*Research Article*

# A Framework for Analyzing Cognitive Demand and Content-Practices Integration: Task Analysis Guide in Science

Miray Tekkumru-Kisa,[1] Mary Kay Stein,[2] and Christian Schunn[3]

[1]*University of Pittsburgh, 807 Learning Research and Development Center, 3939 O'Hara Street, Pittsburgh, Pennsylvania 15260*
[2]*University of Pittsburgh, 828 Learning Research and Development Center, 3939 O'Hara Street, Pittsburgh, Pennsylvania 15260*
[3]*University of Pittsburgh, 821 Learning Research and Development Center, 3939 O'Hara Street, Pittsburgh, Pennsylvania 15260*

*Received 28 October 2013; Accepted 13 December 2014*

Abstract: Many countries, including the United States, emphasize the importance of developing students' scientific habits of mind and their capacity to think deeply about scientific ideas in an integrated fashion. Recent science education policies in the United States portray a related vision of science teaching and learning that is meant to guide the improvement efforts of science teachers, professional developers, and school leaders, as well as the design efforts of curriculum and assessment developers. To understand the extent to which this vision is being enacted in science classrooms, we consider the tasks to which students are exposed as representative of the types of opportunities that they have to think, reason, and engage in disciplinary ideas and practices in science classrooms. The purpose of this article is to advance a framework to analyze science tasks and instruction in terms of two dimensions that are critical for science learning: (1) cognitive demand; and (2) the integration of scientific content and practices. We present the Task Analysis Guide in Science (TAGS) framework through a detailed description of its categories along with concrete examples of science tasks in each category. We compare it to other frameworks related to cognitive demand. We conclude by discussing various ways in which the TAGS can serve as a helpful analytical tool for educational researchers and practitioners. © 2015 Wiley Periodicals, Inc. J Res Sci Teach 52: 659–685, 2015
**Keywords:** cognitive demand; instructional tasks; curricular tasks; task-based framework; scientific practices; integrated instruction

Change is on our doorstep. Research on science instruction and science learning has uncovered fundamental problems with how science is typically taught as well as suggested revisions in the scope and nature of K-12 science education. This revised vision of science teaching and learning is captured in new standards for science, such as the Next Generation Science Standards (NGSS) (NGSS Lead States, 2013) in the United States. Instead of learning

---

All opinions and conclusions in this article are those of the authors and do not necessarily reflect the views of the funding agency.
Contract grant sponsor: National Science Foundation to the Learning Research and Development Center of the University of Pittsburgh; Contract grant number: DRL-1027629.
*Correspondence to*: Miray Tekkumru-Kisa; E-mail: mit34@pitt.edu
DOI 10.1002/tea.21208
Published online 20 March 2015 in Wiley Online Library (wileyonlinelibrary.com).

science by "reading about science" in books or by memorizing the steps of the scientific method, students are now expected to learn science by "doing science."

In many countries, standards for educational practice now emphasize the importance of developing students' ability to use inquiry and engineering design to nurture students' scientific habits of mind (Achieve, 2010; Martin, Mullis, Gonzales, & Chrostowski, 2004). Similarly, new standards like the NGSS emphasize engaging students in scientific practices and the integration of core science ideas with scientific practices while teaching science. This integration is explicit in the ways in which the standard statements are written (e.g., "Plan an investigation to provide evidence *[the practice]* that the change in an object's motion depends on the sum of the forces on the object and the mass of the object *[the content]*"). Although such integration mirrors how science is practiced in the real world, it is far from the norm in many classrooms in which content and practices are often covered separately or the scientific practices are not covered at all (NGSS Lead States, 2013). Lessons often portray science as a static body of knowledge and focus on vocabulary that should be memorized. In such classrooms, the teacher is viewed as possessing scientific knowledge and transmitting it to students (Weiss, Pasley, Smith, Banilower, & Heck, 2003). Even when some integration of practices and content is evident in hands-on inquiry activities, these activities often do not reflect the core attributes of authentic scientific practices that scientists actually engage with when they do science (Chinn & Malhotra, 2002). Similarly, many laboratory guides continue to provide "cook-book" lists for students to follow prescriptively (Hofstein & Lunetta, 2004).

Getting students to *think deeply* about scientific concepts and ideas is another important focus across many countries (Roth & Givvin, 2008). Across a variety of studies, inquiry-based instruction *that emphasizes active thinking* (i.e., using logic, thinking creatively, building on prior knowledge, and/or making deductions) has been shown to be associated with improved student content learning (Minner, Levy, & Century, 2010). The underlying rationale for focusing on a limited set of core ideas is to allow for "deep exploration of important concepts, as well as time for students to develop meaningful understanding" (NRC, 2012, p. 25). Therefore, students in classrooms that are aligned with the new science standards will be expected to engage in cognitively complex thinking and reasoning about the scientific ideas. This means that they should frequently engage in tasks that demand interpretation, flexibility, and the construction of meaning; they must learn how to persevere when the "right" answer or preferred method is not obvious. We know, however, that this is not the case in many classrooms: Students are asked to simply verify knowledge, follow procedures, or carry out activities with weak or no conceptual links (Roth et al., 2006); teachers generally use low level "fill in the blank" questions with a focus on getting the right answer rather than helping students to make sense of the content (Hofstein & Lunetta, 2004; Weiss & Pasley, 2004); classroom practices rely on rote memorization rather than understanding, logical reasoning, and the processes of science (Gallagher & Tobin, 1987). Thus, along with the need to integrate practices and content, the academic rigor embedded in the new science standards will require a transformative change on the part of classroom teachers, not a simple additive fix (Spillane & Zeuli, 1999).

The NGSS portray a particular vision of science teaching and learning that is meant to guide the improvement efforts of science teachers, professional developers, and school leaders, as well as the design efforts of curriculum and assessment developers. As with many policy documents, the NGSS are written to inspire allegiance—among the widest possible group of stakeholders—to a coherent vision of what U.S. students should learn and how they should learn it; they are <u>not</u> a step-by-step map for the implementation of that vision. As new textbooks are written, new curricula are developed, and new assessments are designed, the specifics of what it means to teach and learn in a manner aligned with the NGSS will take shape. In fact, as underscored by the NRC

(2012), "standards will not lead to improvements in K-12 science education unless the other components of the system—curriculum, instruction, professional development, and assessment—change so that they are aligned with the NRC framework's vision" (p. 20).

The purpose of this article is to advance a framework for monitoring the quality of science tasks and instructional practices in terms of their demand on students' thinking about science content and science practices. Given the focus on inquiry and academic rigor in science education in other nations, the framework could also be a useful tool for other countries as they seek to monitor the quality of materials and environments designed to support science teaching and learning. Note that this logic applies regardless of whether the NGSS *per se* are adopted, or whether a related set of standards are adopted that are also inspired by the NRC framework (i.e., emphasizes academic rigor and instruction based on integration of science content and science practices). We also note that the framework is relevant to tertiary science instruction. We use NGSS as a salient example to stand for this new approach to standards calling for integrated content and practices instruction. Monitoring the alignment of these and other activities is important because, as noted above, the emphasis on deep student thinking and integration of content and practices represents a significant shift from the modal form of current instructional practice and, thus, represent a steep learning curve for education professionals and researchers alike.

In that regard, the TAGS can support the work of a variety of stakeholders to promote the uptake of the NGSS vision in more science classrooms in deeper and more meaningful ways. For example, the TAGS can be used by professional developers to support teachers' learning to distinguish between science tasks that may be superficially similar but actually provide very different opportunities for student learning. This kind of support is needed and important because, as underscored by Krajcik (2014), instructional materials aligned with NGSS should (1) engage students in authentic experiences reflecting the practice of science and (2) develop *deeper understanding* of the practices and disciplinary core ideas. Therefore, teachers should know how to select cognitively demanding instructional tasks aligned with the NGSS vision to be able to provide more rigorous opportunities for students' learning in science classrooms. The TAGS could serve as a guide for teachers in this regard. We also foresee the TAGS as supporting the work of other stakeholders, including teachers, curriculum developers, assessment developers, and researchers; we elaborate on these uses in the discussion section.

From past reform efforts similar to those envisioned in the NGSS, we know that it is easy to appropriate the surface features of recommended changes but completely miss the deeper more meaningful dimensions of that change (e.g., the case of Mrs. Oublier; Cohen, 1990). Two dimensions of deep change that we argue cannot be left to chance are as follows: (1) the integration of content and practices; and (2) the cognitive demand of science tasks in which students are asked to engage. While others (e.g., Anderson & Krathwohl, 2001; Porter & Smithson, 2001; Webb, 2007) have developed frameworks for the level of thinking demanded of students, our framework provides for a different conception of cognitive demand and combines cognitive demand with the recent emphasis on the integration of practices and content. Moreover, we argue that curricular and instructional *tasks* are an ideal location to examine these dimensions.

## The Affordances of a Task-Based Framework

Tasks constitute a recognizable and consequential unit of activity across a variety of "channels of influence" (NRC, 2002) including assessments, curricular materials, teacher education programs, and instructional practice. Whether and how the NGSS shape student learning will depend on the nature of tasks across all of these channels. Following research in mathematics education (Stein, Grover, & Henningsen, 1996), we focus here on *instructional* tasks. In classrooms, such tasks are considered as "the basic instructional unit" (Blumenfeld et al., 1991). They are important because working on tasks constitutes what students *do* during the majority of their time in the classroom (Boston & Smith, 2009). Students in the seven countries analyzed in the TIMSS Video Study, including the United States (NCES, 2003), spent over 80% of their time working on tasks.

We define instructional tasks as classroom-based activities, the purpose of which is to focus students' attention on a particular scientific idea and/or practices. These activities incorporate the products that students are expected to produce, the processes that they are expected to use to generate those products, and the resources available to them while they are generating the products (Doyle, 1983). A close examination of instructional tasks provides a window into the kinds of disciplinary ideas to which students are exposed and the types of opportunities they have to engage with these ideas (Boston & Wolf, 2004; Doyle, 1983; Hiebert & Wearne, 1993; Stein et al., 1996).

Therefore, the tasks with which students become engaged provide signals not only to *what* is important to learn, but also *how* one should learn it and to what "counts" as important intellectual activity. For example, science can be thought as both a pursuit and the body of knowledge that results from that pursuit. However, rather than focusing on the pursuit of new knowledge, science education typically focuses—nearly exclusively—on the results, that is, the canon of accepted knowledge (Levin, Hammer, Elby, & Coffey, 2013). Students in classrooms in which science is read from textbooks come to view science as one particular kind of activity (as the results of the pursuit); students in classrooms in which questions are developed and explored through the use of scientific practices come to view science very differently (a pursuit and its results). We argue that instructional tasks are important because they reveal the extent to which the work students do in the classroom supports students' recognition of science as a pursuit, as well as its results. As underscored in the recent Framework for K-12 Science Education, "Any education that focuses predominantly on the detailed products of science labor—the facts of science—without developing an understanding of how those facts were established or that ignores the many important applications of science in the world misinterprets science" (NRC, 2012, p. 43).

A key finding in task-based research is that tasks often change their character once unleashed in real classroom settings. In mathematics education, tasks have been shown to progress through three phases: first, as curricular or instructional materials, second as they are set up by the teacher in the classroom, and, finally, as implemented by students during the lesson (Arbaugh & Brown, 2005; Boston & Smith, 2009; Henningsen & Stein, 1997; Stein et al., 1996; Stein & Kaufman, 2010). This same research demonstrates that teachers often lower the cognitive demands of a task by breaking it into smaller subtasks (Smith, 2000), focusing on the accuracy of procedures and answers rather than thinking and reasoning processes (Doyle, 1983; Henningsen & Stein, 1997), not providing the appropriate scaffolding, and/or "taking over" the thinking and actually doing the work for the students (Henningsen & Stein, 1997; Stein et al., 1996). The ubiquity of tasks, coupled with the power of task implementation to shape student thinking, make teachers' capacity to *maintain* the demand of high-level tasks critical. Another reason that tasks are important is that the level of cognitive demand of enacted instructional tasks is associated with student gains on measures that target high-level thinking and reasoning (as do the new NGSS). Consistent results over the past 25 years have shown that students learn best when they are in classrooms in which a high-level of cognitive demand is maintained throughout lessons (Boaler & Staples, 2008; Hiebert & Wearne, 1993; Stein & Lane, 1996; Stigler & Hiebert, 2004; Tarr et al., 2008).

In mathematics education, the classification of instructional tasks into various levels of cognitive demand using a Task Analysis Guide (TAG) (Stein & Kim, 2009; Stein, Smith, Henningsen, & Silver, 2000) coupled with the tracking of tasks from the pages of textbooks to their actual enactment in classrooms (the Mathematics Task Framework [MTF]) has proven to be useful in both research and practice settings. Not only has it been used in scores of empirical studies (including large-scale, federally funded studies, e.g., the MIST and Scaling Up Mathematics projects), but it has also resonated deeply with teachers and teacher educators. It is used in a broad set of teacher education programs across the United States, as well as in professional development in the United States and abroad. The key resource (Stein, Smith, Henningsen, & Silver, 2009) has gone into a second printing and has been translated into different languages, including Chinese. For researchers, these tools have helped to keep them alert to the fact that one cannot assess the quality of instruction by only attending to the curriculum or even to the set up of a task at the beginning of a lesson. In addition, the TAG provides a unified way to analyze instruction across mathematical topics. Moreover, although the TAG and MTF were not originally intended for practitioners, teachers and teacher educators resonated with this framework immediately, claiming that it provided a language for them to talk about things that happen in the classroom but for which they had no label. Teachers began to ask for feedback on their instruction using the MTF. Researchers answered this demand through practitioner articles (Smith & Stein, 1998; Stein & Smith, 1998) and a casebook for teacher professional development (Stein et al., 2009) that illustrated patterns of task maintenance and decline.

## Advancing a Task-Based Framework in Science

Similar to mathematical instructional tasks, not all science tasks require the same level of cognitive effort from the students. Are the students asked to memorize definitions of the key scientific terminology? Are they asked to apply a set of procedures about scientific practices in a routine manner without understanding the underlying science concepts or ideas? Or are they invited to engage in deep reasoning about disciplinary ideas? We have adapted the TAG to develop a similar framework in science education called the Task Analysis Guide in Science (TAGS). The primary purpose of the TAGS is to identify the *level and kind* of reasoning required of students in order to successfully engage with a task that focuses on science content and/or scientific practices.

As shown in Figure 1, science tasks can be grouped into nine different categories by classifying them according to two critical dimensions: (1) the integration of science content and practices; and (2) cognitive demand. The first dimension, which we will refer to as simply *integration,* is addressed by each of vertical columns. Tasks are categorized as focusing on

| Cognitive Demand Levels | | Scientific Practices (e.g., argumentation and investigation) | Science Content (i.e., scientific body of knowledge) | Integration of Content and Practices |
|---|---|---|---|---|
| 5 | DOING SCIENCE TASKS | | | **Doing Science (DS)** Engaging in practices to make sense of content and recognize how scientific body of knowledge is developed |
| 4 | TASKS INVOLVING GUIDANCE FOR UNDERSTANDING | | | **Guided Integration (GI)** Guidance for working with practices tied to a particular content |
| 3 | | **Guided Practices (GP)** Being guided for understanding practices | **Guided Content (GC)** Being guided for understanding particular content | |
| 2 | TASKS INVOLVING SCRIPTS | **Scripted Practices (SP)** Following a script to work on practices | **Scripted Content (SC)** Following a script about a content | **Scripted Integration (SI)** Following a script to work on practices tied to content |
| 1 | MEMORIZATION TASKS | **Memorized Practices (MP)** Reproducing definitions/ explanations of practices | **Memorized Content (MC)** Reproducing definitions, formulas, or principles about particular content | |

*Figure 1.* Task Analysis Guide in Science.

scientific content (solely), scientific practices (solely), or the integration of the two. The second dimension, *cognitive demand*, is presented in the horizontal layers of the framework. Low-level tasks are presented in the bottom two cognitive demand levels. Level-1 involves memorization tasks and level-2 involves scripted tasks. Tasks in both of these levels provide minimal opportunities for students to engage in thinking and reasoning about science content and/or practices. High-level tasks are presented in the top three rows of Figure 1. Levels 3 and 4 contain tasks that involve guidance for understanding, and level-5 contains ‘‘doing science’’ tasks. Tasks in these three levels provide more substantive opportunities for thinking. Categorizing science tasks using the TAGS requires considering both the integration and the cognitive demand dimensions. For example, a task that focuses solely on content and at a cognitive demand level-3 is categorized as Guided Content.

The TAGS is not meant to imply that low-level tasks are never appropriate for science classrooms. We would expect productive instances of low-level tasks, alongside high-level ones, in curricular materials and during instruction. For example, in spite of their potential for facilitating deep understanding, sometimes Doing Science tasks can be difficult to implement well. Indeed, prior research has identified challenges associated with using inquiry-based curriculum materials such as allocating enough time for in-depth exploration of ideas, classroom management for maintaining productive independent work, and being able to provide appropriate amounts of scaffolding (e.g., Blumenfeld et al., 1991; Marx, Blumenfeld, Krajcik, & Soloway, 1997). However, it is important to understand what kind of learning is facilitated through low-level tasks. Anderson and Schunn (2000) summarized a large body of studies showing that memorization and repeated use strengthen an initially fragile hold on new skills and knowledge, and can optimize performance of those skills. However, pure memorization without understanding is less effective for long-term development and retention of knowledge and will not support developing a generative understanding of an idea or concept. Given the over-abundance of low-level tasks in many science classrooms today, we advocate for developing more opportunities for students' engagement in cognitively demanding tasks, which was shown to be effective in the development of students' learning mathematics with understanding (Boaler & Staples, 2008; Hiebert & Wearne, 1993; Stein & Lane, 1996; Stigler & Hiebert, 2004; Tarr et al., 2008). Similarly, in science, even though not framed as ‘‘cognitively demanding tasks’’, inquiry-oriented, project-based activities that encouraged students to develop a deep understanding of ideas, were found to be effective in supporting increases in overall science achievement (Schneider, Krajcik, Marx, & Soloway, 2002).

## Integration and Cognitive Demand: The Building Blocks of the TAGS

Recent reform documents (e.g., NGSS) as well as science education research (e.g., Berland & Reiser, 2011; Evagorou & Osborne, 2013) highlight the importance of scientific practices. Students, they argue, should learn disciplinary core ideas within the context of scientific practices because ‘‘learning science and engineering involves the integration of the knowledge of scientific explanations (i.e., content knowledge) and the practices needed to engage in scientific inquiry and engineering’’ (NRC, 2012, p. 11). As argued by Bell, Bricker, Tzou, Lee, and Van Horne, (2012), ‘‘in an effort to lend specificity to the broad notion of ‘‘inquiry,’’ the intent behind the practices outlined in the [NRC] Framework is for students to engage in sensible versions of the actual cognitive, social, and material work that scientists do’’ (p. 17). Scientific practices are important for the science education community for two main reasons: (1) engagement in practices is a means to develop students' conceptual understanding; and (2) practices define an important part of what the discipline of science entails (Fortus & Krajcik, 2012). Thus, framing science classrooms as a scientific community similar to communities in professional science (NRC, 2007) encourages students' experiencing *science-as-practice,* that is, doing of science for building rich and elaborated content knowledge (Lehrer & Schauble, 2006).

Because the tasks used in science classrooms vary, the TAGS becomes a useful tool for learning to recognize (and label for discussion across practitioners) **whether or not content and practice are integrated.** Some tasks focus *exclusively* on specific disciplinary core concepts or ideas such as forces and motion, chemical reactions or natural selection. Some tasks are *exclusively* about particular scientific practices that scientists engage as they investigate and build explanations about the world. These tasks can range from memorizing a definition of particular practices to developing skills and knowledge for effectively using them. Finally, tasks that focus on *both* scientific practices and science content are classified as integrated. Note that the tasks that involve the integration of content and practices do not have to give equal weight to content and practices. Some tasks foreground scientific practices more than science content or *vice versa*. Most importantly, integration tasks expose students to *both* scientific practices and disciplinary core ideas.

The Framework for K-12 Science Education, which guided the development of the NGSS, identified eight scientific practices as essential for all students to learn: (1) asking questions; (2) developing and using models; (3) planning and carrying out investigations; 4) using and interpreting data; (5) using mathematics and computational thinking; (6) constructing explanations; (7) engaging in argument from evidence; and (8) obtaining, evaluating, and communicating information (NRC, 2012). To assess inclusion of scientific practices as defined by the NGSS, we recommend referring to the NRC Framework's more detailed specification of these practices. Other standards documents may include other science-related practices or describe critical aspects of these practices in different ways; to judge whether or not a task exposes students to scientific practices relevant in a given context (i.e., state, country, instructional level), the relevant standards specification of critical scientific practices should be consulted.

It is important to underscore that there is large variation among instructional tasks that nominally involve the integration of science content with scientific practices. Prior research has revealed that instructional tasks differ in terms of how similar they are to *authentic scientific inquiry.* For example, many laboratory tasks require students to follow an explicit set of procedures to arrive at expected conclusions, but do not require students to think about the larger purposes of their investigations or use higher-level cognitive skills (Hodson, 1996; Hofstein & Lunetta, 2004; Roth, 1995). Germann, Haskins, and Auls' (1996) study of laboratory tasks in high school biology textbooks in terms of the extent to which they support genuine inquiry showed that they rarely required students to pose questions, investigate natural phenomena, or construct answers or generalizations. These findings are consistent with prior analyses of science laboratory manuals, which showed that students are often provided step-by-step instructions to follow. As a result, students do not have opportunities to "pose a question to be investigated, formulate a hypothesis to be tested, or predict experimental results; to design observations, measurements, and experimental procedures; to work according to their own design; or to formulate new questions..." (Germann et al., 1996, p.482). These studies are consistent with Chin and Malhotra's (2002) analysis of the inquiry activities in nine upper elementary and middle school science textbooks, which revealed that textbook activities consistently failed to incorporate elements of authentic scientific inquiry. For example, none of the textbook activities required students to generate their own research questions. In only 2% of the activities were students given opportunities to select their own variables to investigate.

Students were not often asked to consider control variables. Clearly, the way students were asked to engage in inquiry in these tasks was different than authentic scientific inquiry. With its distinctions between Doing Science, Guided Integration, and Scripted Integration tasks, the TAGS framework captures variation in inquiry tasks that only nominally involve the integration of science content with scientific practices.

The second building block of the framework is the **cognitive demand** or the kind and level of thinking demanded of students to successfully engage with a task (Stein et al., 2000). *Doing science* tasks (level-5) are the most open or unstructured; they require considerable cognitive effort because there is not a well-rehearsed approach or pathway to approaching the task. Rather, students develop a deep understanding of a natural phenomenon by accessing relevant knowledge and engaging in scientific practices. These kinds of tasks require students to self-regulate their own cognitive process in order to monitor, and, if necessary, adjust their approach. *Tasks involving guidance for understanding* (level-3 and level-4) are also considered high-level because they too require cognitive effort, but, in contrast to ''doing science'' tasks, these tasks often have suggested pathways. Thus, there is less ambiguity for students in terms of what to do; nevertheless, the suggested pathways cannot be followed mindlessly. Rather, following them requires students to understand what they are doing and why.

The two lower-level kinds of tasks focus on the correct answer rather than meaning making. In *tasks involving scripts* (level-2), there is no ambiguity in terms of what students need to do because the tasks explicitly tell students what to do. Students can complete the task without needing to think about scientific ideas or principles because simply following the script takes students directly to the correct solution. *Memorization tasks* (level-1; the lowest level) require exact reproduction of previously seen materials (i.e., definitions, rules, formulas, and principles) and what is to be produced is clearly and directly stated.

Note that the level of demand of a given task is not determined purely by the contents of the task in isolation, but rather the relationship of the task to prior experiences. For example, if students have engaged with a particular task before on multiple occasions, a seemingly demanding task can actually become a memorization task. Note that the critical dimension is amount and kind of prior experience not grade level or age *per se*. Thus, a given task may often be less cognitively demanding for 12th graders than 5th graders, but not simply because the students are older, but instead because they have likely amassed relevant experiences in and out-of-school or have learned strategies for dealing with that particular content. However, in-depth prior experiences can trump age-linked patterns on cognitive demand. For example, young children with a large amount of experience with chess find complex chess configurations as having low demand, whereas adults with little chess experiences find these same configurations as having high demand (Chi, 1978).

Overall, the TAGS involves considering both the integration and the cognitive demand dimensions to reason deeply about the type of opportunities that we provide for students' learning in science classrooms. Details about each of the TAGS categories is provided below. In addition, decision trees are provided as supplementary material to assist the classification of tasks and instruction into the TAGS categories.

### *Memorization Tasks: Repeating Facts*

As suggested by the framework, there are two types of memorization tasks. The first type focuses exclusively on content (***Memorized Content***). Memorized Content tasks require students to reproduce previously learned material, often referred to by researchers as the ''scientific body of knowledge.'' This includes rules, formulas, and definitions of scientific terminology. For example, a Memorized Content task could be a crossword puzzle that uses parts of a plant cell

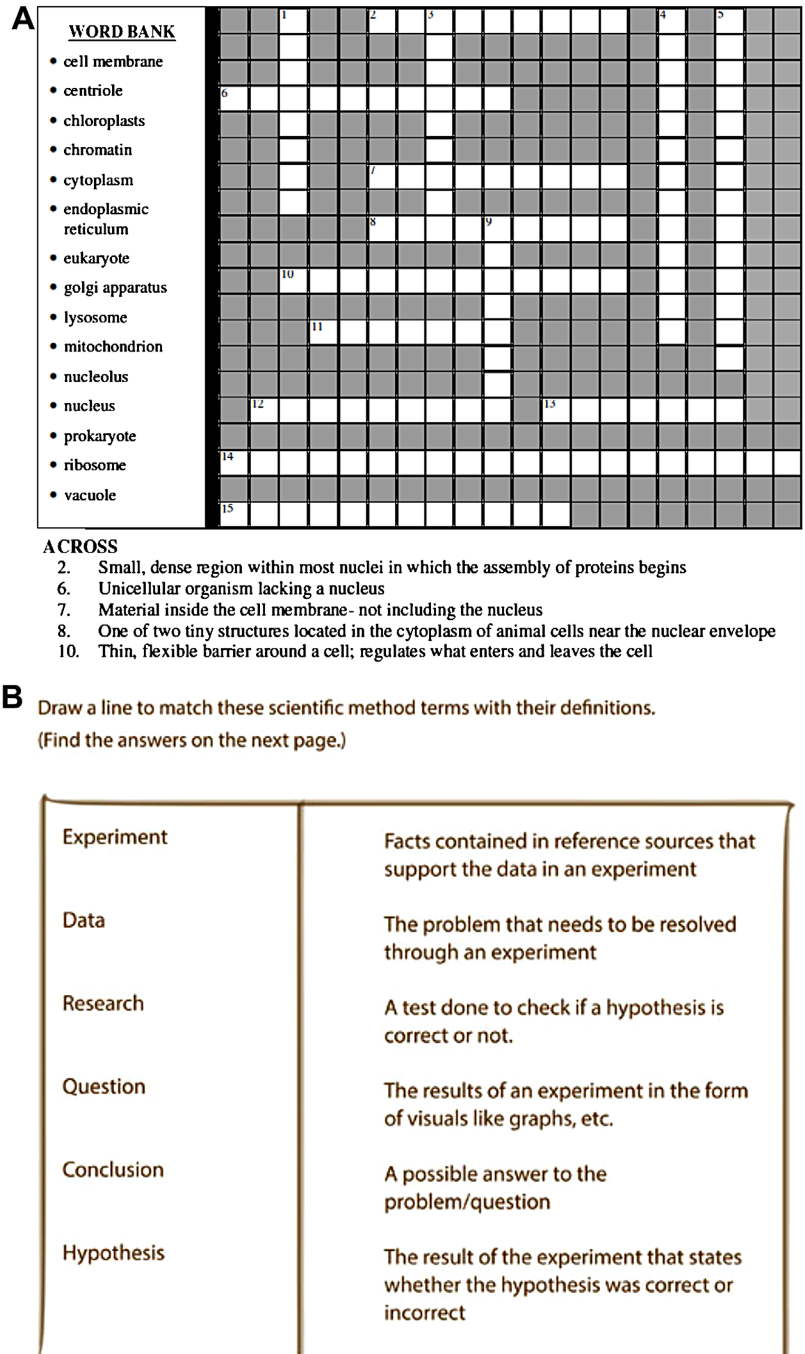


*Figure 2.* (**A.**) An example Memorized Content task. (**B.**) An example Memorized Practices task. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

(''Cell Organelle,'' n.d.) (Figure 2A). Successful completion of the puzzle depends upon having memorized definitions of the terms that appear on the left side in order to match them with the appropriate definitions (e.g., ''unicellular organism lacking a nucleus'' matches with ''prokaryote'').

The second type of memorization tasks focuses exclusively on particular scientific practices (***Memorized Practices***). These tasks require students to reproduce descriptions of scientific practices (or the terminology associated with certain scientific practices) to which they have already been exposed. For example, the task in Figure 2B (''Match the Terms,'' n.d.) requires students to match the scientific method terms with their definitions (e.g., ''the problem that needs to be resolved through an experiment'' matches with ''question'').

Memorized Content and Memorized Practices tasks do not require understanding because students can remember the previously provided definitions; they do not need to make sense of what these definitions mean nor how the ideas or practices could be applied in different settings. Such tasks are not cognitively demanding because what students are supposed to produce is very clear and explicitly stated. These kinds of tasks are generally useful if the goal is retrieving basic facts and definitions (Stein et al., 2000).

Memorization tasks are a common form of teaching and learning science in many U.S. classrooms. Prior research indicates that in many science classrooms, teachers focus more on ensuring that correct vocabulary is used than to students' reasoning (Levin, 2008; Levin, Hammer, & Coffey, 2009). The focus on facts is exemplified by actions during instruction such as ''The students copied notes from the blackboard for half of the lesson, and the remainder of the lesson was spent with the teacher asking them to recall information from the notes'' (Banilower, Smith, Weiss, and Pasley, 2006, p. 98). Similarly, despite no longer being considered as an accurate characterization of science, the fixed list of steps in ''the'' scientific method is still a part of science classrooms, textbooks, and laboratory manuals (Abd-El-Khalick, 2012; Abd-El-Khalick, Waters, & Le, 2008; NRC, 2012).

*Tasks Involving Scripts: Focusing on Desired Answers Rather Than Meaning Making*

The scripted nature of these low-level tasks makes them distinctive from others. ''Script'' refers to a provided, well-defined set of actions or procedures a student needs to take, usually in a given order, to complete a given task. A student can follow those actions and reach the desired answer without really knowing how or why the script leads to that answer. Limited demands are placed on student thinking; they can follow meaningless actions to complete the task. In the ***Scripted Content*** tasks, students are required to use a procedure related to a particular scientific formula or a principle, which they have been taught.

Commonly used Scripted Content tasks require students to solve problems following a well-defined formula (see Figure 3 for an example). The task in Figure 3 asks students to find the speed of a sailboat using the distance traveled during a certain time (''Calculating Speed,'' n.d.). The procedure for solving the task involves identifying the distance and time provided in the problem, and then placing them into the given equation for calculating the speed. The task does not require any reasoning about what speed is, nor about how speed is related to distance and time. Moreover, it does not engage students in scientific practices that are associated with mathematics; students are *not* required to (1) reason from a scientific content perspective about why division is required as a mathematical operator; (2) use the equation to solve a scientific problem; or (3) apply ratios in the context of complicated measurement problems. Following the defined script is all that is needed for students to complete this task.

The kind and level of student thinking that can be categorized as Scripted Content in biology instruction is illustrated in the scenario in Figure 4 (adapted from an activity in ''Simple

<table>
<tr><td>SPEED EQUATION:<br><br>$$Speed = \frac{Distance}{Time}$$<br>Note: The SI unit for speed is meters per second (m/s).</td><td>EXAMPLE:<br><br>Problem: What is the speed of a sailboat that is traveling 100 meters in 120 seconds?<br><br>Step 1: Write down the equation you need to solve the problem.<br>$$Speed = \frac{Distance}{Time}$$<br>Step 2: Place the known information into the equation.<br>$$Speed = \frac{100m}{120s}$$<br>Step 3: Solve. Carefully enter the numbers into your calculator. Remember that this is a division problem. Check to make sure that your solution contains the correct SI unit.<br>The sailboat has a speed of .83m/s.</td></tr>
<tr><td colspan="2">PRACTICE PROBLEMS:<br><br>1. Calculate the speed of a dog running through a field if he is covering 23.7 meters in 54 seconds.<br>2. If a cross-country runner covers a distance of 347 meters in 134 seconds what is her speed?<br>3. What is the speed of a baseball that travels 49 meters in 2.4 seconds?</td></tr>
</table>

*Figure 3.* An example Scripted Content task.

Genetics," n.d.). The scenario shows the limited cognitive demand placed on students that results when the task focuses on successful completion without requiring students to make sense of the underlying scientific ideas. The example student work in Figure 4 required almost no cognitive effort from the student. By following the provided procedure for setting up a Punnett Square and for identifying the percentage of round-seed offspring, students were not required to think about what is being represented in the Punnett Square nor what it means to get a certain percentage of round-seed offspring. The task could be completed successfully by simply following the scripted steps associated with a Punnett Square. Most importantly, the steps could be completed without

> The teacher began the lesson with a description of what a Punnett Square is. She demonstrated on the board how to set up a Punnett Square to solve simple genetics problems. Students were then asked to set up Punnett Squares for the following crosses: 1) Rr X rr, 2) Rr X Rr, and 3) RR X rr with round seeds (dominant) and wrinkled seeds (recessive) to identify what percentage of offspring will be round.
>
> Each student solved the task individually in a notebook. Then, the teacher assigned each cross to one of the students. Students drew the Punnett Square for the crosses on the board, showing how to determine the percentage of offspring that will be round seeds. The teacher decided whether students set up the Punnett Squares correctly and identified the percentages accurately. The rest of the class was asked to correct their solutions on their notebooks if they made a mistake.

*Figure 4.* An example of student thinking categorized as Scripted Content.

**Part C: Mass of Objects**

Tools: Electronic scale or balance.

1. Determine the mass (in grams) of the 3 marbles: _________

2. Determine the mass of 20 ml of water. To do this you will need to weigh an empty graduated cylinder , then add the water and find the difference.

Mass of Graduated cylinder ______
Graduated Cylinder + 20 ml of water ______
Mass of 20 ml of water _____

3. Use the same technique to determine the mass of 50 ml of water: ___________

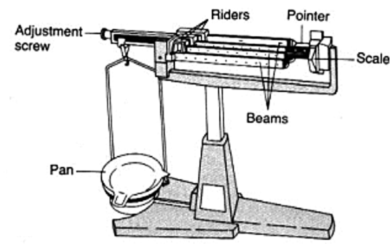


*Figure 5.* An example Scripted Practices task.

consideration of the underlying biology: the alleles and the genes that were used; how they carry the genetic information that the parents have; or how genetic information is transferred to offspring.

Similarly, ***Scripted Practices*** tasks do not require complex thinking. Students can mechanically follow scripted actions acquired through prior instruction. Different from Scripted Content, the script in Scripted Practices tasks *is related to a particular scientific practice*. For instance, in laboratories, students might be given a task that focuses on reliable data collection methods, a key component of "planning and carrying out investigations." For example, they might be shown how to reliably measure the mass of 20 ml of water (see Figure 5; adapted from "Scientific Processes," n.d.). The task shown in Figure 5 does not focus students on understanding mass as a property of matter. Rather, it simply provides a well-defined set of actions that are needed to measure the mass of a liquid reliably. By following these instructions, students can find the mass of 20 ml of water without any cognitive effort. A task like this could help students to develop the skill of measuring the mass of a liquid. However, because the task is not situated in a meaningful context that requires students to think about the reasons for considering the mass of a graduated cylinder and the potential consequences of not considering that in an experiment, students do not develop a deeper understanding of reliable data collection, which is fundamental to scientists' thinking.

The third type of tasks involving scripts is ***Scripted Integration***, also considered to be at a low level of cognitive demand. Although NGSS's focus on the integration of content with scientific practices might be taken to imply that all tasks that require students to engage in scientific practices within the context of science content are of high quality, that is not the case. Scripted Integration tasks involve both science content and scientific practices but at a very superficial level. For example, many cookbook-like, hands-on science activities could be considered to be Scripted Integration tasks because they generally require students to follow a set of actions within the context of particular science content but without requiring students to make sense of the disciplinary ideas. In fact, the TIMSS study indicated that in 27% of U.S. lessons, students carried out activities without any links to science ideas, "for example, students spent an entire lesson following a set of procedures to build rockets, without any consideration given to the science ideas related to the activity" (Roth & Givvin, 2008, p. 26). In other words, although building rockets is an activity that involves engineering design and can draw in Newton's third law of action-reaction, the task was designed such that students paid no attention to those scientific ideas, in addition to not having to think about engineering design practices conceptually because of the script being followed. Due to frequent experiences of this type, students come to perceive "the principal

> Students were introduced to osmosis in the previous lesson. In this lesson, they were told to work on a laboratory task. The teacher told them "try this activity to see how osmosis occurs in an egg cell". Students are handed out the task sheet that presents the materials and step-by-step instructions that are needed to complete the task.
>
> The instructions involve sentences such as:
>
> *"Place the egg into the 500 ml beaker and pour 250 ml of vinegar over the egg. Cover the beaker".*
>
> *"Carefully pour the vinegar into a graduated cylinder, record the amount you now have in the data table".*
>
> Students worked on the lab by following these instructions in small groups. The teacher walked around and made sure students followed the instructions. The lesson was concluded when all the students completed the task and shared with the whole class what their group observed happening to the egg.

*Figure 6.* An example of student thinking categorized as Scripted Integration.

purpose for a laboratory investigation as either following instructions or getting the right answer'' (Hofstein & Lunetta, 2004, p. 38).

Figure 6 presents a classroom scenario involving 7th grade students' work on an osmosis task (adapted from ''Observing Osmosis,'' n.d.). As revealed in this scenario, the level and type of student thinking can be categorized as Scripted Integration. In this task, students were asked to simply follow a set of actions that require almost no reasoning about osmosis. The actions happen to be related to an investigation within the context of particular science content (i.e., osmosis). But the task does not require students to make conceptual links with what they learned about osmosis, so they could complete this lesson without understanding the underlying science ideas or principles. Moreover, this form of students' engagement in scientific practices does not really involve understanding how scientific knowledge is produced through such practices. Thus, even though science content is taught to students within the work surrounding particular scientific practices (i.e., carrying out investigations), tasks at this level focus only on producing the desired answer rather than producing meaning about scientific ideas or how scientific ideas are developed. For example, in a task like this, if students obtained unexpected results, they would potentially think that they did made a simple procedural error, instead of entertaining the possibility that there might be something wrong with their hypothesis or the procedures selected in this investigation (Chinn & Malhotra, 2002; Pickering & Monts, 1982). In general, Scripted Integration tasks aim to engage students in scientific inquiry but fail to engage them in the kind of reasoning processes that are employed in real scientific inquiry (Chinn & Malhotra, 2002); they also fail to engage students in conceptual thinking about the scientific ideas covered in the task. For these reasons, Scripted Integration tasks require little cognitive effort even though they involve some integration of content and practices. As argued by Germann et al., (1996), different types of tasks are needed to transform students' experiences from such cookbook activities in which ''they work like technicians'' to those in which ''they work like scientists'' (p. 496).

*Science as a Pursuit and Its Results: Doing Science and Guided Integration Tasks*

In contrast to the scripted integration tasks (at level-2), there are two types of *cognitively demanding* tasks that involve the integration of science content and scientific practices. The first one is ***Doing Science*** tasks, which require students to work like a scientist. In these tasks, students need to use various scientific practices to be able to develop or deepen an understanding of a scientific idea as they explore a natural phenomenon. Such tasks require students to access relevant

After their exploration by modeling different plate boundaries (i.e. convergent boundaries, divergent boundaries, and transform boundaries), students had a discussion about a question that emerged during their exploration: "With all this plate motion, is the Earth staying the same size or getting bigger or smaller?" There were students who believed the Earth was staying the same size and two thought it was getting bigger. The teacher used students' uncertainty to allocate a productive time for a classroom debate about their question. The following dialogue was taken from part of the debate:

*Pint: You have to dig and dig [to find the dinosaur bones]. So that means the Earth has been getting larger because you have to dig so much to get the bones…*

*Olive: Yeah, we saw Jurassic Park, I guess.*

…

*Fern: I understand how you think of the dinosaur bones. But those are convergent that have covered the dinosaur bones. Bot not all convergence makes mountains. Some meet [gestures that plates meet and stay flat]. So the dinosaur bone was one plate and then that plate kind of moved and then that converged and overlapped.*

…

*Fern: So let's say some dirt moved over here, but then there's dirt not over there. There still might be dirt over there. So it's still even because that dirt over here came from over there. So the world is even, and it's not growing, because the magma might come then it diverges and collapses.*

*Figure 7.* An example of student thinking categorized as Doing Science.

knowledge and make appropriate use of it as they work on an (mostly) authentic problem (i.e., an open scientific question or solving an existing engineering challenge). Tasks that are developed using the principles of designed-based learning (Mehalik, Doppelt, & Schunn, 2008) or project-based learning (Krajcik & Blumenfeld, 2006) can often be categorized as Doing Science.

The classroom discussion summarized in Figure 7 illustrates the level and type of student thinking categorized as Doing Science task in a mixed grade classroom of fifth and sixth students (adapted from a classroom example provided by Reiser, Berland, and Kenyon (2012, p.36) to illustrate meaningful engagement in explanation and argumentation practices). In the classroom discussion, students work to construct an explanation of how tectonic plate movement affects the shape and size of the Earth. The argumentative nature of this discussion shows how students collectively tried to make sense of a scientific idea. As seen in the short snapshot of their discourse, the role of the teacher is minimal; students try to make an explanatory account of a natural phenomenon for a question that they themselves proposed. These are all characteristics of Doing Science tasks and the nature of students' thinking when they work like scientists.

The cognitive demand that is required of students to engage in Doing Science tasks is very high. In many science classrooms, students need more guidance to be able to engage in similar experiences. Indeed many recommendations for effective inquiry in science education mention use of guidance and scaffolds (Lewis & Lewis, 2005; Schwarz & Gwekwerere, 2007; Moog & Spencer, 2008). As shown in the TAGS, such more guided tasks are classified as ***Guided Integration***. The kind of thinking and reasoning required in Guided Integration tasks is very similar to what is required in a Doing Science task. However, in Guided Integration tasks, students are assisted to engage in high-level thinking through scaffolding offered by the teacher (or a more expert peer) or by supports contained within the written task itself. The purpose of these scaffolds and prompts is to provide just enough assistance to allow the student to get started, to make progress if stalled, or to become "unstuck," all the while maintaining many aspects of thinking

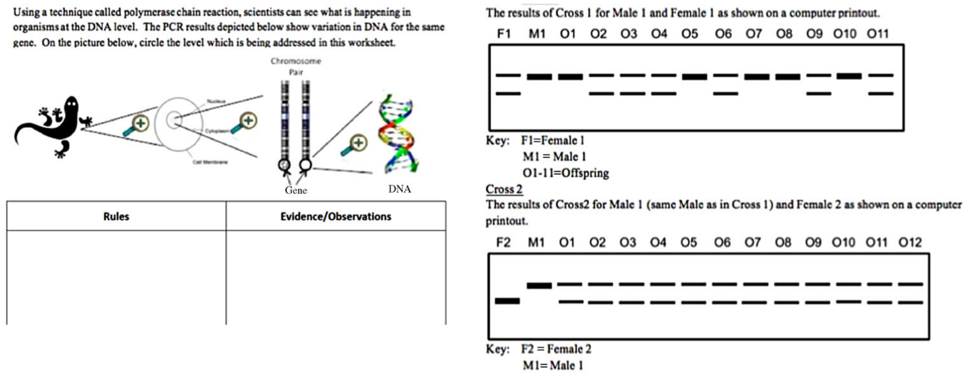


*Figure 8.* An example Guided Integration task. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

and reasoning that are characteristic of Doing Science tasks. Thus, guidance does not involve telling students what to do, nor does it involve providing students with a complete script to follow. Removing these scaffolds can often change a Guided Integration task into a Doing Science task.

The high school biology task in Figure 8 illustrates a Guided Integration task. The task starts with helping students understand what a polymerase chain reaction (PCR) is and what it depicts. Then, students are provided with the PCR results of the two crosses. Their task is to identify patterns in the PCR results and to interpret the patterns in order to derive rules of inheritance. The written task provides various scaffolds such as a picture that helps to situate students' observation at the DNA level, a table to record observations related to each rule that they will generate, and two specific crosses, which are designed to support students' interpretation of the patterns in the PCR results (e.g., according to the first cross, it appears as though some offspring look like the mother and some look like the father, but one cannot say the same thing for the second cross. Thus, it is not a valid conclusion that offspring will either carry the father's or the mother's genes for a particular trait). Even though students' thinking is guided in these ways, the thinking and reasoning about what they see is still left to the students to a considerable extent because without deeply thinking about the patterns, they cannot derive the rules of inheritance (e.g., one line comes from mother and one line comes from father; therefore, both mother and father contribute to the offspring for a particular gene).

Another Guided Integration task example in a very different context comes from a prekindergarten science classroom that used a ''research day'' activity developed by the teacher (adapted from a classroom example provided by Bell et al. (2012, p.18)). The teacher implementedthe Research Day activity as part of a unit about garden ecosystems. Students did research on the questions that they had about insects and other living creatures found in the garden. They were provided with a set of questions to think about while doing their research. Students were given time to look through the books preselected by the teacher and the school librarian, and document what they learned through drawings and dictations to their teacher. Then, they stood in front of the class to share their research papers (their drawings and what they dictated to their teacher to be written on their paper) with their peers. There were various scaffolds in this task for the students, such as providing questions to consider during the research and pre-identifying books that enabled students to focus on productive content. However, students were still provided with opportunities to productively engage in the scientific practices of obtaining, evaluating, and communicating information (NRC, 2012) by deciding which information was necessary and useful for their question, and how to communicate it to others.

Prior research brings attention to the challenges that students experience with Doing Science tasks. Students often find them ambiguous and/or risky because it is generally not clear what to do in these tasks and how to do it (Doyle, 1983,1988; Stein et al., 1996). They are more likely to get frustrated and fail to persist if they cannot do the task (Blumenfeld et al., 1991). Due to such challenges, students often urge teachers to make such tasks more explicit, often by reducing the sense making aspects of the task (i.e., provide a script). Because of their guided nature, we expect that many science teachers prefer using Guided Integration tasks to Doing Science tasks in their classroom because the scaffolding in the task can ameliorate the risk of ''losing some students'' who find the task too challenging. Because of their consistency with the vision of science learning emphasized in the recent science education policies (NGSS, NRC Framework), Guided Integration tasks can still provide similar learning opportunities to those provided by Doing Science tasks.

### *Guided Practices and Guided Content Tasks: Deepening Understanding of Content and Practices*

As suggested by the TAGS, there are also science tasks that can provide high-level thinking and reasoning opportunities for students by focusing on either science content alone or scientific practices alone. These are level-3 tasks that are classified as ***Guided Content*** and ***Guided Practices***. Such tasks assist students in a manner that helps to develop their understanding regarding what they are producing and how they are producing it. They often come in the form of ''application'' activities; students are first introduced to a topic either about a particular scientific idea, such as photosynthesis, or about a particular scientific practice, such as argumentation, and then they are asked to work on various applications of the topic or practices in different contexts or problem situations. Thus, such tasks generally are associated with deepening students' understanding of the topic; similar to Guided Integration tasks, Guided Content and Guided Practices tasks involve *scaffolding* to enable students ''to grasp a particular concept or achieve a particular level of understanding'' (Maybin, Mercer, & Stierer, 1992, p. 188).

The Guided Content task on the left in Figure 9 represents a task that requires thinking deeply about the particular science content presented in the movie. It does not require any engagement in the distinct practices of the discipline. Rather the goal is to encourage students to use what they

| Guided Content | Guided Practices |
|---|---|
| At the end of the lesson about photosynthesis, students were shown a movie about green plants and how they produce oxygen. They were asked to identify inaccuracies in the movie with respect to what was covered in the class about photosynthesis and respiration of plants. | The lesson aimed to help students understand constructing scientific explanations. Students were provided with 1) a worksheet that involved an example for "claim", "evidence" and "reasoning" about the study of regeneration in planarians, and two empty boxes to write claim, evidence, and reasoning, and 2) four short journal papers about different topics in Economist-Science & Technology.<br>The task was: *Focus on two of the papers and for each of them, identify the arguments made by the authors and write down their claim, evidence and reasoning on your worksheet.* |

*Figure 9.* Example tasks involving guidance for understanding but isolating only content or only practices.

learned about photosynthesis to identify inaccuracies in the movie about how green plants produce oxygen. The Guided Practices task in Figure 9, in contrast, illustrates a task that requires students to deepen their knowledge about the scientific practice of "constructing explanations." The goal is not to help students to learn any particular science content even though they can gain new knowledge about the topic of the papers when they read them. Instead, students are required to learn features of scientific explanations and how to identify them in scientific papers. Often the topics of these papers is purposely chosen to be off topic from the current class or to involve relatively trivial content so that the content does not distract students from the goal of learning to construct scientific explanations.

It should be noted that classifying tasks based on the TAGS depends critically on deciding what constitutes scientific content or scientific practices. All tasks inherently ask students to *do* things. Thus, every task has some kind of activity or process embedded in it, but not all of the activities can be categorized as scientific practices. To broadly interpret every action or process that students are asked to do in a task as a *scientific* practice misrepresents scientific practices. In particular, it often misrepresents the particular ways in which science implements more general practices. For example, scientists ask questions but not every question is a scientific question. Similarly, not every explanation about science content is an instance of a scientific explanation, nor is every way of explaining an example of constructing explanations in science. As argued by Reiser et al., (2012), "the scientific practice of explanation goes beyond defining or describing a named process" and instead "links a chain of reasoning to the phenomenon to be explained" (p. 8).

In general, Guided Content tasks require students to engage in high-level cognitive processes —such as finding relations, analyzing information, generalizing to a broader conceptual idea— that could be confused with scientific practices. This confusion would lead to categorizing many Guided Content tasks as Guided Integration tasks. Such confusion is problematic because if students do not engage in actual core practices of science, they will not learn or be able to independently use these practices. To prevent this confusion and likely loss of tasks involving real practices, one must think more deeply about what students are asked to do in a task. The main purpose of engaging students in the practices of science is to help students understand how scientific knowledge develops (NRC, 2012). To decide whether students are really asked to engage in scientific practices in a task, one must judge whether or not students' actions and thinking can help them to understand how scientific knowledge develops. We recommend referring to the scientific practices identified by the NRC (2012) for detailed specifications that are broadly appropriate to science found in K-12 contexts. We also think that as the NGSS aligned materials are designed and used, researchers and practitioners will develop refined understandings of the practices and their key attributes. This will improve our ability in analyzing tasks based on the integration and cognitive demand dimensions in the TAGS.

### *About the Empty Cells in the TAGS Matrix*

After a close look into the nine different TAGS categories, it is useful to reflect on why six of the five demand levels x three integration types combinations are grayed out (see Figure 1). The cells at the intersection of the cognitive demand level-5 and "scientific practices" and "science content" are greyed out because "working like a scientist" inherently constitutes engaging in scientific practices and science content at the same time. Therefore, it is not logically feasible to require students to think like a scientist but solely focus on science content or scientific practices. Integration also inherently involves higher cognitive demand when students are responsible for the integration (i.e., it is not scripted). This explains why Guided Integration generally sits at a higher cognitive demand than does Guided Content or Guided Practice (i.e., why the level 4 content only or practice only cells are gray, and why the level 3 integrated cell is also gray). However, as shown in the TAGS framework, some especially demanding Guided Practices or Guided Content tasks could be more cognitively demanding than some relatively simple Guided Integration tasks. This overlap of cognitive demand levels is shown in Figure 1. Finally, the cell at the intersection of cognitive demand level-1 and integration is grayed out because it is not feasible to require students to reproduce the definitions/explanations of scientific practices and definitions/formulas/principles about a particular content in an integrated way. A task might require students to reproduce the definitions of a list of practices and scientific content in a sequential way, but such a task will not involve any integration of practices *applied to* content.

## The TAGS in Comparison to Existing Frameworks in Science Education

One way to see distinctive characteristics of a tool is to compare and contrast it with other tools that purport to accomplish similar things. Analyses of tasks in science prior to the NGSS have previously paid attention to combinations of science topics and cognitive demand, and so it is important to ask whether those older frameworks could be applied here as well and how the TAGS is different from these exiting frameworks. The two prominent ones are the generic (revised) Bloom's taxonomy (remember, understand, apply, analyze, evaluate, create; Anderson & Krathwohl, 2001; Krathwohl, 2002) and Porter's Survey of Enacted Curricula (SEC) framework (Porter, 2002; Porter & Smithson, 2001; Porter, Smithson, Blank, & Zeidner, 2007), which is a more specific application of Bloom's taxonomy. In our work with science teachers, we find teachers often refer to Bloom-like terminology when they are asked to analyze a science task. Both Bloom and the SEC conceptually apply roughly the same variation of cognitive demand categories to topic categories, but because the SEC framework is more specified in its application to science, we focus here on the SEC.

The SEC has been successfully used to measure alignment (or frequently misalignment) between standards and assessments or between curricula and assessments (e.g., Polikoff, Porter, Smithson, 2011; Porter & Smithson, 2001) and, more importantly, to successfully provide explanations of learning outcome differences (e.g., curriculum A produced inferior learning outcomes to curriculum B on assessment X because the overlap in coverage with assessment X was higher for A than B). The SEC organizes a task, lesson, curriculum, set of standards, or assessment in a two dimensional framework: along the column are the ''Topics'' of science (larger knowledge topics divided into particular smaller concepts and processes of science, but not their integration), and along the rows is a dimension also labeled Cognitive Demand, but it uses categories from Bloom's taxonomy adapted to science (memorize, understand concepts, apply concepts, perform procedures, analyze information, conduct experiments). In the SEC, a given task is situated in this two-dimensional space as having particular content and with a particular cognitive demand (e.g., a task asking students to *memorize* information about *natural selection* or another task asking students to *apply* ideas about the *nature of science* to a novel situation).

Although largely overlapping, Bloom and the SEC are not entirely identical. One difference between typical Bloom applications and the SEC is that typical Bloom applications assume there is a hierarchy, in which some levels are inherently more indicative of a deeper learning (similar to the assumptions of the TAGS), whereas the SEC treats them as simply descriptions of different competencies that are independently learned and independently demonstrated on assessments. There are also small differences in the particular categories of demand, with SEC using ''conduct experiments'' rather than the generic ''creating'' category, and ''applying'' has conceptual and procedural elements (although over the years, different versions of the SEC have used variations of these categories).

One drawback of the SEC or prior Bloom-like analyses of science is that the integrative nature of the content and practices of science is imperfectly reflected. For example, conducting experiments in the SEC is treated as a kind of demand rather than a kind of practice. This approach ignores the special nature of integration as fundamental to learning science. Of course, this problem could be addressed by dropping practices-like categories from SEC and applying either SEC or original Bloom cognitive demand categories to an extended content dimension which, like in the TAGS, treats content, practices, and their integration as one larger dimension.

However, there is a second problem with the SEC and Bloom cognitive demand categories that cannot be addressed by such a reformulation because the problem is fundamentally with the conceptualization of cognitive demand: based on the nearly 60 years of cognitive science research since Bloom's original formulation, there has been refinement to the understanding that cognitive actions (e.g., application, evaluation) do not necessarily constitute cognitive demand. More specifically, each of the Bloom or SEC categories represents *cognitive actions* that, depending upon the nature of the situation, could be high or low cognitive demand.

To illustrate, ‘‘apply’’ can be very high or very low cognitive demand depending upon the similarity of the prior studied experiences to the test situation as well as the overall complexity of the test situation (Paas, Renkl, & Sweller, 2003; Sweller, 1988; Sweller et al., 1998); yet we find teachers are lured into thinking any application is high demand because of the Bloom formulation. For example, application does not require a lot of cognitive effort in a situation in which students are required to use the well-defined force equation to which they have already been introduced to solve a problem about the increase in the rate of the speed that is experienced while riding a roller coaster (adapted from ‘‘How Roller Coasters Work,’’ n.d.). If the problem is only ‘‘response’’ driven and does not require students to create an explanation, it could be categorized as Scripted Content based on the TAGS. By contrast, application requires considerable cognitive effort in a task in which students are asked to use what they learned about force to design a roller coaster that illustrates different forces at work in people's body. Depending on the amount of guidance students are provided in this task, it can be considered as either a Guided Integration or a Doing Science task because the task requires students to deepen their understanding of force by engaging in design work. Such design work requires students to develop a model that shows the relationship between force, mass, and acceleration as part of an explanation of a real-world phenomenon. Therefore, even though both tasks require students to ‘‘apply’’ what they learned about force in a different problem context, because of the difference in the complexity of the problem context, these tasks place very different cognitive demands on students' thinking.

Further, ‘‘understand’’ is not a cognitive action, but rather understanding is a general and fundamental result that is built through actions like application (Fortus, Dershimer, Krajcik, Marx, & Mamlok-Naaman, 2004) or evaluation (Chi, De Leeuw, Chiu, & LaVancher, 1994). The example above about designing a roller coaster can develop students' understanding of the concept of force and its relationship to mass and acceleration. Similarly, students can develop an understanding of transfer of energy through their evaluation of wind turbines at different locations in terms of their effectiveness for transferring wind energy into electricity (see Lord, 2014 for the details of such a task). Across these two examples, understanding of scientific ideas resulted through actions like application and evaluation rather than from an ‘‘understand’’ action.

## Conclusion

Our goal in this paper was to introduce a task-based framework for analyzing the type of opportunities for thinking and reasoning that students are exposed to through classroom instruction and curricular materials. Tasks were selected as our unit of analysis because they provide a window into the opportunities that students have to engage in thinking and reasoning about the disciplinary ideas and practices as well as the ways in which students engage in these ideas and practices (Blumenfeld et al., 1991; Boston & Wolf, 2004; Doyle, 1983; Hiebert & Wearne, 1993; Stein et al., 1996). The framework that we advanced (i.e., the Task Analysis Guide in Science) analyzes science tasks by drawing attention to the two critical dimensions for science learning: (1) the extent to which science content and scientific practices are integrated in the task; and (2) the cognitive demand of the task. Although integration can influence cognitive demand, the two dimensions are separable and important; various combinations of integration types and cognitive demand levels result in nine different types of science tasks.

The utility of the TAGS is demonstrated by what it helps us notice that otherwise might remain hidden. First, the TAGS makes clear that not all tasks that integrate science content with the practices foster high-level thinking and reasoning. We suspect that Scripted Integration tasks can be easily mistaken for high-level tasks because they aim to engage students in scientific practices within the context of science content as recommended by the NRC Framework. The TAGS helps us to notice, however, that the integration occurs at a very shallow level; students are not required to make sense of what they are asked to do in the task. They can complete it without having to invest cognitive effort toward developing an understanding of the scientific ideas. Without the TAGS framework, this shallow implementation task type might not be noticed.

Second, the TAGS helps us to see important similarities and differences between the two high-level integration tasks: Doing Science and Guided Integration. Doing Science tasks require students to work on non-routine, ill-structured, complex activities. Many teachers are reluctant to use these kinds of tasks, believing that their students are not equipped to do the task on their own (Marx et al., 1994). Guided Integration tasks, because they can look very similar to Doing Science tasks, may cause some teachers to avoid using them as well or to use them without committing to the maintenance of their cognitive demand. The TAGS identifies Guided Integration tasks as having similar affordances as Doing Science tasks in terms of the integration of content and practices as well as engaging students in complex forms of thinking and reasoning—affordances that help students to develop deeper, more generative understandings regarding the nature of science, concepts, ideas and principles. However, the TAGS also highlights the ways in which Guided Integration tasks provide more guidance for students' engagement in scientific ideas and the practices of the discipline, which may reduce teachers' anxiety about their students' ability to engage with the task. This guidance will relieve some of the cognitive effort demanded of students, but not in ways that take away the sense making. All of this suggests that being able to recognize the similar benefits but different support structures associated with Doing Science vs. Guided Integration tasks may lead to more teachers selecting Guided Integration tasks than might otherwise be the case.

## Discussion: How Can the TAGS Support Researchers' and Practitioners' Work?

The development of a task-based framework in science was initially motivated by our observation of the value of the Task Analysis Guide for both mathematics education researchers and practitioners. With the release of the NRC (2012) framework, we felt that we had the skeleton (that is, a clear vision of science teaching and learning that promotes complex thinking about the disciplinary ideas AND the use of scientific practices) on which to build the TAGS.

As noted earlier, the TAG has played a variety of roles within mathematics education. It has served as an analytical tool for researchers to examine the cognitive demand of curricular materials (e.g. Jones & Tarr, 2007; Stein & Kim, 2009; Ubuz, Erbas, Cetinkaya, & Ozgeldi, 2010) and the quality of mathematics instruction (Boston & Smith, 2009; Stein & Kaufman, 2010). Moreover, it has served as the basis for a standardized instrument—the Instructional Quality Assessment (IQA) (Boston & Wolf, 2004) that has been used in several small (Boston & Smith, 2009, 2011) and large-scale studies that focus on improving mathematics teaching and learning at scale (e.g., MIST project; Cobb & Jackson, 2011). The TAG has also been used widely in professional development and teacher education programs across the United States as well as in studies of teachers' learning to maintain cognitive demand of tasks (e.g., Arbaugh & Brown, 2005; Boston & Smith, 2009; Silver, Clark, Ghousseini, Charlambous, & Sealy, 2007). We foresee the TAGS playing similar roles in science education research and improvement efforts.

The TAGS can be used to guide the design and evaluation of science curricula in various ways. It can be used to monitor the proportion of different types of tasks included in curricula. Not all tasks in the curricula may be designed at a high-level (i.e., Guided Content, Guided Practices, Guided Integration, or Doing Science tasks). Moreover, the proportion of different kinds of high-level tasks could differ (e.g., Guided Integration tasks may occur more often than do Guided Content tasks). Some topics within the curriculum may be covered at a surface level, mostly through Memorization Tasks or Scripted Tasks but others may be covered at a deeper level by requiring students to engage in high-level cognitive processes about the ideas and practices. Such proportional distribution of different types of tasks in the curriculum could provide insight about the quality of the curriculum and its overall focus.

The TAGS could be used to assess curriculum materials in terms of their alignment with standards. With the recent release of the NGSS, we believe that just focusing on content will not be adequate for such curriculum evaluation work. As stated in the recent standards documents, "The Framework specifies that each performance expectation must combine a relevant practice of science or engineering, with a core disciplinary idea and crosscutting concept. . . That guideline is perhaps the most significant way in which the NGSS differs from prior standards documents." While examples of other frameworks for judging the alignment of curricula to standards exist (Kesidou & Roseman, 2002; Stern & Roseman, 2004), we are not aware of any that specifically target the integration of content and scientific practices, especially at the level of individual tasks. For example, although the Educators Evaluating the Quality of Instructional Products (EQuIP) rubric (2014), which was recently developed by NSTA and Achieve, calls for judging the extent to which a lesson or unit "provides opportunities to develop and use specific elements of the practice(s) to make sense of phenomena and/or design solutions to problems," it does not provide details regarding how to evaluate integration at the task level. Moreover, by providing insight into the cognitive demand levels of tasks as well as identifying whether or not tasks integrate content and practices, our framework has the capacity to ferret out integration tasks that would be called "NGSS-aligned" but that would only engage students at a superficial level.

If the curriculum is designed to be educative (i.e., designed to be directly helpful to teachers, Ball & Cohen, 1996; Davis & Krajcik, 2005), the TAGS can provide clues to curriculum developers in terms of the tasks in the curriculum that will require more scaffolding for teachers to support effective implementation. For example, prior research revealed the challenges of implementing project-based activities, which can often be categorized as cognitively demanding (Guided Integration or Doing Science). Classroom management for maintaining productive independent work and being able to provide appropriate amounts of scaffolding were found to be difficult during the enactment of such materials (e.g., Blumenfeld et al., 1991; Marx et al., 1997). In recent work, we have begun to analyze videos of biology instruction in terms of the quality of enactment of a cognitively demanding Guided Integration task (Tekkumru-Kisa, Stein, & Schunn, 2015). This work reveals that students are sometimes reluctant to work on these tasks because they find them very challenging and different from what they are accustomed to, and teachers find it difficult when students do not make immediate progress and often prematurely step in and "take over the thinking." students' thinking on these kinds of tasks can decline into reiterating what they were told, or—with little help from the teacher—unsystematic and nonproductive exploration.

Such findings from research regarding the use of cognitively demanding tasks in science classrooms suggest that educative curriculum materials designed particularly for Guided Integration and Doing Science tasks will require support for teachers' effective enactment of these kinds of tasks in their classrooms. To illustrate, effective enactment of such tasks requires knowing how to help students understand authentic activities of the discipline such as how scientific knowledge is developed (i.e., pedagogical content knowledge aboutthe scientific practices). In this regard, several of the design heuristics for educative science curriculum materials developed by Davis and Krajcik (2005) are about supporting teachers' pedagogical content knowledge for scientific inquiry. They suggest that educative curriculum materials should support teachers for engaging students in designing investigations, making explanations based on evidence, collecting and analyzing data. Therefore, the TAGS could help to identify the tasks in the curriculum that will require this extra support for teachers (i.e., Guided Integration and Doing Science task).

The TAGS could be used to reveal the alignment between different types of tasks in the curriculum and the goals for students' learning. This is an important aspect of educative curriculum because it helps teachers by explicitly stating *what* students should learn and *how* (Stein, Remillard, & Smith, 2007). For example, if the goal is to increase students' accuracy in solving routine problems, it will be better to use Scripted Content tasks. If the goal is to help students understand how scientific knowledge develops, on the other hand, selecting or designing Guided Integration or Doing Science tasks will be more appropriate. Providing teachers with such rationales underlying the design of particular tasks in the educative curriculum materials can support teachers' learning to carefully select and use tasks based on the cognitive goals for students' learning.

The TAGS can also be used to support teacher learning. The vision of teaching and learning portrayed by the NGSS will not magically appear in science classrooms. Teachers will need support about the instructional practices that they should adopt to achieve NGSS vision in their classrooms (e.g., Osborne, Simon, Christodoulou, Howell-Richardson, & Richardson, 2013). For example, recently Kloser (2014) presented a set of core science teaching practices that might best help teachers support current goals in science education, such as ‘‘constructing and interpreting models’’ and ‘‘engaging students in investigations.’’ We claim that familiarity with the TAGS can also help teachers to monitor the opportunities for learning they are providing for their own students and help them to design, select, and implement tasks that are aligned with their goals. Developing such skills will not happen over night, however. In a recent study, we found that teachers had a difficult time identifying the scientific practices involved in some of the tasks used in their professional development (Tekkumru-Kisa, Stein, & Schunn, 2013). Teachers will need support to distinguish science tasks that provide different opportunities for student learning, especially those that engage students in scientific practices. In another related study, which was designed to support teachers' learning within a video-based professional development, teachers learned to notice important aspects of classrooms in which cognitively demanding tasks are enacted as emphasized in the NGSS. The TAGS was instrumental in the selection of the video cases for the professional development as well as the facilitation of the discussions about them (Tekkumru-Kisa, 2013; Tekkumru-Kisa & Stein, 2015).

Finally, the TAGS can be used for examining the quality of science instruction. It can be used to classify the type of tasks selected and set up in lessons as well as the extent to which the demands of high-level tasks are maintained during the enactment of those tasks. We have recently designed an instructional quality observation instrument based on the TAGS that captures if and how a teacher's instruction maintains the high levels of cognitive demand of tasks that are built into an engineering-based, high school biology unit developed as part of the Biology Levers Out Of Mathematics (BLOOM) project. The protocol provides insight into how and why students end up memorizing or simply following scripted procedures versus, sometimes, stopping to think and reason about the ideas or engaging in authentic scientific practices (Tekkumru Kisa et al., 2015).

A final advantage of the TAGS is that—while specific to the discipline of science—it is not topic- or grade-specific. It can be used to observe any type of science classrooms at any grade level, to design or evaluate curricular materials for any science content, and to plan for and enact professional development of teachers and coaches across science content areas. Moreover, because the TAGS names scientific practices which could be (re) defined in any given science education context, the TAGS could be used in higher education as well.

To conclude, we anticipate the TAGS playing a broad range of roles in science education research and improvement efforts. As noted above, we have already begun to use TAGS to support teachers' learning and curriculum development as well as to examine quality of science instruction. But more research is needed to examine the relationship between tasks within TAGS categories and student learning outcomes, as well as to better understand and refine the ways in which TAGS support research and improvement efforts in science education.

## References

Abd-El-Khalick, F., Waters, M., & Le, A. (2008). Representation of nature of science in high school chemistry textbooks over the past four decades. Journal of Research in Science Teaching, 45, 835–855.

Achieve. (2010). Taking the lead in science education: Forging Next-Generation Science Standards. Washington, DC: Author. Retrieved from http://achieve.org/files/InternationalScienceBenchmarkingReport.pdf.

Anderson, L.W., & Krathwohl, D.R. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. New York: Longman.

Arbaugh, F., & Brown, C. A. (2005). Analyzing mathematical tasks: A catalyst for change. Journal of Mathematics Teacher Education, 8(6), 499–536.

Ball, D. L., & Cohen, D. K. (1996). Reform by the Book: What is: or might be: The role of curriculum materials in teacher learning and instructional reform. Educational Researcher, 25(9), 6–14.

Bell, P., Bricker, L. A., Tzou, C., Lee, T., & Van Horne, K. (2012). Engaging learners in scientific practices related to obtaining, evaluating, and communicating information. The Science Teacher, 79(8), 17–22.

Berland, L. K., & Reiser, B. J. (2011). Classroom communities' adaptations of the practice of scientific argumentation. Science Education, 95(2), 191–216.

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. Educational Psychologist, 26(3&4), 369–398.

Boaler, J, & Staples, M. (2008). Creating mathematical futures through an equitable teaching approach: The case of Railside school. Teachers' College Record, 110(3), 608–645.

Boston, M., & Wolf, M. K. (2004). Using the Instructional Quality Assessment (IQA) Toolkit to assess academic rigor in mathematics lessons and assignments. Paper presented at the Annual Meeting of the American Educational Research Association Meeting, San Diago.

Boston, M. D., & Smith, M. S. (2009). Transforming secondary mathematics teaching: Increasing the cognitive demands of instructional tasks used in teachers' classrooms. Journal for Research in Mathematics Education, 40(2), 119–156.

Boston, M. D., & Smith, M. S. (2011). A 'Task-centric approach' to professional development: Enhancing mathematics teachers' ability to implement cognitively challenging mathematical tasks. The International Journal on Mathematics Education, 43(6–7), 965.

Calculating Speed. (n.d.). Retrieved from http://www2.franciscan.edu/academic/mathsci/mathscienceintegation/MathScienceIntegation-827.htm.

Cell Organelle Crossword Puzzle. (n.d.). Retrieved from http://images.pcmac.org/SiSFiles/Schools/NC/OnslowCounty/NewBridge/Uploads/Forms/Cell%20Organelle%20Crossword%20Puzzle.pdf.

Chi, M. T., De Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. Cognitive Science, 18(3), 439–477.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. Science Education, 86(2), 175–218.

Cobb, P., & Jackson, K. (2011). Towards an empirically grounded theory of action for improving the quality of mathematics teaching at scale. Mathematics Teacher Education and Development, 13 (1), 6–33.

Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. Educational Evaluation and Policy Analysis, 12(3), 311–329.

Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. Educational Researcher, 34(3), 3.

Doyle, W. (1983). Academic work. Review of Educational Research, 53(2), 159.

Doyle, W. (1988). Work in mathematics classes: The context of students' thinking during instruction. Educational Psychologist, 23(2), 167–180.

Educators Evaluating the Quality of Instructional Products (EQuIP) Rubric for Lessons and Units: Science (2014). Retrieved from: http://www.nextgenscience.org/sites/ngss/files/EQuIP%20Rubric%20for%20Science%20October%202014_0.pdf.

Evagorou, M., & Osborne, J. (2013). Exploring young students' collaborative argumentation within a socioscientific issue. Journal of Research in Science Teaching, 50(2), 209–237.

Fortus, D., Dershimer, R. C., Krajcik, J., Marx, R. W., & Mamlok-Naaman, R. (2004). Design-based science and student learning. Journal of Research in Science Teaching, 41(10), 1081–1110.

Gallagher, J. J., & Tobin, K. (1987). Teacher management and student egagement in high school science. Science Teacher Education, 71(4), 535–555.

Germann, P. J., Haskins, S., & Auls, S. (1996). Analysis of nine high school biology laboratory manuals: Promoting scientific inquiry. Journal of Research in Science teaching, 33(5), 475–499.

Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. Journal for Research in Mathematics Education, 28(5), 524–549.

Hiebert, J., & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. American Educational Research Journal, 30(2), 393–425.

Hodson, D. (1996). Laboratory work as scientific method: Three decades of confusion and distortion. Journal of Curriculum studies, 28(2), 115–135.

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. Science Education, 88(1), 28–54.

How roller coasters work. (n.d.). Retrieved from http://science.howstuffworks.com/engineering/structural/roller-coaster4.htm.

Jones, D. L., & Tarr, J. E. (2007). An examination of the levels of cognitive demand required by probability tasks in middle grades mathematics textbooks. Statistics Education Research Journal, 6(2), 4–27.

Kloser, M. (2014). Identifying a core set of science teaching practices: A delphi expert panel approach. Journal of Research in Science Teaching, 51(9), 1185–1217.

Kesidou, S., & Roseman, J. E. (2002). How well do middle school science programs measure up? Findings from Project 2061's curriculum review. Journal of Research in Science Teaching, 39(6), 522–549.

Krajcik, J. S., & Blumenfeld, P. C. (2006). Project-based learning. In R. K. Sawyer (Ed.), The Cambridge handbook of the learning sciences (pp. 317–333). New York: Cambridge University Press.

Krajcik, J. (2014). How to select and design materials that align to the Next Generation Science Standards. Retrieved from http://nstacommunities.org/blog/2014/04/25/equip/.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. Theory into practice, 41(4), 212–218.

Levin, D. M. (2008). What secondary science teachers pay attention to in the classroom: Situating teaching in institutional and social systems. Unpublished Doctoral Thesis, University of Maryland at College Park.

Levin, D. M., Hammer, D., & Coffey, J. E. (2009). Novice teachers' attention to student thinking. Journal of Teacher Education, 60(2), 142–154.

Levin, D., Hammer, D., Elby, A., & Coffey, J. (2013). Becoming a responsive science teacher: Focusing on student thinking in secondary science. Arlington, Virginia: NSTA Press.

Lewis, S. E., & Lewis., J. E. (2005). Departing from lectures: An evalua- tion of a peer-led guided inquiry alternative. Journal of Chemical Education, 82(1), 135–139.

Lord, M. (2014). Wind power for your home. Retrieved from http://teachers.egfi-k12.org/wind-power-for-home/.

Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Chrostowski, S. J. (2004). TIMSS 2003 international science report. Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Marx, R. W., Blumenfeld, P. C., Krajcik, J., Blunk, M., Crawford, B., Kelly, B., & Meyer, K. M. (1994). Enacting project-based science: Experiences of four middle grade teachers. Elementary School Journal, 517–538.

Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., & Soloway, E. (1997). Enacting project-based science. The Elementary School Journal, 97(4), 341–358.

Match the Terms. (n.d.) In School of Dragons. Retrieved from http://www.schoolofdragons.com/resources/match-the-terms-view.

Mehalik, M. M., Doppelt, Y., & Schunn, C. D. (2008). Middle-school science through design-based learning versus scripted inquiry: Better overall science concept learning and equity gap reduction. Journal of Engineering Education, 97(1), 71–85.

Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis years 1984 to 2002. Journal of Research in Science Teaching, 47(4), 474–496.

Moog, R. S., & Spencer, J. N. (2008). Process oriented guided inquiry learning (POGIL). New York: Oxford University Press.

National Center for Education Statistics. (2003). Teaching mathematics in seven countries: Results from the TIMSS video study. Washington, DC: U.S. Department of Education.

National Research Council. (2002). Investigating the influence of standards: A framework for research in mathematics, science, and technology education. Washington, DC: The National Academies Press.

National Research Council. (2007). Taking science to school: Learning and teaching science in grades K-8. Washington, DC: Committee on Science Learning, Kindergarten Through Eight Grade.

National Research Council. (2012). A Framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Committee on Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and education. Washington, DC: The National Academies Press.

NGSS Lead States. (2013). Next Generation Science Standards: For States, By States. Achieve, Inc. on behalf of the twenty-six states and partners that collaborated on the NGSS.

Observing Osmosis (n.d.). Retrieved from http://www.biologycorner.com/worksheets/observing_osmosis.html#.U5Hbo5RdVD4.

Osborne, J., Simon, S., Christodoulou, A., Howell-Richardson, C., & Richardson, K. (2013). Learning to argue: A study of four schools and their attempt to develop the use of argumentation as a common instructional practice and its impact on students. Journal of Research in Science Teaching, 50(3), 315–347.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. Educational Psychologist, 38(1), 1–4.

Pickering, M., & Monts, D. L. (1982). How students reconcile discordant data: A study of lab report discussions. Journal of Chemical Education, 59, 794–796.

Polikoff, M. S., Porter, A. C., & Smithson, J. (2011). How well aligned are state assessments of student achievement with state content standards? American Educational Research Journal, 48(4), 965–995.

Porter, A. C., & Smithson, J. L. (2001). Are content standards being implemented in the classroom? A methodology and some tentative answers. Yearbook-National Society For The Study Of Education, 2, 60–80.

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. Educational Researcher, 31(7), 3–14.

Porter, A. C., Smithson, J., Blank, R., & Zeidner, T. (2007). Alignment as a teacher variable. Applied Measurement in Education, 20(1), 27–51.

Reiser, B. J., Berland, L. K., & Kenyon, L. (2012). Engaging students in the scientific practices of explanation and argumentation. Science and Children, 49(8), 8–13.

Roth, W. M. (1995). Authentic science: Knowing and learning in open-inquiry science laboratories. Dordrecht: Kluwer.

Roth, K. J., Druker, S. L., Garnier, H. E., Lemmens, M., Chen, C., Kawanaka, T., & Gallimore, R. (2006). Highlights from the TIMSS 1999 Video Study of eighth-grade science teaching (NCES 2006-17). U. S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Roth, K., & Givvin, K. B. (2008). Implications for math and science instruction from the TIMSS 1999 video study. Principal Leadership, 8(9), 22–27.

Schneider, R. M., Krajcik, J., Marx, R. W., & Soloway, E. (2002). Performance of students in project-based science classrooms on a national measure of science achievement. Journal of Research in Science Teaching, 39(5), 410–422.

Schwarz, C. V., & Gwekwerere, Y. N. (2007). Using a guided inquiry and modeling instructional framework (EIMA) to support preservice K-8 science teaching. Science Education, 91(1), 158–186.

Scientific Processes - Tools and Measurements. (n.d.) Retrieved from http://www.biologycorner.com/worksheets/scientific_processes.html.

Silver, E. A., Clark, L. M., Ghousseini, H. N., Charalambous, C. Y., & Sealy, J. (2007). Where is the mathematics? Examining teachers' mathematical learning opportunities in practice-based professional learning tasks. Journal of Mathematics Teacher Education, 10(4–6), 261–277.

Simple Genetics Practice Problems. (n.d.) Retrieved from http://www.biologycorner.com/worksheets/genetics_practice.html.

Smith, M. S., & Stein, M. K. (1998). Selecting and creating mathematical tasks: From research to practice. Mathematics Teaching in the Middle School, 3(5), 344–350.

Smith, M. S. (2000). Balancing old and new: An experienced middle school teacher's learning in the context of mathematics instructional reform. Elementary School Journal, 100(4), 351–375.

Spillane, J. P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. Educational Evaluation and Policy Analysis, 21(1), 1–27.

Stern, L., & Roseman, J. E. (2004). Can middle-school science textbooks help students learn important ideas? Findings from Project 2061's curriculum evaluation study: Life science. Journal of Research in Science Teaching, 41(6), 538–568.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classroom. American Educational Research Journal, 33(2), 455–488.

Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. Educational Research and Evaluation, 2(1), 50–80.

Stein, M. K., & Smith, M. S. (1998). Mathematical tasks as a framework for reflection. Mathematics Teaching in the Middle School, 3(4), 268–275.

Stein, M. K., Smith, M. S., Henningsen, M. A., & Silver, E. A. (2000). Implementing standards-based mathematics instruction: A casebook for professional development (First Edition). New York, NY: Teachers College Press.

Stein, M. K., Smith, M. S., Henningsen, M. A., & Silver, E. A. (2009). Implementing standards-based mathematics instruction: A casebook for professional development (Second Edition). New York, NY: Teachers College Press.

Stein, M. K., & Kaufman, J. H. (2010). Selecting and supporting mathematics curricula at scale. American Educational Research Journal, 47(3), 663–693.

Stigler, J. W., & Hiebert, J. (2004). Improving mathematics teaching. Educational Leadership, 61(5), 12.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive science, 12(2), 257–285.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. Educational psychology review, 10(3), 251–296.

Tarr, J. E., Reys, R. E., Reys, B. J., Chavez, O., Shih, J., & Osterlind, S. (2008). The impact of middle-grades mathematics curricula and the classroom learning environments on student achievement. Journal for Research in Mathematics Education, 39, 247–280.

Tekkumru-Kisa, M. (2013). Science teachers' learning to notice from video cases of the enactment of cognitively demanding instructional tasks. Unpublished Doctoral Thesis, University of Pittsburgh.

Tekkumru-Kisa, M., Stein, M. K., & Schunn, C. (2013). Teachers' learning to analyze cognitive demand of science tasks. Poster presented at the meeting of the American Educational Research Association, San Francisco.

Tekkumru-Kisa, M., & Stein, M. K. (2015). Learning to see teaching in new ways a foundation for maintaining cognitive demand. American Educational Research Journal. 52(1), 105–136.

Tekkumru-Kisa, M., Stein, M. K., & Schunn, C. (2015). Quality of science instruction during the enactment of NGSS-aligned, cognitively demanding science tasks. Paper to be presented at the National Association for Research in Science Teaching (NARST) International Conference, Chicago, IL.

Ubuz, B., Erbas, A. K., Cetinkaya, B., & Ozgeldi, M. (2010). Exploring the quality of the mathematical tasks in the new Turkish elementary school mathematics curriculum guidebook: The case of algebra. ZDM, 42(5), 483–491.

Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. Applied Measurement in Education, 20(1), 7–25.

Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). Looking inside the classroom: A study of K-12 mathematics and science education in the United States. Chapel Hill, NC: Horizon Research, Inc.

Weiss, I. R., & Pasley, J. D. (2004). What Is High-Quality Instruction? Educational Leadership, 61, 5.

## Supporting Information

Additional Supporting Information may be found in the online version of this article.