huge wood block into smaller and smaller pieces until I have just tiny bits. But each bit of wood will still have the same density as the original large block of wood. This is because even though the mass of one tiny piece of wood is much smaller than the huge block, so is the volume."		

MAT_0006

Ms. Ellis wants to introduce the concept of density of solid materials to her students. She is considering using the diagrams below to help her students begin to understand differences in the density of various solid materials. These two-dimensional diagrams are representations used to show the particles in solid matter. Each circle represents one particle in the solid material while each diagram represents a different solid material.



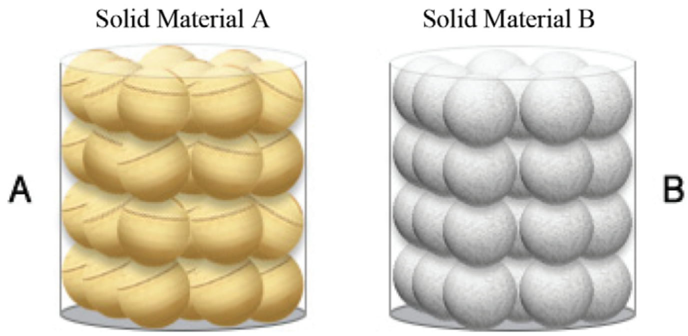
Ms. Ellis wants to use these diagrams to help her students understand the differences in the density of various materials.

Which set of diagrams could be used to help students understand that differences in the arrangement of particles account for differences in the density of various solid materials?

A. Diagram A and Diagram B
B. Diagram A and Diagram C
C. Diagram B and Diagram D
D. None of these diagrams could be used to help students understand that differences in the arrangement of particles account for differences in the density of various solid materials.

MAT_0007

Mr. Ohm decides to use a model to help his students understand why the densities of solid materials differ. In his model, Mr. Ohm uses different types of balls to represent what two different solid materials (A and B) are composed of—wooden balls for solid material A and foam balls for solid material B. The balls in each container represent the particles in each solid material. He uses the same number and size of solid balls in two identically sized, clear cylindrical containers.



After presenting the two cylinders to his students and passing them around for the students to hold, Mr. Ohm asks his students to use the model to explain why solid materials can have different densities.

Which of the following students shows the best understanding of how Mr. Ohm's particle model helps explain why these two solid materials (Solid Material A and Solid Material B) have different densities?

A. Devon: "This model shows me how even with different solids the particles can be arranged in the same way."
B. Sarah: "This model helps me see how different types of particles can be the same size."
C. Jonah: "This model helps me see how the particles of different solids are compacted differently."
D. Terence: "This model shows me that particles of different solids can have different masses."

MAT_0009

Mr. Liu is teaching a lesson to his fifth-grade class about measuring the weight of matter. The students are working in small groups to weigh different solid objects using a variety of instruments.
One student, Jessica, asks a question: "On TV last night, I saw how astronauts jump around so easily on the Moon. My mom says that they can do that because they weigh less up there. How come they weigh less in space?"
To answer Jessica's question, Mr. Liu decides to explain the difference between the weight and mass of an object on the Moon and on Earth.

Write a response that Mr. Liu could use to answer Jessica's question. In your response, make sure to explain the difference between the weight and mass of an object on the Moon and on Earth.
The response should address the underlying science and be accessible to fifth-grade students.

Response Box:

MAT_0010

Ms. Chen is teaching a series of lessons on using different methods to determine volume. Today, her students will be using a mathematical formula to calculate the volume of rectangular prisms. After some initial whole-class instruction, she has the students move into triads and gives each triad the same size prism and a ruler, like the one below. Ms. Chen asks them to solve the following problem.

"Measure the prism using your rulers and determine its volume in cubic centimeters using the formula: V = area of base x height."

(The rectangular prism has a height of 6 cm and square bases with 4 cm sides.)

Two triads of students in the class get the same wrong answer of 175 cubic cm. Which of the following is the most likely error the six students made?

A. Measuring the sides and height by starting at the 25 cm mark instead of at the 0 cm mark.
B. Measuring the sides and height by starting at the 1 cm mark instead of at the 0 cm mark.
C. Measuring the sides and height by using inches instead of centimeters.
D. Measuring the sides and height by starting from the very left end of the ruler, prior to the 0 cm mark.

MAT_0011a

Ms. Andujar's class is measuring volumes of irregular solid objects using the water displacement method. At the beginning of the lesson, she wants to check the students' ability to accurately read volume amounts in a glass graduated cylinder.

She places three 100 ml graduated cylinders at a station in the classroom. One cylinder has 29 ml of water in it, the second has 44 ml of water in it, and the third has 89 ml of water in it.

The students take turns reading the amounts of water in each cylinder, and then they record their results and turn them in to Ms. Andujar.

A quarter of the students in the class get the same wrong answers: 30 ml, 45 ml, and 90 ml.

For each of the following options, indicate whether it is a reason that these students could be incorrectly reading the amounts of water in the glass graduated cylinders.

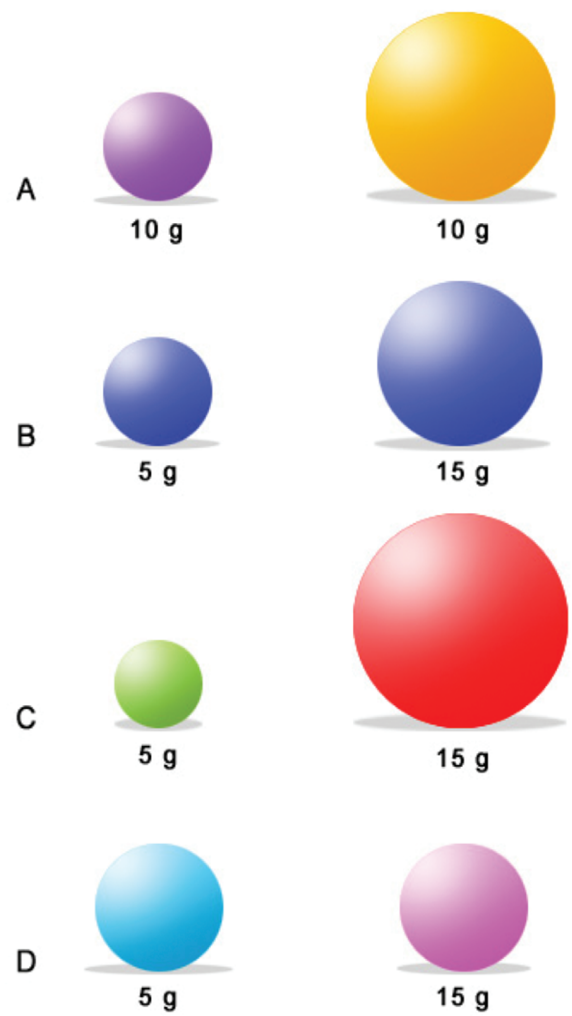
	Reason for Incorrect Measurement	**Not a Reason for Incorrect Measurement**
A. Reading the amounts at the top of the meniscus		
B. Reading the amounts at the bottom of the meniscus		
C. Reading the amounts by looking only at the nearest 5 ml tick marks		

MAT_0013

As part of his science unit on the structure and properties of matter, Mr. Singh teaches a series of lessons about density. He remembers from his work with previous classes that students understand the difference between the volume and density of solid materials, but sometimes they fail to distinguish between the mass and density of solid materials. So, at the end of this lesson series, Mr. Singh decides that he wants to check whether any students confuse the relationship between mass and density of solid materials.

Mr. Singh shows his students a set of two spheres and asks them to predict (without holding the spheres) which one, if any, is denser. He mentions that the two spheres could have the same density. He tells them the mass of each sphere and that each color represents a different solid material.

Which of the following sets of spheres is most useful in determining which students confuse the relationship between mass and density of solid materials?
Note: The mass of each sphere is given, and each color represents a different solid material. Also, the relative sizes of the spheres in each set are as they appear.



Appendix C

Ecosystems Practice-Based Content Knowledge for Teaching Items and Item Key

Item ID	Item key
ECO_0014	A
ECO_0015	C
ECO_0016a_Row1	Will not need
ECO_0016a_Row2	Will not need
ECO_0016a_Row3	Will need
ECO_0016a_Row4	Will need
ECO_0017	C
ECO_0019a	B
ECO_0021	A
ECO_0022_Row1	Inaccurate
ECO_0022_Row2	Accurate
ECO_0022_Row3	Inaccurate
ECO_0022_Row4	Inaccurate
ECO_0022_Row5	Accurate
ECO_0022_Row6	Inaccurate
ECO_0023b	A
ECO_0026 _Accessibility	See Appendix F
ECO_0026 _Accuracy	See Appendix F
ECO_0028a	D
ECO_0028b	A

ECO_0014

Ms. Krauser's fifth-grade students have been investigating different ecosystems and the interdependent relationships between organisms in ecosystems. In today's lesson, Ms. Krauser wants her students to focus on the role of carnivores in an ecosystem. Since this lesson is the first one in the unit that specifically addresses carnivores, the teacher asks students to get into small groups to discuss their current understanding of carnivores' role within an ecosystem. Ms. Krauser walks around the classroom and listens to each small group's conversation for a few minutes to get a sense of her students' current understanding.

Which of the following student comments demonstrates the best understanding of the role carnivores play within an ecosystem?

A. Juan: "Carnivores are at the top of the food chain and have small populations compared to the organisms they eat."
B. Brigid: "There are no animals that eat carnivore animals, so they are sort of the rulers of nature and food webs."
C. Thomas: "Carnivores chomp on everything else inside an ecosystem, so they are worry free."
D. Lei: "If something crazy happened to kill all the plants, carnivores could still live on forever, because they eat only animals."

ECO_0015

Mr. Griffith is beginning his new science unit on plant growth and development by having his fourth-grade students discuss how plants get their food. As students share their ideas in small groups, Mr. Griffith walks around the room and listens in on their conversations in order to assess his students' beginning ideas.

Which of the following student comments demonstrates the best understanding of how plants get their food?

A. Tino: "Plants suck up nutrients from the soil."
B. Brianna: "Plant roots take in things like water and minerals to help plants grow big and tall."
C. James: "Plants use water and air to make their own food."
D. Lara: "Plants get their food from a lot of different places—water, soil, fertilizer, and the air."

ECO_0016a

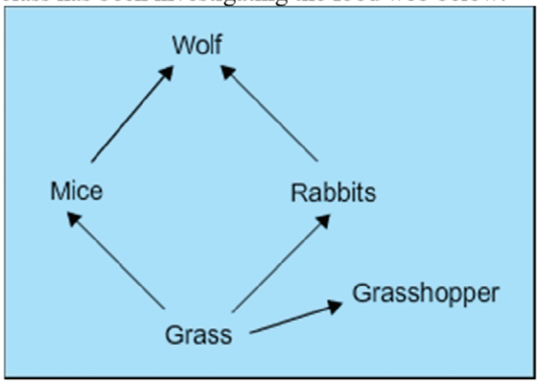
Through a series of lessons, Mr. Yamaguchi's students have been learning about woodland decomposers and their food sources. In order to check for student understanding, Mr. Yamaguchi asks his students to complete an exit ticket near the end of a lesson. In this short assessment, groups of students need to give an example of at least one woodland decomposer and describe in a sentence how the decomposer gets its food.

In the next day's lesson, Mr. Yamaguchi plans on providing additional assistance to each group whose response shows a weak understanding of decomposers. Based on the groups' responses, identify whether the student groups will need additional support in the next day's lesson. For each group, place an X in the appropriate box.

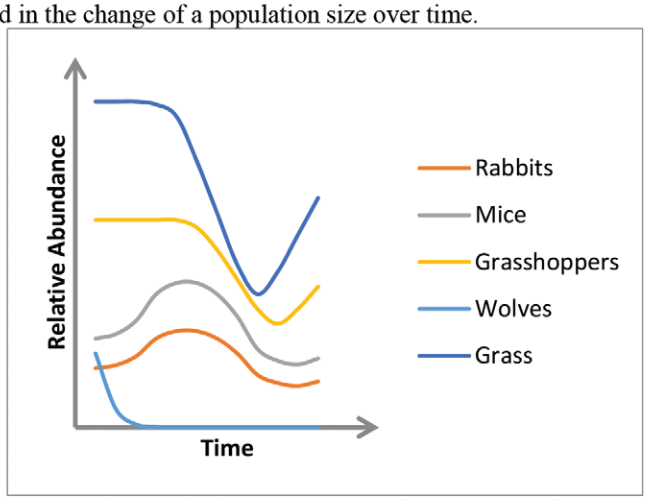
Student Group	Decomposers and Their Food Sources	Will Need Additional Support	Will Not Need Additional Support
1	Mushrooms and other fungi are examples of decomposers in the woodlands. To get their food, they munch on dead plants and animals by breaking them into very small pieces.		
2	Bacteria are decomposers in the woodland areas. They break down dead animals and plants to get their food.		
3	In the woodlands there are lots of mushrooms, which are decomposers. They make their own food using sunlight, air, and water.		
4	Bacteria are a type of a woodland decomposer. They create their own food supply with the Sun, water, and air.		

ECO_0017

Ms. Goldberg has been using food webs to help her students learn about the interdependent relationships among organisms in ecosystems. Her class has been investigating the food web below:



After discussing the relationships among the grass, grasshoppers, rabbits, mice, and wolves in this ecosystem, Ms. Goldberg gives her students a culminating unit task that uses a virtual simulation. In this simulation, students can manipulate the number of organisms in each population by dragging different slider buttons in the simulation (e.g., they can change the number of organisms in the rabbit population by moving the rabbit slider button from very high to very low). Ms. Goldberg asks a group of students to use the slider button to decrease the wolf population dramatically. They are able to use the graph produced below to see what happens to the size of all the populations in the food web as a result of this change to the wolf population. The lines in the graph represent the overall trend in the change of a population size over time.



At the end of the activity, Ms. Goldberg asks her students to write an explanation that accounts for the patterns in the rabbit and mice populations after the wolf population starts to crash.

Based on the written student explanations below, which student has the best understanding of the interdependent relationships of organisms in this ecosystem?

A. Ravi: “When the wolves decrease, the rabbits and mice increase at first because they no longer have a predator to eat them. But then they decrease because they get into fights over the grass and kill each other off.”

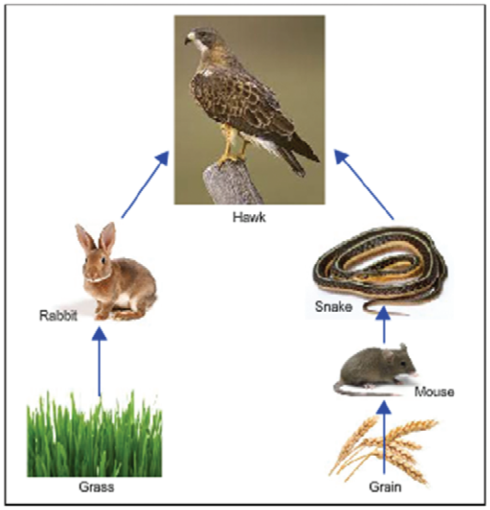
B. Maria: “When the wolves crash, that leaves more rabbits and mice around at first. Then the rabbits and mice decrease because after they eat all the grass they start to eat the grasshoppers, which might be toxic to them.”

C. Kayla: “As the wolves die off, the rabbits and mice increase at first because they no longer have a predator to eat them. But that means the rabbits and mice eat more grass, and when the grass is very low, there is not enough food for both the rabbit and mice, and they decrease.”

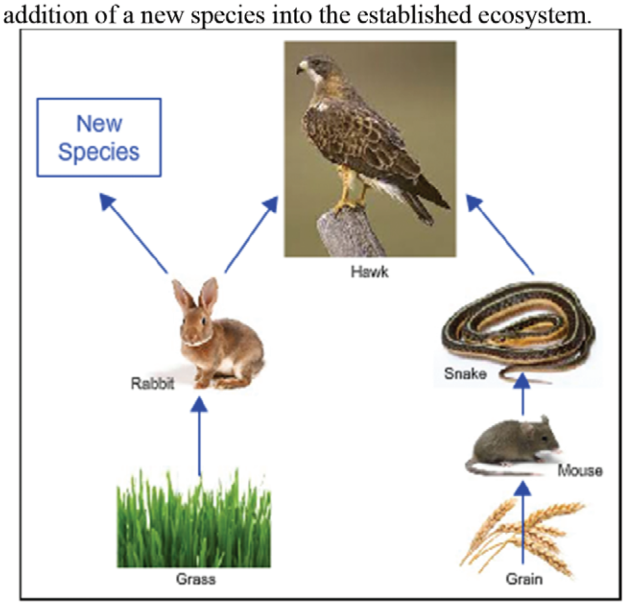
D. Pedro: “When the wolves go down, all the populations go down, because wolves prey on all the organisms in the food web and they are at the top of the food chain.”

ECO_0019a

His students have learned about food webs, and now Mr. Hasan's class is investigating the consequences of introducing a new species into an established ecosystem.
Initially, Mr. Hasan presents them with the food web below.



Then, he shows them another food web (see below). The second food web is similar to the first one, except that the second one shows the addition of a new species into the established ecosystem.



Mr. Hasan finds that many students incorrectly conclude that the introduction of a new species in the ecosystem will result in no change in the population of mice.

Which of the following common student errors most likely leads to the students' incorrect conclusion?

A. Students are thinking that top predators eat everything in a food web.
B. Students are focusing just on direct, one-step connections in the food web.
C. Students are thinking about only individual organisms rather than an entire population.
D. Students are interpreting the arrows in the food web as going from predator to prey.

ECO_0021

Mr. Suthers' class has been investigating plant growth and development and he has been focusing on helping his students understand how plants get their food. To get a sense of their current understanding, Mr. Suthers asks students to develop a model for how plants get their food that shows the necessary inputs and outputs involved in the process.

Model for How Plants Get Their Food
Input → (Plant) → Output

Which of the following students demonstrates the best understanding of how plants get their food?

	Student Models for How Plants Get Their Food		
Student	**Inputs →**	**Plant**	**→ Outputs**
Student A	carbon dioxide, H_2O, sunlight		gas, food
Student B	gas, water, minerals		O_2, sugar
Student C	CO_2, water, plant food		oxygen, sugar
Student D	air, H_2O, soil, sunlight		food

A. Student A
B. Student B
C. Student C
D. Student D

ECO_0022

In her unit on food webs, Ms. Fox is considering using different comparisons to help her students better understand food sources for less familiar organisms. In order to do this, she wants to use examples of food sources for organisms that her students are more familiar with and compare them to the food sources for less familiar organisms.

For each of the following statements, indicate whether or not it is an accurate example that the teacher could use to help her students understand food sources for less familiar organisms.

Example	**Accurate Example**	**Inaccurate Example**
Grass is food for rabbits as sunlight is food for plants.		
Berries are food for deer as dead trees are food for fungi.		
Rabbits are food for hawks as soil is food for plants.		
Deer are food for wolves as water is food for plants.		
Rodents are food for snakes as sugars are food for plants.		
Worms are food for birds as heat is food for bacteria.		

ECO_0023b

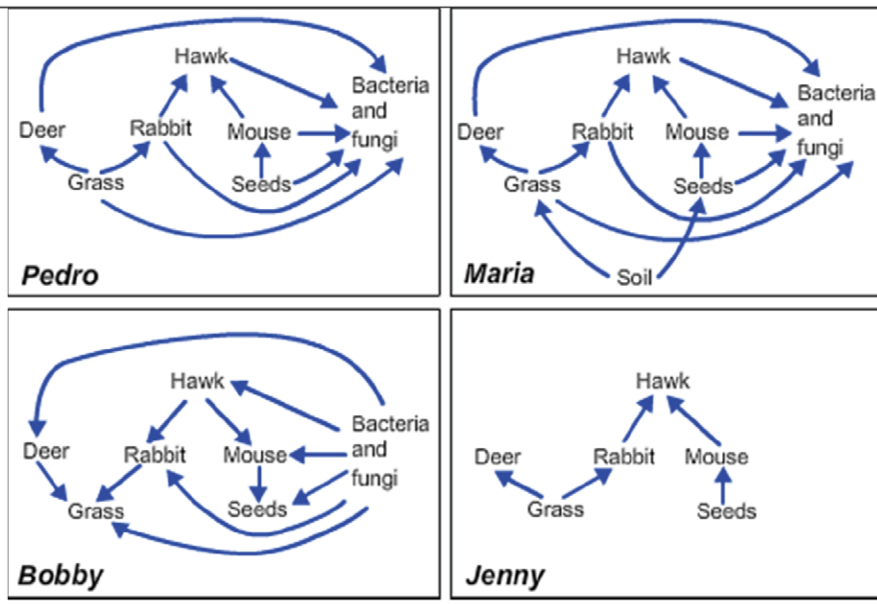
Mr. Patel is teaching a unit on ecosystems and his class has been discussing the roles of producers, consumers, and decomposers in the transfer of matter.

Mr. Patel has given his class an assignment: to construct a food web from the potential components of an ecosystem listed below. Students can choose to include some, but not all of the components, in their food webs.

Hawk
Rabbit
Mouse
Deer
Grass
Seeds
Bacteria and fungi
Soil

Mr. Patel asks the students to create a food web to represent the transfer of matter in this ecosystem. He reminds them that the arrows indicate that one component is a food source for the next component.

Below are four students' food webs. Based on their food webs, which student has the best understanding of the transfer of matter in the ecosystem from the perspective of what the sources of food are and where the food goes?

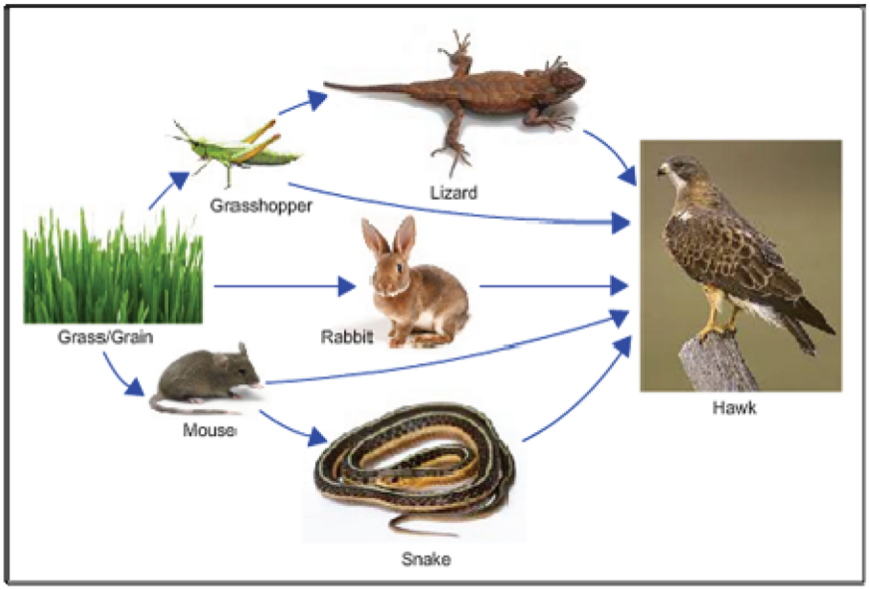


ECO_0026

The students in Ms. Colt's fifth-grade class have been conducting multiple controlled experiments to learn about plant growth. Through their experiments, they found that plants do not continue to grow without sunlight, water, or air. Ms. Colt explained the experiments provide evidence that plants use sunlight, water, and air to make their own food. Some of the students wondered about how fertilizer helps plants grow. To examine the effects of fertilizer on plant growth, the students have just conducted another investigation to compare plants grown with and without fertilizer. The students observed that the plants grown with fertilizer looked very healthy with green leaves and thick, strong stems while the plants grown without fertilizer had leaves that were turning yellow and thin, weak stems. Since the fertilizer helped the plants grow better, some students have concluded that fertilizer is also food for plants. During the next few lessons, Ms. Colt will have her students further investigate plant growth, and she plans on explaining why fertilizer does help plants grow but is not considered food for plants.
Write a brief explanation that the teacher could give as part of her next lesson to explain why fertilizer does help plants grow but is not considered food for plants. The explanation should address the underlying science and be accessible to fifth-grade students.
Enter your response in the box below:

ECO_0028a

Ms. Rivera is going to teach a lesson that focuses on the connections between organisms in an ecosystem, and she plans on beginning her lesson by showing her students the food web below.



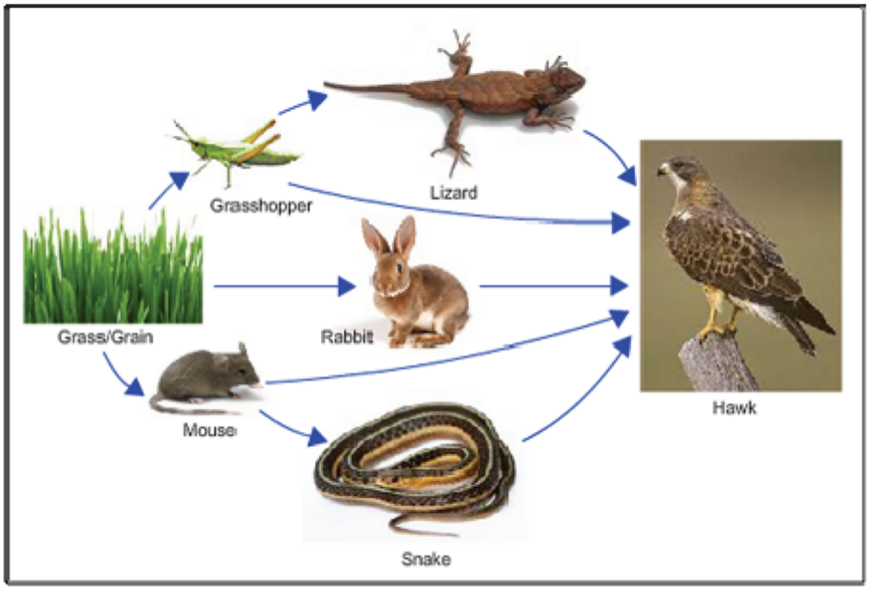
Ms. Rivera knows that there are strengths and limitations of using food webs to model the connections between organisms in an ecosystem.

Which of the following statements is a strength of using food webs, such as the one pictured above, to model connections between organisms in an ecosystem?

A. All of the possible feeding relationships within the ecosystem can be seen.
B. The food web shows the relative sizes of the population of each organism.
C. Top-level consumers are shown as directly feeding on every organism.
D. The transfer of matter can be followed throughout the entire food web.

ECO_0028b

Ms. Rivera is going to teach a lesson that focuses on the connections between organisms in an ecosystem, and she plans on beginning her lesson by showing her students the food web below.



Ms. Rivera knows that there are strengths and limitations of using food webs to model the connections between organisms in an ecosystem.

Which of the following statements is a limitation of using food webs, such as the one pictured above, to model connections between organisms in an ecosystem?

A. An entire population of organisms is represented by an image of an individual from that population.
B. Some organisms are shown as the food source for many different organisms.
C. Arrows are used to show the direction that matter flows to the various feeding levels.
D. Some organisms are shown consuming at more than one feeding level.

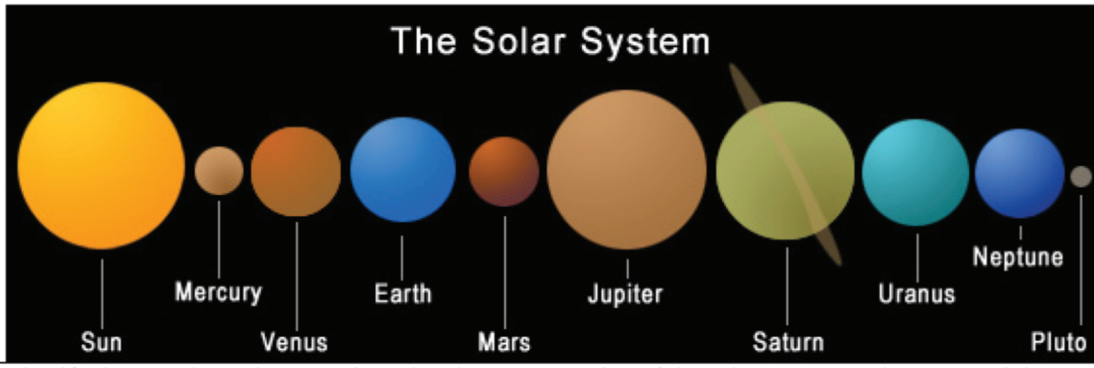
Appendix D

Earth's Place Practice-Based Content Knowledge for Teaching Items and Item Key

Item ID	Item key
EAR_0029a	See Appendix G
EAR_0030	A
EAR_0031_Row1	Yes
EAR_0031_Row2	No
EAR_0031_Row3	No
EAR_0031_Row4	Yes
EAR_0032a	C
EAR_0034_Row1	Effective
EAR_0034_Row2	Not effective
EAR_0034_Row3	Effective
EAR_0034_Row4	Not effective
EAR_0035_Row1	Does not demonstrate
EAR_0035_Row2	Demonstrates
EAR_0035_Row3	Does not demonstrate
EAR_0035_Row4	Does not demonstrate
EAR_0035_Row5	Demonstrates
EAR_0036a	C
EAR_0036b_Row1	A, C, D
EAR_0036b_Row2	A, C
EAR_0036b_Row3	A, B
EAR_0036b_Row4	none
EAR_0038_Accessibility	See Appendix H
EAR_0038_ Accuracy	See Appendix H
EAR_0039_Row1	C
EAR_0039_Row2	D
EAR_0039_Row3	B
EAR_0039_Row4	A
EAR_0040_Row1	B
EAR_0040_Row2	A
EAR_0040_Row3	C
EAR_0040_Row4	D

EAR_0029a

Ms. Washington will be teaching a series of lessons on the solar system—the basic components, the orbits of different objects around the Sun, and the subsequent observable patterns. She is looking for a representation of the solar system that will help her when teaching the lessons. In her search, Ms. Washington comes across the following poster of the solar system and notices that the representation on the poster has some weaknesses.



Identify three student misconceptions that the representation of the solar system on the poster might cause or reinforce.

Enter your response in the box below:

EAR_0030

Ms. Garcia is beginning some lessons about the patterns in the Moon's appearance across the day, month, and year. In order to plan her lessons, she first wants to find out their current ideas about the motion of the Moon every 24 hours. Ms. Garcia asks her students the following question.

> If you are looking at the Moon from your house, which of the following happens to the Moon every 24 hours? Select all that apply.
> ☐ The Moon can be seen in the sky only during the night.
> ☐ The Moon can be seen in the sky sometimes during the day and sometimes at night.
> ☐ The Moon appears to rise and set in the sky approximately every 24 hours.
> ☐ The Moon always appears in the same position in the sky.

Based on the following student responses that Ms. Garcia collected, which student has the best understanding of what happens to the Moon every 24 hours?

Julia:
☐ The Moon can be seen in the sky only during the night.
☒ The Moon can be seen in the sky sometimes during the day and sometimes at night.
☒ The Moon appears to rise and set in the sky approximately every 24 hours.
☐ The Moon always appears in the same position in the sky.

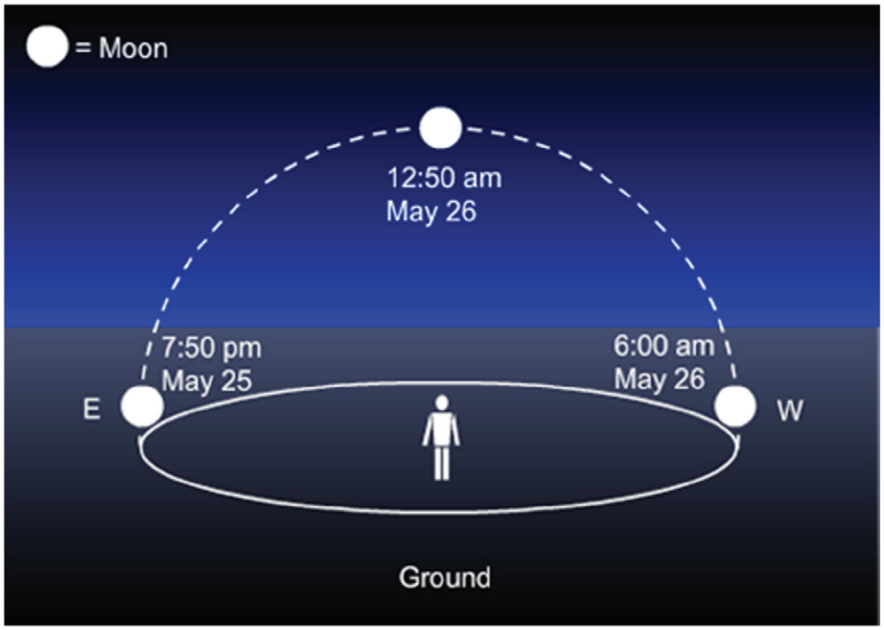
Bobby:
☒ The Moon can be seen in the sky only during the night.
☐ The Moon can be seen in the sky sometimes during the day and sometimes at night.
☐ The Moon appears to rise and set in the sky approximately every 24 hours.
☒ The Moon always appears in the same position in the sky.

Toby:
☐ The Moon can be seen in the sky only during the night.
☒ The Moon can be seen in the sky sometimes during the day and sometimes at night.
☐ The Moon appears to rise and set in the sky approximately every 24 hours.
☒ The Moon always appears in the same position in the sky.

Yuwei:
☒ The Moon can be seen in the sky only during the night.
☐ The Moon can be seen in the sky sometimes during the day and sometimes at night.
☒ The Moon appears to rise and set in the sky approximately every 24 hours.
☐ The Moon always appears in the same position in the sky.

EAR_0031

Ms. Chang is teaching a lesson on how the position of the Moon changes daily. She first asks her students to predict what they think happens to the full Moon from evening to morning. Then, Ms. Chang shows her students a drawing representing data that she downloaded from a Web site and asks them whether the data is consistent or inconsistent with their predictions.
The drawing representing the data is shown below.



The drawing represents a field of view in the sky and shows a person positioned in the center, standing on the ground. The three time stamps represent when the person observed the full Moon.

Based on the responses below, which students have effectively used the data to evaluate their predictions? Select <u>all</u> that apply.

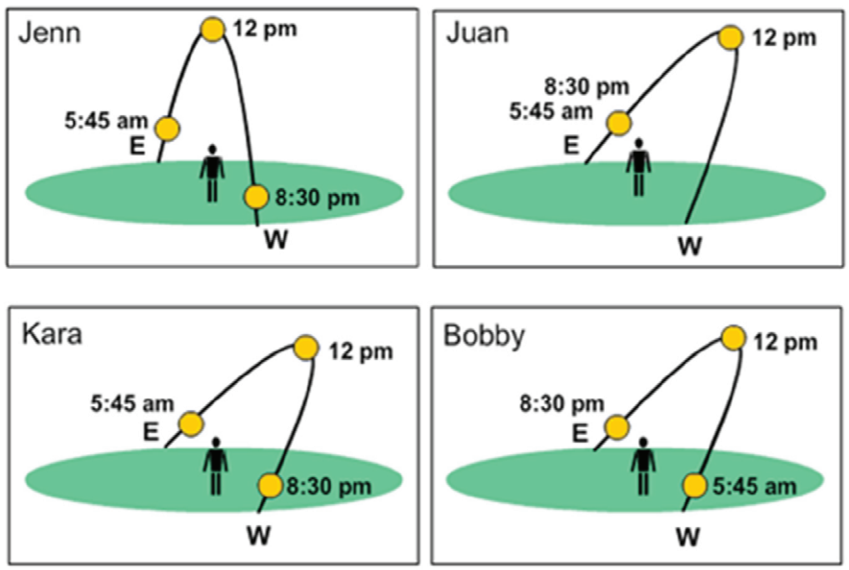
- ☐ Kelly: "I predicted that the full Moon would rise, move across the sky, and set, just as the Sun rises and sets (rises in the east and sets in the west). These data confirm my prediction."
- ☐ Sammy: "I predicted the full Moon would stay in one position in the sky from evening to morning. These data are consistent with my prediction."
- ☐ Kasandra: "I predicted the full Moon would move across the sky and set in the opposite direction of the Sun by rising in the west and setting in the east. These data confirm my predicition."
- ☐ Jordan: "I predicted the full Moon would rise on one side of the sky, move half way across the sky, and then set in the same position that it rose. These data are inconsistent with my prediction."

EAR_0032a

Mr. Calderon's science class is investigating the Sun's changing position in the sky. He decides to first probe his students' ideas about the Sun's position across the sky in one day. Mr. Calderon gives his students the following task to complete.

> From a set of four diagrams, select the diagram that best represents the Sun's position in the sky at 5:45 A.M., 12:00 P.M., and 8:30 P.M. on June 21st in the continental United States. Each diagram shows an observer standing on the ground. Direction is indicated by "W" for west on one end of the arc and by "E" for east on the other end of the arc. All times are Standard Time.

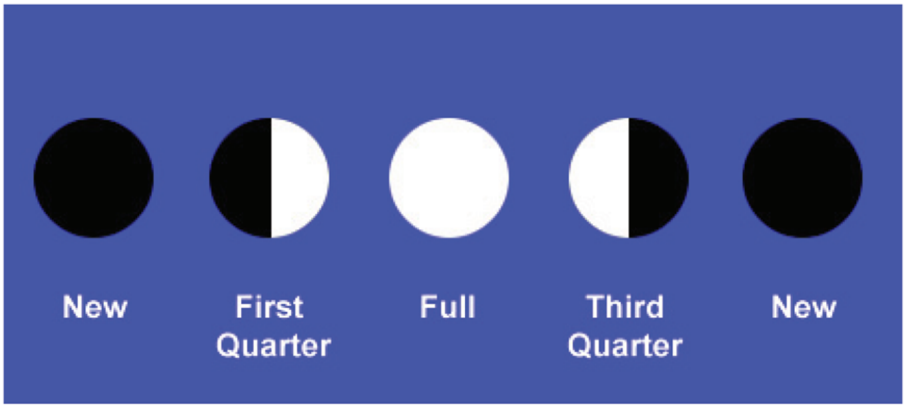
Each of the four students below selected a different diagram. Which student has the best understanding of the Sun's position across the sky in one day?



EAR_0034

Mr. Nguyen is teaching some lessons on changes in the appearance of the Moon over time. Before having his students look at data, he asks them to predict the observable monthly patterns of change in the Moon's appearance. Then, Mr. Nguyen gives his students the data below showing how the Moon appears to change from new Moon to full Moon and back to new Moon each month.

He tells the students that in the diagram, The Phases of the Moon, the white portion represents the part of the Moon that is visible from Earth.



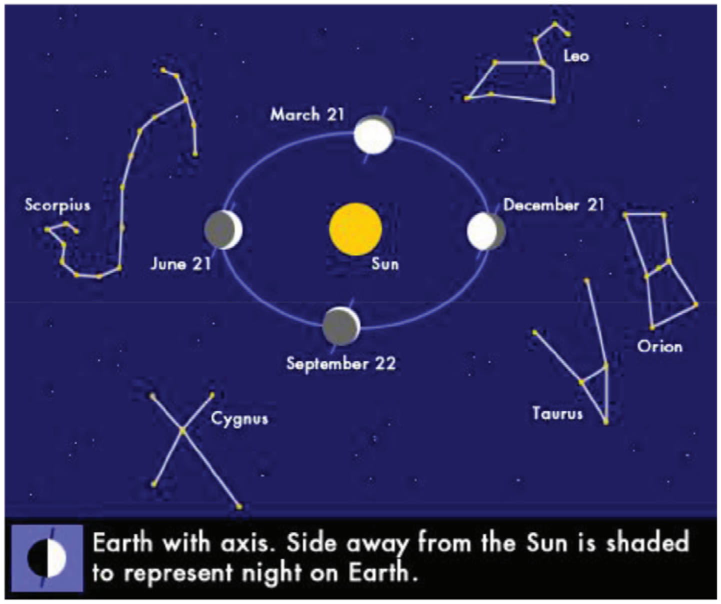
Mr. Nguyen then asks his students to explain whether the data support their predictions.

Four students gave the explanations below. For each student, determine whether he or she effectively used the data to evaluate his or her prediction. For each student, place an X in the corresponding box.

Student	Effectively Used the Data to Evaluate His or Her Prediction	Did Not Effectively Use the Data to Evaluate His or Her prediction
Larry: "I said that the full Moon should be completely shaded**,** the new moons should have no shading, the first quarter Moon should be half-shaded on the left, and the third quarter Moon should be half-shaded on the right. The data do not support my prediction."		
Mattias: "I thought that the new moons ought to be completely shaded, the full Moon should have no shading, the first quarter Moon should be shaded on the left half, and the third quarter Moon should be shaded on the right half. The data do not support my prediction."		
Tina: "I said that the new moons need to be completely shaded, the full Moon should have no shading, the first quarter moon should be half-shaded on the left, and the third quarter moon should be half-shaded on the right, as in the diagram. The data support my prediction."		
Jing: "I thought that the full Moon should be completely shaded, the new moons ought to have no shading, the first quarter Moon should be half-shaded on the right side, and third quarter Moon should be half-shaded on the left side. The data support my prediction."		

EAR_0035

Mr. Kurk is introducing a lesson about the seasonal appearance of some stars in the night sky. He groups his students into triads and gives each group a copy of the following diagram.



Mr. Kurk explains that the diagram shows some seasonal constellations that are visible from the Northern Hemisphere. For example, Scorpius is only visible in the summer, and Orion is only visible in the winter. He writes the following question on the board.

Why are we able to see only certain constellations during specific times of the year?

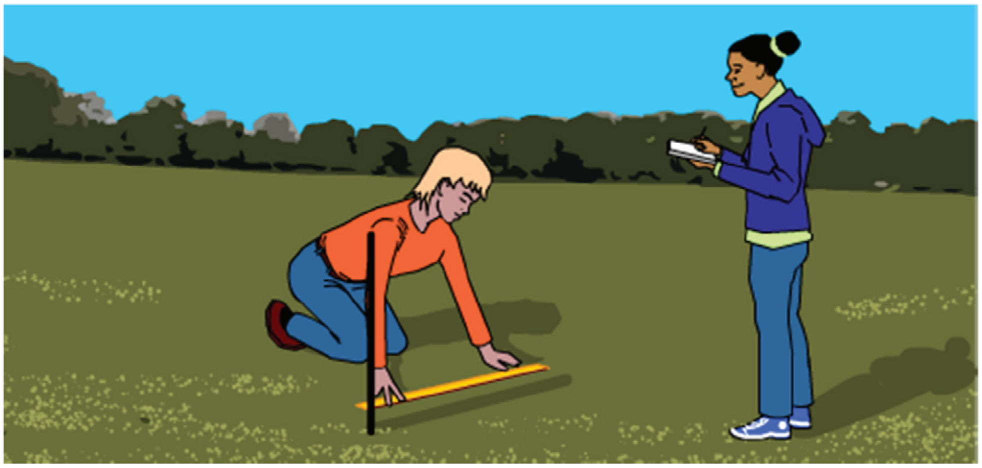
As he walks around the classroom to the various triads, Mr. Kurk listens to conversations and writes down some of the student responses.

For each of the following student responses, indicate whether or not it demonstrates an understanding of the scientific phenomenon central to Mr. Kurk's question. For each response, place an X in the corresponding box.

Student Responses	Demonstrates an Understanding of the Scientific Phenomenon	Does <u>Not</u> Demonstrate an Understanding of the Scientific Phenomenon
Aiden: "Because in winter the Earth is tilted back on its axis away from the Sun, but in summer the Earth's tilt changes, exposing other parts of the universe."		
Briniqua: ". . . because at different times of the year, Earth's night side faces different directions."		
Charlie: "As Earth rotates, people see different objects in the night sky."		
Devin: "The stars are always in motion, and they all move together in the same patterns around Earth."		
Evelynne: "It's similar to a girl sitting facing out on the edge of a merry-go-round. Her back is facing the center, the Sun, and her head is facing outward toward the night sky. As the merry-go-round turns and she goes around the Sun, different objects come into her view."		

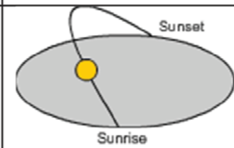
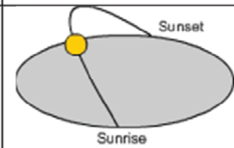
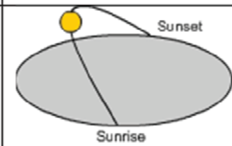
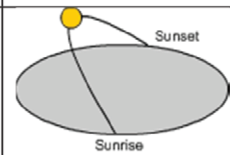
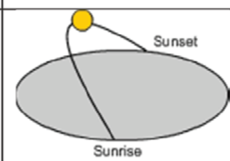
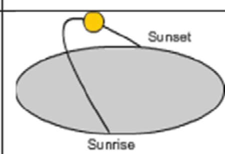
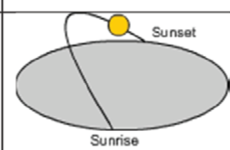
EAR_0036a

Mr. Yamamoto is having his students investigate the relationship between the Sun's position in the sky and the length and position of shadows on the ground. He gives each pair of students a meter stick and a measuring tape. Then he directs them to find a place outside that will get Sun all day, to put one end of the stick in the ground so it can stand upright on its own, and to measure the length of the stick's shadow in centimeters every hour throughout the school day.



Mr. Yamamoto discusses with his students the importance of not looking at the Sun while they are outside taking their shadow measurements. Instead, in order to learn about the Sun's position in the sky, Mr. Yamamoto tells his students that he will provide them with a completed Sun path diagram each time they take a shadow measurement (see the third column in the chart below). Each Sun path diagram has a line representing the Sun's apparent path from sunrise to sunset and a small yellow circle representing the Sun. The yellow circle (the Sun) is at a precise place along the line to show where the Sun is located in the sky at a specific time.

Students record their shadow measurements each hour and then return to the classroom to enter their data in a spreadsheet that calculates the hourly average shadow length across the class data. The data on shadow lengths collected by the students and the data on the Sun's position provided by the teacher are listed in the chart below.

Standard Time	**Students' Collected Data: Shadow Length** (centimeters)	**Data Provided by Teacher: Sun's Position**
9:00 A.M.	267 cm	
10:00 A.M.	196 cm	
11:00 A.M.	163 cm	
12:00 P.M. (noon)	158 cm	
1:00 P.M.	173 cm	
2:00 P.M.	218 cm	
3:00 P.M.	340 cm	

The day after the students' investigations, four students tell the class how they think the Sun's position affects shadows. Based on the student comments below, which student demonstrates the least understanding of how the Sun's position in the sky has an effect on the shadows it casts during the day?

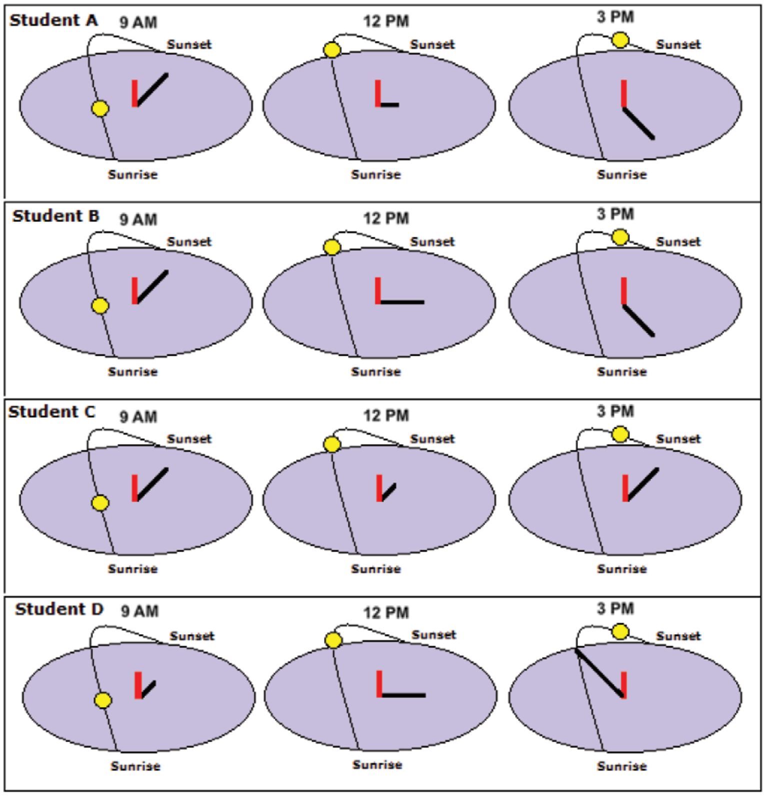
A. **Addy:** "Shadows are shortest around 12:00 P.M., because that is when the Sun is highest in the sky."
B. **Ben:** "I think that at 4 P.M. shadows will be longer than at 3 P.M., because the Sun is lower in the sky at 4 P.M."
C. **Carolyn:** "Shadows get longer throughout the day as the Sun's position changes."
D. **David:** "The lower the Sun is in the sky, the longer the shadows are."

EAR_0036b

For the last few weeks, students in Mr. Yamamoto's class have been investigating the relationship between the Sun's position in the sky and the length and position of shadows on the ground. In order to check for student understanding near the end of the lesson series, Mr. Yamamoto asks his students to complete an exit ticket. In this short assessment, Mr. Yamamoto has his students use Sun path diagrams to draw the relative lengths and positions of the shadows that would be cast by the Sun at 9 a.m., 12 p.m., and 3 p.m.

In the Sun path diagrams below, the red line represents a meter stick, and the black line represents the shadow cast by the Sun. In addition, each diagram has a line representing the Sun's apparent path from sunrise to sunset and a small yellow circle representing the Sun. The yellow circle (the Sun) is at a precise place along the line to show where the Sun is located in the sky at a specific time.

The Sun path diagrams below contain the drawings of four students.



In the chart below, select the student responses, if any, that provide evidence that the student understands the concept indicated. For each concept, select all that apply. Evaluate each concept separately.

Concepts	Student Responses				
	A	B	C	D	None of the responses
1) Shadows' lengths change as the Sun's position changes.					
2) Shadows are longer when the Sun is positioned lower in the sky.					
3) The shadows cast by the Sun are in the opposite direction from the Sun's position in the sky.					
4) Patterns in the daily changes to the length and direction of shadows are caused by Earth's rotation.					

EAR_0038

During a lesson about the orbit of Earth around the Sun, Ms. Van Buren's fifth-grade students are working in pairs to analyze data from various Web sites.

As part of this work, the students are looking at pictures and diagrams showing the orbits of different planets around the Sun. One of the students, Shelton, notices that the Sun, which is one of the stars in the solar system, and the planets look fairly similar in these pictures and diagrams. He wonders why that is and asks Ms. Van Buren, "What is the difference between a star and a planet?"

Write an explanation that Ms. Van Buren could give to answer Shelton's question. The explanation should address the underlying science and be accessible to fifth-grade students. Be sure that the explanation includes two distinct differences between a star and a planet.

Question: "What are at least two distinct differences between a star and a planet?"

Response Box:

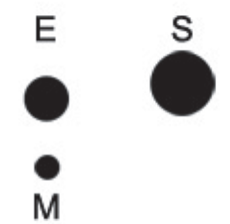
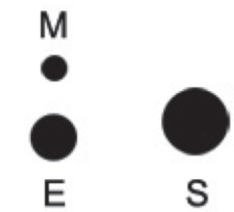
EAR_0039

Mr. Harrison's class has been investigating the basic lunar phases. To check for understanding, he asks his students to complete the following exit ticket at the end of class.

Please draw two diagrams. For the first diagram, show the positions of Earth, the Moon, and the Sun during a new Moon. For the second diagram, show the positions of Earth, the Moon, and the Sun for a full Moon. Be sure to label Earth (E), the Moon (M), and the Sun (S) in each of your diagrams.

The diagrams from four of Mr. Harrison's students are below.

	Moon Phase	
Student	**New**	**Full**
Shana		
Lei		
Bret		

Ahmad			

After seeing the students' diagrams, Mr. Harrison gives each student a feedback prompt to address what is shown in the student's response. Match each feedback prompt with the student for whom it is most likely to be helpful. Each student should be matched to one of the feedback prompts.

"What is special about the alignment of Earth, the Moon, and the Sun during the new Moon and full Moon phases?"

"What object in the sky does the Moon revolve around?"

"Why can we not see the Moon during the new Moon phase?"

"Now can you diagram the positions of Earth, the Moon, and the Sun during the first quarter Moon phase?"

Click on a feedback response and then drag to the appropriate box below.

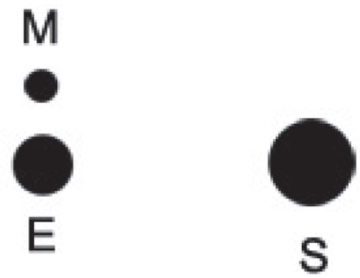
Student	**Feedback**
Shana	
Lei	
Bret	
Ahmad	

EAR_0040

Mr. Brenner's class has been investigating the basic lunar phases. To check for understanding, he asks his students to complete the following exit ticket at the end of class.

Draw a diagram that shows the positions of Earth (E), the Moon (M), and the Sun (S) during the first quarter Moon phase.

The diagrams from four of Mr. Brenner's students are below.

Student	First Quarter Moon Phase
Jana	
Tia	
Lucius	
Marcus	

After seeing the students' diagrams, Mr. Brenner gives each student a feedback prompt to address what is shown in the student's response. Match each feedback prompt with the student for whom it is most likely to be helpful. Each student should be matched to one of the feedback prompts.

"Why can we not see the Moon during the new Moon phase?"

"Now can you diagram the positions of Earth, the Moon, and the Sun during the third quarter Moon phase?"

"Why do we see the whole Moon disc lit up during the full Moon phase?"

"What object in the sky does the Moon revolve around?"

Click on a feedback response and then drag to the appropriate box below.

Student	**Feedback**
Jana	
Tia	
Lucius	
Marcus	

Appendix E

Constructed-Response Scoring Rubric for Matter Question 0009 (Weight and Mass Explanation)
Holistic Rubric 1: Scientific Accuracy
(To what extent does the response address the underlying science accurately and completely?)

Level (score)	Description	Exemplar
Complete (2)	Explanation includes 2 key ideas ([1] weight is different because there is less gravity on the Moon; [2] mass of an object would be the same on Earth and any smaller object) to explain why the weight of a person is different on the Moon and describe the difference between the weight and mass of an object on the Moon and on Earth. Explanation includes no major misconceptions.	Mass is the measurement of how much matter is in an object. It is the combined measure of the number of atoms in an object combined with the density of those atoms. Mass does not change because the amount of matter inside you will not change unless matter is taken away. Whether you travel to the Moon or stay here on Earth, your mass will remain the same. Weight is different from mass. Weight is the measure of the force of gravity on an object. The mass of an object will never change, but the weight of an object can change based on the location of the object being measured. For example, you may weigh 60 pounds on Earth, but on the Moon you would only weigh 10 pounds because the gravity of the Moon is 1/6th that of the Earth. However, you will always have the same mass on Earth as you have on the Moon.
Partial (1)	Explanation includes 1 key idea (idea 1 or 2) to explain why the weight of a person is different on the Moon and describe the difference between the weight and mass of an object on the Moon and on Earth. Explanation includes no major misconceptions.	Mass is the same wherever you are—whether it's on the Moon or on Earth—because the amount of stuff you are made of does not change. Weight depends on how much gravity is acting on you. So you would weigh less on the Moon than on Earth.
Unsatisfactory/incorrect (0)	Explanation does not accurately explain why the weight of a person is different on the Moon or describe the difference between the weight and mass of an object on the Moon and on Earth.	There is a lack of gravity on Moon.

Holistic Rubric 2: Accessibility for Fifth-Grade Students
(To what extent is the explanation clear, comprehensible, and easily understood by fifth-grade students?)

Level/score	Description	Exemplar
Completely accessible (2)	Explanation is likely to be understood easily by fifth-grade students • Uses language that students at this grade level are likely to understand • Analogies or examples used to explain concepts/ideas are comprehensible (as needed) • Strong coherence/connectivity across different parts of the explanation (logical progression of ideas)	Everything has mass, things with more mass have more gravity, and things with less mass have less gravity. Since the Moon has less mass than Earth, it has less gravity. Weight is a measure of how much force gravity uses to pull you to the Earth. So if you are on the Moon, since it has less gravity, there is less of a force pulling you toward the Moon, and your weight is less. Mass is different than weight because mass is a measurement of how much "stuff" is in an object. So on the Earth, if an object has a mass of 1 kg that is saying it is made up of 1 kg of "stuff." If the same object were on the Moon, it would still be made up of the same amount of "stuff," so its mass would be the same. Its weight would be less on the Moon because the smaller amount of gravity on the Moon pulls on it less.
Somewhat accessible (1)	Explanation is somewhat likely to be understood easily by fifth-grade students • Mostly uses language that students at this grade level are likely to understand (combines accessible and inaccessible language) • When used, contains analogies or examples that are confusing or irrelevant • Gaps in coherence/connectivity across different parts of the explanation.	On Earth we have gravity. Gravity makes us stay put instead of floating away. Gravity also affects our mass which in turn equates to our weight. We still have the same mass on the Moon but because there is less gravity there our weight appears less than when we are on Earth.
Inaccessible (0)	Explanation is unlikely to be understood easily by fifth-grade students • Frequently uses language that students at this grade level are unlikely to understand • Majority of response provides information about concepts or ideas that students at this grade level are unlikely to understand (e.g., mass is measured in grams; weight is measured in Newtons; mass is a numerical measure of inertia; weight is equal to mass times the acceleration of gravity) • Limited coherence/connectivity across different parts of the explanation	Weight is a measure of force pulling on a mass based on gravity. Larger objects have such as the Earth have a stronger pull from gravity than the Moon does. Although the objects masses do not change, the weight does change because weight is measured with the acceleration and pull of gravity.

Appendix F

Constructed-Response Scoring Rubric for Ecosystem Question 0026 (Plant and Food Explanation)
Holistic Rubric 1: Scientific Accuracy
(To what extent does the response address the underlying science accurately and completely?)

Level/score	Description	Exemplar
Complete (2)	Explanation includes 2 key ideas (#1: food provides energy for plant growth or fertilizer does not provide energy for plant growth; #2: fertilizer provides important nutrients/minerals to help plants grow) explaining why fertilizer does help plants grow but is not considered food for plants. Explanation includes no major misconceptions.	Sunlight, air and water are used by a plant to make its food, which is a sugar. Fertilizer contains nutrients. These nutrients contribute to the overall health of the plant, but don't give its cells energy. Nutrients help the plant utilize its energy better, but don't give it energy.
Partial (1)	Explanation includes 1 key idea (idea #1 or #2) explaining why fertilizer does help plants grow but is not considered food for plants. Explanation contains no major misconceptions.	Food provides organisms with energy and matter for growth. Did the plants stop growing without fertilizer? (no) Food is necessary for organisms to live and grow. The fertilizer is not necessary, but it is certainly helpful!
Unsatisfactory/incorrect (0)	Explanation does not include any key ideas explaining why fertilizer does help plants grow but is not considered food for plants.	Fertilizer cannot be effective in the absence of sunlight, water, and air.

Holistic Rubric 2: Accessibility for Fifth-Grade Students
(To what extent is the explanation clear, comprehensible, and easily understood by fifth-grade students?)

Level/score	Description	Exemplar
Completely accessible (2)	Explanation is likely to be understood easily by fifth-grade students • Uses language that students at this grade level are likely to understand • Analogies or examples used to explain concepts/ideas are comprehensible (as needed) • Strong coherence/connectivity across different parts of the explanation (logical progression of ideas)	Food is defined as what an organism consumes to grow and use for energy. Fertilizer cannot be used to get energy from. Kind of like taking vitamins. The vitamins helps my body complete its necessary functions more completely, but the body cannot get energy from the vitamins. I would have to consume other food for that to happen.
Somewhat accessible (1)	Explanation is somewhat likely to be understood easily by fifth-grade students • Mostly uses language that students at this grade level are likely to understand (combines accessible and inaccessible language) • When used, contains analogies or examples that are confusing or irrelevant • Gaps in coherence/connectivity across different parts of the explanation	Minerals in the fertilizer aid the plant growth and health by supplying nutrients that make the plant healthier.
Inaccessible (0)	Explanation is unlikely to be understood easily by fifth-grade students • Frequently uses language that students at this grade level are unlikely to understand (e.g., “optimize food production”) • Majority of response provides information about concepts or ideas that students at this grade level are unlikely to understand (e.g., discussion of particular molecules without any explanation; formal discussion of photosynthesis; why certain plants require particular minerals, distinction between organic/inorganic fertilizers) • Limited coherence/connectivity across different parts of the explanation	I would talk about CO_2 and how CO_2 is what makes up the plant material. I would start out with an activity about what is CO_2, where does it come from, etc. Maybe using an indicator that shows that CO_2 comes from our breath. Then I would talk about how plants use CO_2 and not fertilizer for plant matter.

Appendix G

Constructed-Response Scoring Rubric for Earth's Place Question 0029A (Solar System Poster)

Level/score	Description	Exemplar
Complete (2)	Response identifies and clearly explains three aspects of the poster that does not accurately represent the components of the solar system, how they are arranged, or how they interact; *and* these aspects relate to student misconceptions that the poster might cause or reinforce.	1. Students believe that planets are all lined up in order, as shown in the poster. 2. Students believe all planets are the same distance apart and that the distances between stars and planets are not very great. 3. Students do not understand how large the Sun is in comparison to the Earth.
Partial (1)	Response identifies, but does not clearly explain, two or three aspects of the poster that does not accurately represent the components of the solar system, how they are arranged, or how they interact; *and* these aspects relate to student misconceptions that the poster might cause or reinforce.	1. Distances between the planets and the Sun 2. Relative sizes of the planets and the Sun 3. The orbits of the planets around the Sun
Unsatisfactory/incorrect (0)	Response identifies, but does not clearly explain, one or no aspects of the poster that does not accurately represent the components of the solar system, how they are arranged, or how they interact; *and* these aspects relate to student misconceptions that the poster might cause or reinforce.	Don't know, I don't teach space/solar system and don't recall from my elementary education.

Appendix H

Constructed-Response Scoring Rubric for Earth's Place Question 0038 (Star and Planet Explanation)

Holistic Rubric 1: Scientific Accuracy

(To what extent does the response address the underlying science accurately and completely?)

Level/score	Description	Exemplar
Complete (2)	Explanation includes scientifically accurate and complete description of 2 distinct differences between a star and a planet and no major misconceptions	A star is usually very large in size and is made up of only gas. It also emits light. Meanwhile, a planet, which is usually smaller in size relative to a star, can be made up of solid or gas. Additionally, a planet cannot emit its own light, but rather it reflects light.
Partial (1)	Explanation includes scientifically accurate and complete description of 1 distinct difference between a star and a planet and no major misconceptions	A star has its own light, and a planet has no light of its own. It shines by reflecting light off of the Sun.
Unsatisfactory/incorrect (0)	Explanation does not include scientifically accurate and complete description of 1 or 2 distinct differences between a star and a planet	Stars are formed when a cloud of gas collapses under the force of gravity.

Holistic Rubric 2: Accessibility for Fifth-Grade Students
(To what extent is the explanation clear, comprehensible, and easily understood by fifth-grade students?)

Level/score	Description	Exemplar
Completely accessible (2)	Explanation is likely to be understood easily by fifth-grade students • Uses language that students at this grade level are likely to understand • Analogies or examples used to explain concepts/ideas are comprehensible (as needed) • Strong coherence/connectivity across different parts of the explanation (logical progression of ideas—e.g., comparison structure *plus* significant details to explain key distinctions)	Do you remember learning that the Moon reflects the Sun's light? Well, the planets in our solar system do the same thing. They reflect the light of the Sun. But they are so far away that they appear as tiny specks of light, like a star. You know that stars, like our Sun are huge, much larger than planets. They're also so hot they glow, so unlike planets, they make their own light. The reason they look similar to planets is that they are so far away. If you looked at your thumb close up, you could use it to "cover" your friend's whole body when he stood across the room from you. Your friend is much bigger than your thumb, but your thumb is closer to you, so it appears larger. So summing up, planets are smaller and shine with reflected light. Stars are much larger and create their own light. Both appear the same size because of how close or far they are from us.
Somewhat accessible (1)	Explanation is somewhat likely to be understood easily by fifth-grade students • Mostly uses language that students at this grade level are likely to understand (combines accessible and inaccessible language) • When used, contains analogies or examples that are confusing or irrelevant • Gaps in coherence/connectivity across different parts of the explanation	Two distinct differences between a star and planet is that star have a much larger size and gravitational pull which is caused by their size.
Inaccessible (0)	Explanation is unlikely to be understood easily by fifth-grade students • Frequently uses language that students at this grade level are unlikely to understand • Majority of response provides information about concepts or ideas that students at this grade level are unlikely to understand (e.g., how stars and planets are formed; why stars emit light; nuclear fusion) • Limited coherence/connectivity across different parts of the explanation	A star is a ball of gas that is constantly exploding, and although some planets are composed of gas, they are not combusting. Planets revolve around a star.

Appendix I

Item Statistics

I.1: Matter Item Statistics

Item ID	Item name	Item type	Percent correct	Biserial or polyserial correlation[a]
MAT_0001a	Exponents and cylinder volumes	Multiple choice	35	0.31
MAT_0001b	Exponents and cylinder volumes	Multiple choice	30	0.31
MAT_0003	Confusing weight and volume	Multiple choice	42	0.26
MAT_0004_Row1	Discussing the volume of irregular-shaped solids	Table	69	0.28
MAT_0004_Row2	Discussing the volume of irregular-shaped solids	Table	38	0.18
MAT_0004_Row3	Discussing the volume of irregular-shaped solids	Table	38	0.41
MAT_0004_Row4	Discussing the volume of irregular-shaped solids	Table	43	0.26
MAT_0005_Row1	Relationship among weight, volume, and density	Table	71	−0.06[b]
MAT_0005_Row2	Relationship among weight, volume, and density	Table	70	0.20
MAT_0005_Row3	Relationship among weight, volume, and density	Table	73	0.17
MAT_0006	Density dot diagrams	Multiple choice	27	0.15
MAT_0007	Density containers	Multiple choice	63	0.37
MAT_0009_Accessibility	Weight and mass explanation	Constructed response	n/a (0.68 average score)	0.46
MAT_0009_Accuracy	Weight and mass explanation	Constructed response	n/a (0.65 average score)	0.59
MAT_0010	Incorrect use of a ruler	Multiple choice	56	0.23
MAT_0011a_Row1	Reading a graduated cylinder	Table	91	0.16
MAT_0011a_Row2	Reading a graduated cylinder	Table	92	0.33
MAT_0011a_Row3	Reading a graduated cylinder	Table	82	0.06
MAT_0013	Density lab	Multiple choice	Removed[c]	Removed[c]

[a]Biserial correlations were calculated between each non-constructed-response item and the overall scale score for that item set. Polyserial correlations were calculated between each constructed-response item and the overall scale score for that item set. [b]These item statistics were reported when all of the items were used in the scoring model. However, these items were removed from the final scoring model due to negative biserial or polyserial correlations. [c]This item was removed from the final scoring model for the matter item set due to inaccurate programming of answer options.

I.2: Ecosystems Item Statistics

Item ID	Item name	Item type	Percent correct	Biserial or polyserial correlation[a]
ECO_0014	Carnivores rule	Multiple choice	96	0.05
ECO_0015	Plant growth materials	Multiple choice	48	0.42
ECO_0016a_Row1	Decomposers	Table	64	0.32
ECO_0016a_Row2	Decomposers	Table	92	0.51
ECO_0016a_Row3	Decomposers	Table	89	0.54
ECO_0016a_Row4	Decomposers	Table	86	0.63
ECO_0017	Cycling of matter, interpreting food webs	Multiple choice	94	0.55
ECO_0019a	New species	Multiple choice	72	0.28
ECO_0021	Plant growth models	Multiple choice	65	0.13
ECO_0022_Row1	Food comparisons	Table	54	0.22
ECO_0022_Row2	Food comparisons	Table	94	0.21
ECO_0022_Row3	Food comparisons	Table	83	0.46
ECO_0022_Row4	Food comparisons	Table	83	0.43
ECO_0022_Row5	Food comparisons	Table	71	0.32
ECO_0022_Row6	Food comparisons	Table	89	0.38
ECO_0023b	Constructing food webs	Multiple choice	48	0.34
ECO_0026 _Accessibility	Plant food explanation	Constructed response	n/a (0.51 average score)	0.25
ECO_0026 _Accuracy	Plant food explanation	Constructed response	n/a (0.35 average score)	0.40
ECO_0028a	Food web strengths and limitations	Multiple choice	52	−0.003[b]
ECO_0028b	Food web strengths and limitations	Multiple choice	70	0.29

[a]Biserial correlations were calculated between each non-constructed-response item and the overall scale score for that item set. Polyserial correlations were calculated between each constructed-response item and the overall scale score for that item set. [b]These item statistics were reported when all of the items were used in the scoring model. However, these items were removed from the final scoring model due to negative biserial or polyserial correlations.

I.3: Earth's Place Item Statistics

Item ID	Item name	Item type	Percent correct	Biserial or polyserial correlation[a]
EAR_0029a	Solar system poster	Constructed response (2 points)	n/a (1.46 average score)	0.29
EAR_0030	Moon motion	Multiple choice	87	0.25
EAR_0031_Row1	Moon positions	Select all that apply	66	0.39
EAR_0031_Row2	Moon positions	Select all that apply	96	0.09
EAR_0031_Row3	Moon positions	Select all that apply	70	0.38
EAR_0031_Row4	Moon positions	Select all that apply	29	0.15
EAR_0032a	Sun motion	Multiple choice	43	0.13
EAR_0034_Row1	Arguments regarding the moon changes	Table	53	0.13
EAR_0034_Row2	Arguments regarding the moon changes	Table	75	0.38
EAR_0034_Row3	Arguments regarding the moon changes	Table	83	0.33
EAR_0034_Row4	Arguments regarding the moon changes	Table	87	0.49
EAR_0035_Row1	Constellations	Table	51	0.36
EAR_0035_Row2	Constellations	Table	59	0.19
EAR_0035_Row3	Constellations	Table	62	0.33
EAR_0035_Row4	Constellations	Table	93	0.36
EAR_0035_Row5	Constellations	Table	84	0.01
EAR_0036a	Sun shadow	Multiple choice	83	0.51
EAR_0036b_Row1	Sun shadow	Table	56	0.29
EAR_0036b_Row2	Sun shadow	Table	68	0.48
EAR_0036b_Row3	Sun shadow	Table	74	0.46
EAR_0036b_Row4	Sun shadow	Table	55	0.29
EAR_0038_Accessibility	Star and planet explanation	Constructed response	n/a (0.50 average score)	0.33
EAR_0038_ Accuracy	Star and planet explanation	Constructed response	n/a (0.56 average score)	0.39
EAR_0039_Row1	New and full moon feedback	Matching	58	0.58
EAR_0039_Row2	New and full moon feedback	Matching	59	0.64
EAR_0039_Row3	New and full moon feedback	Matching	92	0.61
EAR_0039_Row4	New and full moon feedback	Matching	65	0.58
EAR_0040_Row1	First quarter moon feedback	Matching	80	0.62
EAR_0040_Row2	First quarter moon feedback	Matching	63	0.60
EAR_0040_Row3	First quarter moon feedback	Matching	64	0.58
EAR_0040_Row4	First quarter moon feedback	Matching	91	0.66

[a]Biserial correlations were calculated between each non-constructed-response item and the overall scale score for that item set. Polyserial correlations were calculated between each constructed-response item and the overall scale score for that item set.

Suggested citation:

Mikeska, J. N., Phelps, G., & Croft, A. J. (2017). *Practice-based measures of elementary science teachers' content knowledge for teaching: Initial item development and validity evidence* (Research Report No. RR-17-43). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12168

Action Editor: Heather Buzick

Reviewers: Gabrielle Cayton-Hodges and Hui Jin