UC Irvine
UC Irvine Previously Published Works

Title
Creating Opportunities for Students to Show What They Know: The Role of Scaffolding in Assessment Tasks

Permalink
https://escholarship.org/uc/item/1q97v3bz

Journal
Science Education, 98(4)

ISSN
0036-8326

Authors
KANG, HOSUN
THOMPSON, JESSICA
WINDSCHITL, MARK

Publication Date
2014-07-01

DOI
10.1002/sce.21123

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Science
Education

LEARNING

Creating Opportunities for Students to Show What They Know: The Role of Scaffolding in Assessment Tasks

HOSUN KANG,[1] JESSICA THOMPSON,[2] MARK WINDSCHITL[2]
[1]*School of Education, University of California–Irvine, Irvine, CA 92697, USA;* [2]*College of Education, University of Washington, Seattle, WA 98195, USA*

Received 26 March 2013; accepted 27 March 2014
DOI 10.1002/sce.21123
Published online 22 May 2014 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: This study examines the ways in which teachers provide students with written scaffolds in assessment tasks and the impact of these on students' abilities to demonstrate a core disciplinary proficiency—constructing evidence-based explanations. Data include 76 assessment tasks designed by 33 science teachers and 707 samples of student work. We found five types of scaffolding embedded in assessments that allowed students to make their reasoning explicit: (a) using contextualized phenomena, (b) rubrics, (c) checklists, (d) sentence frames, and (e) encouraging students to draw explanatory models in combination with written explanation. Analyses showed that all five forms of scaffolding were significantly associated with the quality of student explanation even when controlling for teacher variance and student background. Providing contextualized phenomena had the greatest impact on the quality of student explanations, both by itself and in combination with other scaffolding. The results indicate that strategic combinations of scaffolds can prompt students across all achievement levels to more readily use what they know to produce evidence-based explanations, but that the scaffolding must be of high quality.
© 2014 Wiley Periodicals, Inc. *Sci Ed* **98:**674–704, 2014

Correspondence to: Hosun Kang; e-mail: hosunk@uci.edu
Contract grant sponsor: National Science Foundation.
Contract grant number: DRL-0822016.
Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of these organizations.

© 2014 Wiley Periodicals, Inc.

INTRODUCTION

Reform visions of science learning that focus on explanatory reasoning about complex bodies of knowledge raise expectations for student performance beyond basic competency levels (National Research Council [NRC], 2012a; Resnick, 2010). These expectations require new forms of teaching expertise in which professionals frame challenging tasks for students while creating varied opportunities for them to demonstrate what they know. Teachers can support both students' intellectual engagement and their demonstration of deeper learning by providing well-designed assessments.

Yet the power of assessment to reveal and support learning depends on how well student responses to tasks authentically reflect their thinking and understanding (Shepard, 2005). When such assessments are well designed, teachers gain insights into students' current ideas, gaps in understanding, and reasoning processes. With this information, teachers can adapt instruction based on learners' needs and strategically move students toward more advanced thinking. We refer here to structured forms of written assessment (as opposed to assessing through instructional conversations, e.g.) that are designed in advance by the teacher (see Ruiz-Primo & Furtak, 2007). Depending upon their scope and timing, such assessments may be considered formative or summative. However for the purposes of this study we make the assumption that revealing what students know is important at multiple times throughout a unit. In many of the cases we present later, the teacher has administered the assessment near the end of a unit (normally a characteristic of summative assessment) but used the responses to understand the impact of their instructional choices as well as to provide valuable feedback to students on their thinking (normally a characteristic of formative assessment). In this study then, we will focus more on *the potential of tasks to illuminate students' scientific understandings and abilities to construct evidence-based explanations*, and less on the question about where the use of such tasks might be labeled as formative or summative.

Little is known about the types of scaffolding teachers use within assessments, especially when designing written tasks, and how this can support student learning. We do know that the design qualities of the assessments themselves significantly affect the quality of produced responses (Herman, 1992; Supovitz, 2012), especially when the task requires higher levels of intellectual work and language use. Most previous research into the design qualities of assessments, however, has focused on externally developed assessments, not tasks designed by teachers (Supovitz, 2012).

In this study, we examined "medium-cycle assessments" (Supovitz, 2012) that teachers construct and administer within a unit of instruction to provide information about students' understanding of content. All our participating teachers used a framework to organize sequences of learning activities that were intended to support students' understanding of "big" science ideas by engaging them in the construction and revision of evidence-based explanations for a selected phenomenon throughout the unit (Windschitl, Thompson, Braaten, & Stroupe, 2012). The assessment tasks analyzed in this study were given to students typically 1 or 2 days before the last day of the unit to make the progress of students' ideas visible. The teachers used the information from the assessments in varied and productive ways, such as providing specific feedback to students on their responses, reviewing previous activities that students were confused about, having students talk to peers who constructed different explanations, and encouraging students to revise their explanation. We focused on teachers' use of scaffolding in assessment tasks to create opportunities for students to show what they had learned through 2 or 3 weeks of instruction. Specifically, we explore the impact on students' construction of evidence-based explanations—a core proficiency as described in recent consensus reports (NRC, 2012b). We examined the scaffolding embedded in 76 assessment tasks designed by science teachers and 707 samples of student work. These samples represent a range of student academic backgrounds (i.e., students who learn with difficulty, who are typical, and who learn easily).

THEORETICAL FRAMEWORK

Scaffolding and Assessment Tasks

Contemporary learning theories highlight two key ideas about the process of learning. First, learning involves the construction of knowledge. Second, learning and development are culturally embedded, socially supported processes (Bransford, Brown, & Cocking, 1999; Sawyer, 2006; Shepard, 2005). When learners actively participate in constructing knowledge in a culturally and socially supported learning environment, they gain deeper understandings, more generalizable knowledge, and greater motivation to use that knowledge in other settings (Brown & Campione, 1994; Smith, Maclin, Houghton, & Hennessey, 2000). These insights into the processes of learning draw researchers' attention to essential supports, or scaffolding, which must be provided in the learning environment.

Drawing upon sociocultural learning theory and Vygotsky's (1978) zone of proximal development, Shepard (2005) points out that scaffolding and formative assessment are "essentially the same thing" (p. 66) in that both are strategies that teachers use to move learning forward within the zone of proximal development. In formative assessment, teachers collect and analyze information about students' learning, then use that information to provide additional support to meet students' changing needs. With supports, learners can achieve a learning goal or produce what they cannot produce alone (or can only with difficulty). Scaffolding refers to various forms of material, social, linguistic, or conceptual assistance that can support students' reasoning, participation, and learning (Sawyer, 2006). When teachers engage in effective forms of assessment, they are likely to provide scaffolding, such as prompts or response structures that address learners' difficulties and are informed by previous student responses and classroom talk.

A critical component of scaffolding that distinguishes it from other forms of support is "fading" (Pea, 2004). According to Pea, work on tasks that are assisted by scaffolding should be achieved later without such assistance when the learner becomes proficient. From an instructional point of view, the decision of when and how to remove a support is crucially important in shaping learning opportunities—as much as the decisions about the forms of scaffolding themselves. In a science classroom, fading will be dependent upon the complexity of the tasks as well as the degree of learners' progress. Constructing evidence-based explanation is not only complex and intellectually challenging but also new to most students in K-12 science classrooms. The literature is unclear about forms of scaffolding that assist students' mastery of such intellectually challenging tasks, and how the scaffolds might be faded.

Studies of English language learners (ELLs) provide particularly useful insights of scaffolding in science classrooms because engaging with wide-ranging bodies of knowledge and the explanatory practices of the discipline impose significant linguistic demands on these students. The theories behind ELL scaffolding, however, are applicable to all students. Walqui (2006) proposed six types of instructional scaffolding for ELLs. Four of the six are particularly relevant to constructing evidence-based explanations. First, *instructional modeling* is providing clear examples of what is requested of students for emulation. The objects of this modeling (not to be confused with scientific modeling) encompass tasks and activities, but also "appropriate language use for the performance of specific academic functions" (p. 171). Constructing evidence-based explanations requires particular ways of using language, and instructional modeling as one form of scaffolding enables teachers to help students employ appropriate disciplinary discourses. *Bridging* refers to providing support that helps students connect their previous knowledge and understandings with new concepts and language. An important aspect of bridging is establishing a personal link between the student and the subject matter by "showing how new material is relevant to the student's life, as an individual, here and now" (p. 172). Traditional science teaching frequently fails to make connections to children's day-to-day lives (Moje et al., 2004; Warren, Ballenger, Ogonowski, Rosebery, & Hudicourt-Barnes, 2001). *Contextualizing* refers to the use of language in concrete sensory contexts. Examples of this form of scaffolding include using material manipulation, pictures, a few minutes of a film, and other types of realia. In science, students need help working with the decontextualized, situation-independent, and dense academic language; contextualizing makes ideas and language more accessible and engaging for students. Finally, *developing metacognition* involves supporting learners' ability to monitor their current level of understanding and to decide when it is not adequate for a specified task. Examples of scaffolding here are rubrics or lists of steps for the routine being practiced. Variants and combinations of these forms of support can make it possible for students to productively engage in cognitively and linguistically challenging assessment tasks (Nasir, Rosebery, Warren, & Lee, 2006).

Disciplinary Proficiency Projected in a Written Assessment Work: Constructing Evidence-Based Explanations

The construction of explanations is an essential feature of science, as well as a fundamental classroom activity that engages students in epistemic practices of the discipline (Knorr-Cetina, 1999; Latour & Woolgar, 1979; Nersessian, 2005; Pickering, 1995). Recent reform documents, including the *Next Generation Science Standards* and its associated *K-12 Framework*, have highlighted causal explanation as a central practice (NRC, 2012b, 2013). There is a growing consensus among science educators that student-produced explanations—as opposed to the reproduction of textbook explanations—are evidence of deep science learning (Braaten & Windschitl, 2011; Ford & Wargo, 2012; NRC, 2000, 2007, 2012b). Written explanations, as artifacts of this core disciplinary practice, reflect not only students' conceptual understanding and reasoning but also their epistemic commitment to a specialized form of knowledge building (Ford & Wargo, 2012; Sandoval, 2003).

Scientific explanations are causal accounts for phenomena reflecting how one makes sense of events and processes in the natural world. Thoughtful scientific explanations articulate the underlying mechanisms, going beyond "explicating" observable phenomena to "theorizing" how and why things happen (Braaten & Windschitl, 2011). Constructing scientific explanation involves positing particular kinds of relationships, specifically that natural processes or events are attributed to a set of factors that produced the phenomenon (Ohlsson, 2002). During the process, unobservable mechanisms are distinguished from observable events or processes and connected to each other through coherent and principled reasoning. When students (and scientists) construct scientific explanations, they can engage in either reasoning from data or with known scientific ideas to develop/revise a coherent, causal explanatory model ("explanatory hypothesis") about how and why things happen. In this way, justifying elements of an explanatory model with sufficient evidence is one essential feature of scientific explanation. A good scientific explanation accounts for patterns in data and links claims in these accounts with relevant evidence (Sandoval & Millwood, 2005; Sandoval & Reiser, 2004).

In this study, we use the idea of "evidence-based explanation" rather than just "scientific explanation" to identify core disciplinary proficiencies projected in a written assessment task. Currently, the meaning of scientific explanation is underconceptualized, especially in relation to the practice of argumentation with evidentiary warrants.[1] Many researchers have combined the goals of explanation and argumentation, and then characterized scientific explanation as that which justifies explanations of scientific phenomena where claims are supported with appropriate evidence and reasoning (Furtak & Ruiz-Primo, 2008; McNeil & Krajcik, 2006; Ruiz-Primo, Li, Tsai, & Schneider, 2010). The quality of student-written explanations, then, have been examined referencing three components—claim, evidence, and reasoning—that originated from Toulmin's (1958) argument structure (Furtak & Ruiz-Primo, 2008; Ruiz-Primo et al., 2010). For example, Ruiz-Primo and colleagues (2010) evaluated the quality of student explanations in terms of (a) focus and accuracy of the claim, (b) type, nature, and sufficiency of evidence, and (c) the alignment and quality of the link between evidence and claim in reasoning. Although the three components are frequently cited as essential in scientific explanations (Kenyon & Reiser, 2006; McNeil & Krajcik, 2006; Sandoval & Reiser, 2004), using the components of argumentation to evaluate the quality of scientific explanations gives rise to several issues, such as looking past students' opportunities to theorize "how and why" things happen—beyond justifying a claim.

By using the idea of evidence-based explanation, we intend to examine the quality of student explanations, considering students' capability to theorize how and why things happen as well as justify their working theories with the use of evidence. The following describes these criteria in detail.

Conceptualizing the Quality of Student Explanations

Building on previous studies and grounded in the analysis of 707 samples of student work, we propose four dimensions of explanation that mark important differences in the quality of written accounts:

1. *The conjectural framing of explanations*: How observable natural phenomena and unobservable processes or ideas are treated in the explanation
2. *The role of evidence*: How explanations are supported with observation or data
3. *The depth of explanation*: The degree to which explanations provide comprehensive and gapless accounts for focal phenomena, including causal relationships and underlying mechanisms
4. *Causal coherence*: How explanations are logically consistent with data, observation, evidence (i.e., internal coherence) as well as generally accepted scientific principles and theories (i.e., external consistency)

[1] Scientific explanation is similar to argumentation in that both practices involve reasoning from the data, giving special credit to evidentiary support and generating a tentative conclusion. In the case of scientific explanation, this tentative conclusion is perceived as a current best explanatory hypothesis. However, scientific explanation and argumentation differ in their goals and the entities that invoke each practice. The goal of scientific explanation is "To provide an account that offers a plausible causal mechanism" about a natural phenomenon (Osborne & Patterson, 2011, p. 634). The entity for scientific explanation is the feature or phenomenon that is observed. Therefore, typical scientific explanations consist of a statement of the feature or a phenomenon and causal accounts for it. In contrast, the goal of argumentation is to persuade or "To provide incontrovertible warrants that support the claim and to show that it is a justified belief" (Osborne & Patterson, 2011, p. 634). The entity that invokes argumentation is the validity of claims or any explanation, not natural events or phenomena. The primary focus of argumentation is to examine the link between claims and evidentiary warrants (i.e., whether the claim is well supported with quality evidence) with the goal of evaluating or justifying the validity of claim, not understanding the relevant natural phenomena.

This framework guides the evaluation of student explanations. The following describes four dimensions of evidence-based explanations that reflect the quality of reasoning by students.

The Conjectural Framing of Explanation. Explanations can be characterized as either narrated or constructed as reflecting the modes of thought involved in the processes (Bruner, 1985). *Narrated explanations* take a form of a "correct version" of a story about some natural phenomenon. Students (re)produce uniform textbook-like explanations without significant variation. The links between observable phenomena and unobservable scientific ideas are not clear, and the tentative, revisable, and testable features of the explanation are not evident. In contrast, *constructed explanations* show causal links between observable phenomena and a proposed explanatory mechanisms. These explanations incorporate claims and reasoning. In some cases, students' explanations are constructed by reasoning through data and at other times by principled reasoning with scientific ideas. Students usually reveal a wider spectrum of understandings when they construct explanations about natural phenomena as opposed to reproducing textbook explanations.

The Role of Evidence. The second dimension of explanation is the role that evidence plays. Evidentiary support is a key feature of explanation that indicates students' conceptions of the epistemic nature of the discipline (McNeil & Krajcik, 2006; Sandoval, 2003). In science explanation studies, two characteristics of evidence—appropriateness and sufficiency—are frequently used (Kenyon & Reiser, 2006; McNeil & Krajcik, 2008; Sandoval, 2003). Appropriateness concerns whether the data cited are relevant to the problem, and sufficiency involves whether sufficient and credible data are provided to warrant the claim. Ruiz-Primo and her colleagues extended this framework to better capture the quality of evidence. They examined (a) type (i.e., what type of evidence did the student provide?), (b) nature (i.e., did the student focus on patterns of data or isolated examples?), and (c) sufficiency (i.e., did the student provide enough evidence to support the claim?). Building on these previous studies, we have characterized explanations with special attention to the roles of evidence: (a) explanations with no support (i.e., no use of evidence), (b) evidence-referring explanations, and (c) evidence-based explanations. *Evidence-referring explanations* have some forms of evidence, but the connection between evidence and claim is not sufficient or appropriate. For example, students simply refer to activities, information, or data as evidence without elaborating key patterns of the activity and how they support their claim. In contrast, *evidence-based explanations* are supported by data or observations that are directly related to the object of explanation.

The Depth of Explanations. The third dimension of student explanation is the degree to which explanations provide comprehensive and in-depth accounts of observable phenomena or events. The depth of explanations—providing a best explanatory model of how and why things happen—is the heart of scientific explanation. Drawing upon Braaten and Windschitl's (2011) framework, we have characterized explanations as (a) "what" explanations, (b) simple causal explanations or "how" explanations, and (c) "why" explanations. "*What" explanations* focus on describing observations in terms of patterns, without suggesting cause. *Simple causal explanations* focus on a causal relationship of one observable event affecting another with little attention to underlying mechanisms or principles (i.e., what fundamentally influences observations). "*Why" explanations* provide causal stories for a phenomenon and use unobservable or theoretical ideas or processes to explain patterns of observations.

Causal Coherence. A good scientific explanation is internally consistent, thus providing well-connected accounts for focal phenomena. Such an explanation is also consistent with generally accepted scientific principles and theories. Using causal coherence as an analytic feature, student explanations can be characterized as (a) explanations with no causation expressed, (b) partially coherent explanations, or (c) coherent and consistent causal explanations. First, *explanations with no causation* have bits and pieces of information relevant to the problem, but the information is not organized so as to reveal the causal relationships. Such an explicating-type explanation (i.e., "what" explanation) is likely to provide no causation. *Partially coherent explanations* show a chain of reasoning that is not fully coherent internally or not consistent with other science ideas. For example, students may produce an interesting and plausible theory of how and why things happen, but such theory may be scientifically inaccurate (i.e., inconsistent with generally accepted scientific ideas). In other cases, some part of the explanation may be coherent and logical, but other parts conflict with the referenced data or evidence. *Coherent and consistent causal explanations* show a logical chain of reasoning; the link between evidence and an explanatory model is adequate and substantial (i.e., internally consistent), and the proposed explanatory model is consistent with generally accepted scientific ideas (i.e., externally consistent).

The Use of Scaffolding and Opportunities to Construct Evidence-Based Explanations

Studies have documented various forms of scaffolding used to support students' authentic disciplinary talk and writing. Overall, the scaffolding discussed in the science education literature can be characterized as structure-oriented claim, evidence, reasoning (CER) and explanation-oriented scaffolding.

Structure-oriented CER scaffolding is the most popular form of support and focuses primarily on providing *the structure* of scientific explanations inspired by Toulmin. Sometimes, researchers also provide additional support to help students incorporate concepts into explanations (see Kenyon & Reiser, 2006; McNeil & Krajcik, 2006; Songer & Gotwals, 2012). For example, Songer and Gotwals (2012) investigated how "scaffold-rich assessments" support young students' explanations. One task was to construct an explanation given a specified scientific question, "Is the large fish [in an ecosystem] a producer or a consumer?" This assessment task was laid out with three boxed spaces that had the subheadings of "Make a CLAIM," "Give your REASONING," "Give your EVIDENCE." The authors also provided detailed prompts for each component (e.g., "Make a CLAIM: Write a sentence that answers the scientific questions") along with concept support (e.g., "Hint: Think about how producers and consumers get energy"). Similarly, Kenyon and Reiser (2006) provided both an "explanation framework" consisting of CER and criteria for evaluating the quality of each structural component. CER scaffolding provides opportunities for students to structure their explanations, reminding them of relevant scientific concepts that they need to consider. Students are provided opportunities to learn what scientific explanations look like and to craft their explanations following this guidance. Using these supports, Songer and Gotwals (2012) reported that students' conceptual understanding increased based on pre- and posttest comparisons.

Explanation-oriented scaffolding has a similar objective to CER structure-oriented scaffolding but tends to feature the blending of conceptual and epistemic scaffolds. More of an emphasis is placed on explaining authentic scientific phenomena, rather than a single lab-based classroom experiences. For example, Sandoval and Reiser (2003, 2004) developed a computer-based tool, "Explanation Constructor," to support students' construction and evaluation of explanations about a situated example of natural selection (e.g., "Why are the finches that survive able to survive?"). Explanation Constructor provides conceptual scaffolds, such as a series of sentence frames that guide students to consider key patterns (e.g., "The existing variation in the population before the pressure is . . . "), how things happen (e.g., "How has the distribution of organisms in the population with this trait changed?), and why (e.g., "The survivors are the most fit under this pressure because they have these traits . . . .that enable them to . . . "). Explanation Constructor also provides evidence such as data or figures adjacent to the explanation space, thus prompting students to use evidence in writing their explanation. As Sandoval and Reiser (2004) noted, these computer-based scaffolds mediate students' activity by acting as "enablers," but their role depends on students' understanding of the purpose of their work and the affordances for action that are shaped by social interaction between teacher and students.

The nature of opportunities created with the specific use of scaffolding has significant implications for the participation of students in disciplinary activities, especially students for whom science represents a different way of knowing, talking, or doing than is prevalent in their life experiences (Moje, Collazo, Carrillo, & Marx, 2001; Rosebery, Ogonowski, DiSchino, & Warren, 2010). The construction of written explanations involves negotiating meaning, mediated by new language and varied texts. Determining how to negotiate the multiple texts, discourses, and knowledge available within the learning community can be challenging to students from nondominant cultural and linguistic backgrounds (Moje et al., 2001; Rosebery et al., 2010). Importantly, well-designed scaffolding makes it possible to provide academically challenging instruction for students who typically are underserved in secondary schools (Walqui, 2006).

RESEARCH QUESTIONS

The following research questions are addressed:

1. What types and combinations of scaffolding do teachers most often use when designing written assessment tasks?
2. How does each type of scaffolding relate to the extent and quality of students' scientific explanations?
3. Do certain levels of quality and combinations of scaffolding influence the quality of students' explanations more than others?

METHODS

Research Context and Participants

We employed a mixed-methods approach (Creswell & Plano Clark, 2011) to study how and why particular forms of scaffolding embedded in assessments support students' construction of written evidence-based explanations. Assessment tasks and samples of student work were collected from 33 first-year science teachers who participated in induction activities between 2010 and 2012. All the teachers graduated from a teacher education program at a public university in the United States between 2010 and 2011 and taught science at the secondary level in local communities. They received support from the university in their first year of teaching. Throughout the 2 years from preparation through the first year of teaching, they were exposed to the resources and tools for reform-oriented science teaching that emphasized students' construction of evidence-based explanations. The teachers participated in three sessions of collegial analyses of student work artifacts, facilitated by the university research team during their first year of teaching (for details, see Thompson et al., 2009; Windschitl et al., 2011). On these occasions, teachers were asked to bring samples

TABLE 1
The Coding Scheme About the Characteristics of Scaffolding in Assessment Tasks

Type of Scaffolding	Level 0 (code = 0)	Level 1 (code = 1)	Level 2 (code = 2)
Drawing in combination with writing	No drawing	Generic drawing/ Posterizing	Modeling
Contextualizing phenomena	Generic phenomena	Contextualizing	NA
Checklist	No checklist	Simple words checklist	Explanation checklist
Rubric	No rubric	Generic rubric	Comprehensive rubric
Sentence frame	No sentence frame	Focusing	Connecting

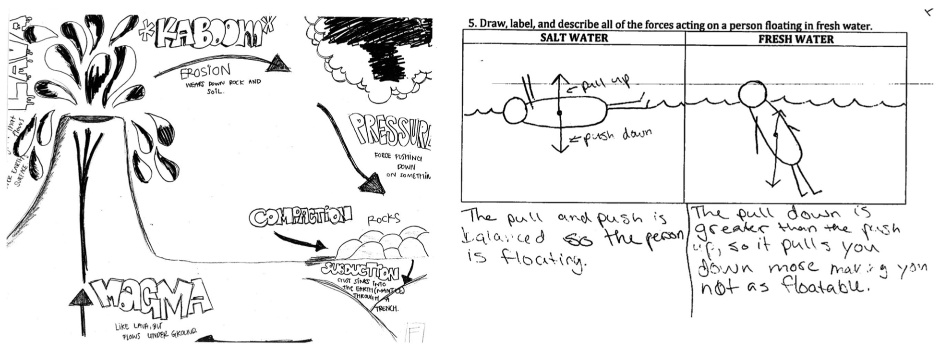
of student explanations. Teachers were provided a rubric for evaluating samples of student work prior to the collaborative analysis of these artifacts. Specifically the rubric asked teachers to evaluate their own students' work based on the (1) degree to which the student made comparisons among pieces of evidence, (2) degree of depth in student's explanation, and (3) degree to which evidence and explanations were integrated in written products. In each session, teachers brought their assessment tasks along with 9–12 samples of work from students with varied academic backgrounds. They were asked to bring three to four samples of student work from students who appeared to learn new ideas easily, the same number from students who were in the midrange of academic achievement, and the same number who appeared to learn with difficulty in their classrooms. We also asked teachers to include one or two samples of students who had special needs, such as ELLs. The 76 assessment tasks were approximately evenly distributed across the 33 teachers. These assessments and the samples of student work from the induction activities became the sources of data.

Data Sources and Measures

We analyzed 76 assessment tasks and 707 copies of student work. The collected student work consisted of even percentages of students with different academic backgrounds (students who learn with difficulty (30.8%), typical students (29.3%), students who learn easily (29.7%)). Of the student work, 72 copies did not have identification (10.2%) and, thus, were excluded from regression analyses. The types of scaffolding in the assessment tasks and the quality of student explanation in student work were coded using schemes described in the following section—one set around scaffolding and one set around the quality of student explanation.

Use of Scaffolding in Written Assessment Tasks

We found five salient modes of scaffolding from this initial analysis: (a) allowing students to draw in combination with writing, (b) contextualizing the explanation within a focal phenomenon or event, (c) providing checklists, (d) using rubrics, and (e) providing sentence frames. Within each type of scaffolding, there existed different levels of design sophistication. The main features of each type of scaffolding, and how they were converted quantitatively, are shown in Table 1.



Posterizing **Modeling**

Figure 1. Drawing in combination with writing.

Scaffolding by Prompting Drawing in Combination With Writing. Many assessment tasks allowed students to explain focal phenomena with drawing and writing. We found two different categories for prompts for drawing: One was generic drawing or posterizing (these were at the lower level of sophistication), and the other was modeling (see both examples in Figure 1). Generic drawing asked students to illustrate any aspect of the focal phenomena without clear guidance. Posterizing prompted students to reproduce models that could be found in the textbook and that illustrated some known set of discrete/unproblematic relationships, such as in the rock cycle or a "standard" volcano eruption. Both generic drawing and posterizing were categorized as lower level (Level 1) because this form of drawing, as scaffolding, simply provided alternative ways to express canonical ideas. In contrast, when students engaged in modeling, the work of drawing itself engaged them in higher levels of intellectual work.

Drawing as modeling usually provided designated spaces, structures, or templates for drawing, such as boxes, an outline of the sun, human body, or an enlarged blank inset. In contrast, the samples of student drawing that were coded as posterizing often fail to include any structure or prompt in the task design as shown in Figure 1. A few noticeable characteristics appeared in the prompts for modeling. First, students were prompted to illustrate *unobservable underlying mechanisms* that cause an observable event or phenomenon. For example, in one assessment of cell membrane mechanisms in ninth-grade biology (see Figure 2), students were prompted to "draw [a] scientific diagram" showing "what is happening that we can't see!" In some assessments, students were prompted to illustrate something over changes in time (e.g., draw what happens before, during, and after), temperature, and concentration (e.g., low vs. high). Another characteristic was that students were prompted to illustrate how events happened at an appropriate *scale* (e.g., cellular level, molecular level). Occasionally, students were not prompted to draw at a particular scale; in these cases, students did not use their drawing to explain ideas effectively. For example, in an assessment about cancer, students were prompted to draw what would happen to someone who had cancer before, during, and after, the onset of the disease but without specification of the scale. In this case, students illustrated a person at the organism level (i.e., drawing of human body), instead of what would happen at the cellular level, which made the drawings less useful for revealing in-depth explanations of cancer growth. We coded models using a scale of 0–2 (0 = no drawing, 1 = generic/posterizing, 2 = modeling) for regression analysis.

Scaffolding by Contextualizing a Focal Phenomenon. There was substantial difference in both the nature of explanation and the ways in which focal phenomena for explanation

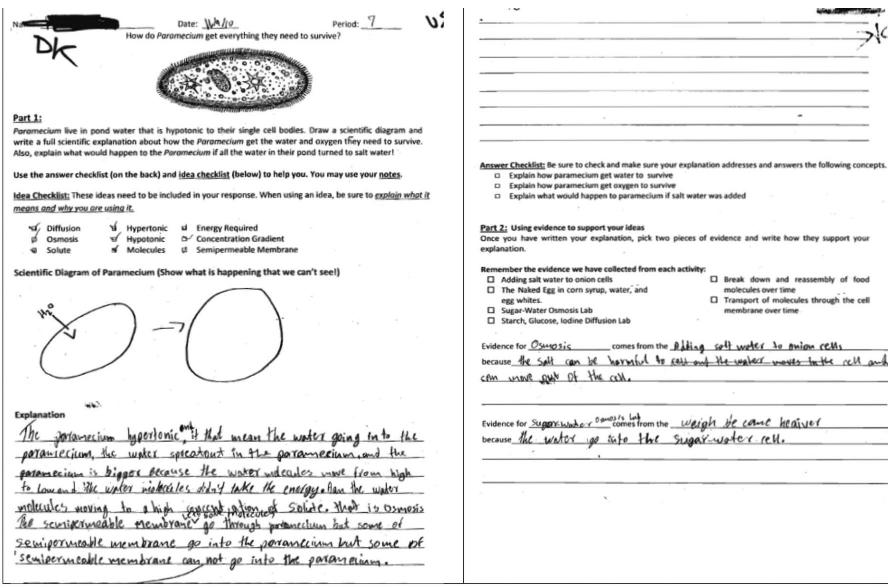


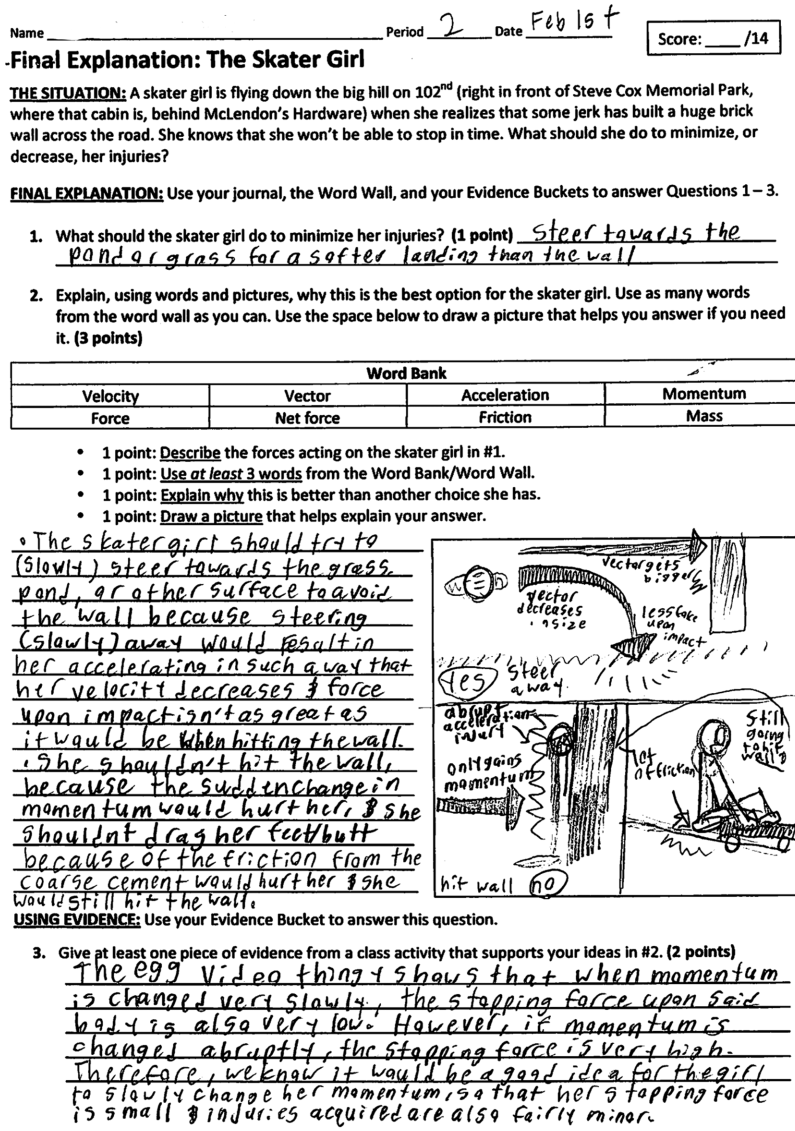
Figure 2. Cell membrane assessment: How does the paramecium get everything it needs to survive?

were framed in assessments. One group of assessment tasks asked students to explain general phenomena, "Why do siblings look different?," "Why is the equator hotter than the poles?," and "Why do the seasons change?" Often these assessments asked students to explain scientific ideas rather than an observable event, such as "What is homeostasis and why is it important to our body?" Representations of those events or phenomena usually appeared in the textbook. In contrast, some teachers contextualized a phenomenon or event in a particular time, place, and situation. For example, instead of asking about generic seasonal changes, one teacher asked, "Why don't countries near the equator, like Samoa, seem to have seasons like we do here in Seattle?" In another assessment in a unit on force and motion for the seventh grade, a teacher contextualized the physics in the form of "the skater girl"—a young woman in a local community where the school is located (see Figure 3):

> A skater girl is flying down the big hill on 102nd (right in front of Steve Cox Memorial Park, where that cabin is, behind McLendon's Hardware) when she realizes that some jerk has built a huge brick wall across the road. She knows that she won't be able to stop in time. What should she do to minimize, or decrease, her injuries? Explain why this is the best option for the skater girl.

For regression analysis, a noncontextualized idea or event in assessments was coded as 0 and the contextualized phenomenon or event was coded as 1.

Scaffolding by Providing a Checklist. The third form of scaffolding in the assessment task was providing one or a set of idea checklists to be referenced while constructing the explanation. We found two different kinds of checklists. The first is a "simple checklist" that lists concepts or scientific terms. This was often provided as a word bank in a box (see Figure 3). The other was an "explanation checklist" that prompted students to explain multiple aspects of the focal event as well as some relationships among ideas,


Name ______________________ Period 2 Date Feb 1st | Score: ____/14

Final Explanation: The Skater Girl

THE SITUATION: A skater girl is flying down the big hill on 102nd (right in front of Steve Cox Memorial Park, where that cabin is, behind McLendon's Hardware) when she realizes that some jerk has built a huge brick wall across the road. She knows that she won't be able to stop in time. What should she do to minimize, or decrease, her injuries?

FINAL EXPLANATION: Use your journal, the Word Wall, and your Evidence Buckets to answer Questions 1 – 3.

1. **What should the skater girl do to minimize her injuries? (1 point)** Steer towards the pond or grass for a softer landing than the wall

2. **Explain, using words and pictures, why this is the best option for the skater girl. Use as many words from the word wall as you can. Use the space below to draw a picture that helps you answer if you need it. (3 points)**

Word Bank			
Velocity	Vector	Acceleration	Momentum
Force	Net force	Friction	Mass

- 1 point: Describe the forces acting on the skater girl in #1.
- 1 point: Use *at least* 3 words from the Word Bank/Word Wall.
- 1 point: Explain why this is better than another choice she has.
- 1 point: Draw a picture that helps explain your answer.

• The skater girl should try to (slowly) steer towards the grass, pond, or other surface to avoid the wall because steering (slowly) away would result in her accelerating in such a way that her velocity decreases & force upon impact isn't as great as it would be when hitting the wall. • She shouldn't hit the wall, because the sudden change in momentum would hurt her, & she shouldn't drag her feet/butt because of the friction from the coarse cement would hurt her & she would still hit the wall.

USING EVIDENCE: Use your Evidence Bucket to answer this question.

3. **Give at least one piece of evidence from a class activity that supports your ideas in #2. (2 points)** The egg video thingy shows that when momentum is changed very slowly, the stopping force upon said body is also very low. However, if momentum is changed abruptly, the stopping force is very high. Therefore, we know it would be a good idea for the girl to slowly change her momentum, so that her stopping force is small & injuries acquired are also fairly minor.

Figure 3. Force and motion assessment: What should the skater girl do to minimize her injuries?

observations, and key patterns. For example, in a cell membrane assessment (see Figure 2), a teacher asked students to explain how a paramecium survives in pond water, providing the following explanation checklist. This scaffolding was framed by the teacher as an "Answer Checklist":

Answer Checklist: Be sure to check and make sure your explanation addresses and answers the following concepts:
☐ *Explain how paramecium gets water to survive.*
☐ *Explain how paramecium gets oxygen to survive.*
☐ *Explain what would happen to paramecium if salt water is added.*

The explanation checklist occasionally appeared along with the simple checklist. The checklist as a form of scaffolding was coded using a scale of 0–2: 0 = no checklist, 1 = simple checklist, and 2 = explanation checklist.

Scaffolding by Providing a Rubric. Rubrics provided information about the essential attributes of a high-quality evidence-based explanation. A "simple rubric" (Level 1) provided prompts, such as "Once you have written your explanation, pick two pieces of evidence and write how they support your explanation." Sometimes, a Level 1 rubric provided a more elaborated list for getting full credit, as shown in the skater girl's assessment (see Figure 3):

- 1 point: Describe the forces acting on the skater girl in #1.
- 1 point: Use at least three words from the Word Bank/Word Wall.
- 1 point: Explain why this is better than another choice she has.

A simple rubric is similar to checklist except that a rubric shows associated credit points. In contrast, a "comprehensive rubric" (Level 2) took the form of a table with multiple rows and columns, elaborating expected levels of performance in detail. The performance expectation was developed by the teacher and, thus, did not always match with our criteria for evaluating the quality of an explanation. We coded different forms of rubrics using a scale of 0–2: 0 = no rubric, 1 = simple rubric, and 2 = comprehensive rubric.

Scaffolding With Sentence Frames. Two different forms of sentence frames appeared in the assessments: focusing versus connecting. Focusing sentence frames (Level 1) prompted students to draw their attention to the phenomena and explain them by providing linguistic lead-ins. For example, in the assessment about the gas laws using the phenomenon of changes in a balloon sitting on a table, students were prompted with a sentence frame like this: "What I saw was__________," "Inside [the balloon] the particles were__________." and "I know this because ___________." In contrast to focusing sentence frames, connecting sentence frames (Level 2) prompted students to make deeper connections among key components of scientific explanation, such as evidence and reasoning. For example, in a cell membrane assessment (see Figure 2), the teacher provided the following prompt to support students' use of evidence: Evidence for ________ comes from the _ [activity or reading]_ because___." Sentence frames were coded using a scale of 0–2: 0 = no sentence frame, 1 = focusing, and 2 = connecting.

Quality of Student Explanations

The quality of student responses was evaluated with respect to the four dimensions of good scientific explanation discussed in the conceptual framework (see Table 2). The four dimensions were (a) conjectural framing of explanation, (b) role of evidence, (c) depth of explanation, and (d) causal coherence. Each of the four dimensions was specified into three different levels. The first dimension, conjectural framing of explanation, had only two levels (0 and 1). For the three other dimensions, we assigned scores from 0–2 at each level depending on its sophistication (i.e., least sophisticated = 0, most sophisticated = 2). The unit of analysis was one sample of student work. The highest composite score was 7 when a sample was evaluated as being at the most sophisticated level of response across the four dimensions (i.e., a constructed, coherent and causal explanation that is strongly supported by evidence). The following shows three examples that illustrate different levels of sophistication (high, middle, and low). The first two examples are student explanation

TABLE 2
Criteria to Score Qualities of Explanations

Dimension	Level 1 (Code = 0)	Level 2 (Code = 1)	Level 3 (Code = 2)
Conjectural framing of explanation	Narrated • Mostly restating textbook explanation as a form of an unproblematic story • Observable phenomenon/event(s) is not distinguished from unobservable mechanism or scientific ideas	Constructed • Constructing explanation by reasoning through data • Constructing explanation through principled reasoning (applying) using scientific theories/ideas/models	
Role of evidence	Explanation with no support • Explanation is not supported by any form of evidence	Evidence-referring explanation • Simply referring activities or data as evidence • Some form of evidence is referred in the explanation, but the connection between evidence and explanation is weak. For example, referring the topic of activity rather than the key patterns of the activity that are related to the object of explanation	Evidence-based explanation • Highlighting key patterns of data or observations, or in activities to support a claim • Explanation is supported by strong evidence that is directly related to the object of explanation
Depth of explanation	"What" explanation • Describing observations in terms of patterns, without suggesting cause • "Explanation as explication" (Braaten & Windschitl, 2011)	Simple causal explanation or "how" explanation • Focusing on a causal relationship of one observable event affecting another with little attention to underlying mechanisms or principles (i.e., what fundamentally influence observation)	"Why" explanation • In-depth constructed explanations that provide full causal stories for a phenomenon, and use unobservable or theoretical events or processes to explain patterns of observations

(*Continued*)

TABLE 2
Continued

Dimension	Level 1 (Code = 0)	Level 2 (Code = 1)	Level 3 (Code = 2)
		• "Explanation as simple causation" (Braaten & Windschitl, 2011)	
Causal coherence (logical consistency, internal, and external coherence)	Explanation with no causation • Explanation covers bits or pieces of information about phenomena/events (illogical or incoherent list of information); there is no causal explanation about focal phenomena/event.	Partially coherent explanation • Partial explanation about the posed phenomena ("plausible-but-incorrect scientific explanation"); Some part of explanation is incoherent or illogical, conflict with data/evidence	Coherent and consistent causal explanation • Gapless and integrated explanation about the posed phenomena • Coherent and logical • No conflict with or among data/evidence ("causal coherence" internally consistent as well as consistent with generally accepted scientific principles and theories)

from the Skater Girl assessment (Figure 3), and the last one is a student response to an assessment task of a ninth-grade biology unit.

> **Example 1**—highly sophisticated explanation: A constructed, evidence-based, and in-depth "why" explanation with coherent and integrated reasoning (score = 7 of 7)
> The skater girl should try to (slowly) steer towards the grass, pond, or other surface to avoid the wall because steering (slowly) away would result in her accelerating in such a way that her velocity decreases & force upon impact isn't as great as it would be when hitting the wall. She shouldn't hit the wall, because the sudden change in momentum would hurt her, & she shouldn't drag her feet/butt because of the friction from the coarse cement would hurt her & she would still hit the wall. [Prompt: Give at least one piece of evidence from a class activity that supports your ideas] The egg video thingy shows that when momentum is changed very slowly, the stopping force upon said body is also very low. However, if momentum is changed abruptly, the stopping force is very high. Therefore, we know it would be a good idea for the girl to slowly change her momentum, so that her stopping force is small & injuries acquired are also fairly minor.

This is an example of student explanation that is scored as the highest level of sophistication. The observable event (steering slowly toward the grass, drag her feet/bottom, accelerating, hitting the wall, etc.) is clearly distinguished from unobservable mechanism or scientific ideas (changes in momentum, friction) (code = "constructed"). The observable events are linked to unobservable ideas through a coherent chain of cause and effect relationship (code = "coherent"). The explanation provides a relatively full causal story about various possibilities that the skater girl can choose (code = in-depth "Why"). Finally in terms of use of evidence, not only does the explanation refer to the source of evidence (e.g., the egg video) but the student also thoroughly describes key patterns appearing in the video that support the claim about the suggested skater girl's choice (code = "evidence-based explanation").

> **Example 2**—midrange level explanation: A narrated, evidence-referring, and simple "how" explanation with coherence (score = 4 of 7)
> I think she should roll on the grass to minimize the force. If she rolls onto the grass the force would spread all over her body so it wouldn't hurt as much. When you roll on the floor there is friction so it slows you down if you have a lot of ways. [Prompt: Give at least one piece of evidence from a class activity that supports your ideas] roll a ball onto the floor, it will stop at one point.

This is an example of explanation at the midrange level of sophistication, but still above the average scores of the total (average = 2.1). It was coded as "narrated," not "constructed" because of the connections between observable phenomenon and unobservable science ideas. In the previous example, the changes of speed and motion (observable/conceptual) are clearly discerned from the ideas of momentum and friction (unobservable), and the unobservable concepts were used to describe the mechanism for changes in speed and motion. In contrast, those epistemic distinctions are unclear among the ideas of "roll onto the grass," "the force would spread all over her body," and the process of getting injured. This explanation shows a causal coherence beyond describing "what" happens, but not all presented variables are taken into account, therefore, losing the point for "full, gapless in-depth why explanation." Finally, with respect to the use of evidence, one piece of evidence is cited, but there is no justification for how this pattern supports the proposed claim.

> **Example 3**—low level of sophistication: A narrated, without evidential support and what explanation with no causation (score = 0 of 7)

Because all student responses to the Skater Girl assessment were scored above the average, we pulled out an example of low-level sophistication from the other assessment task. The prompt was "describe how they (a bit of energy) travel through a biological system."

> [Prompt: "Write a story in your journal, pretend you are a bit of energy in this ecosystem. Tell me exactly how you got there, what processes you went through, and how you ended up (a story that will be continued later). Use as many details as you want, turn it into a comic, or make it into a song."]
> Student explanation: I am a sun light molecule, now I am going into this plant through photosynthesis. Now I am going through cellular respiration. I am turned into energy. Then an animal eats me and I am now glucose (sugar) and once again cellular respiration takes place turning me into energy.

In this example, the distinction between the observable phenomenon and unobservable ideas is unclear (code = "narrated"). The explanation is rather explicating a science story than constructing causal relationships (code = "what" explanation, explanation with no causation). There is no support with evidence (code = explanation with no support).

Analytical Approach

The first stage of data analysis focused on understanding descriptively how teachers used scaffolding in designing their assessments. The frequencies of both types and combinations of scaffolding embedded in assessment tasks were calculated. To determine the predictors that explain quality of student explanation, we computed Spearman's correlation coefficients among five types of scaffolding and quality of student explanation. The resulting scatter and box plots indicated a general linear relationship between each of the five scaffolding types and quality of student explanation.

Next, we examined the association between each type of scaffolding and quality of student explanation using hierarchical multiple regression analysis. We intended to use an initial model to examine whether the types of scaffolding would significantly predict the quality of student explanation if other factors were accounted for in advance. Both teacher effects and students' academic background (i.e., students who easily learn, are typical, and are underserved) were controlled as covariates. We created 33 teacher dummy variables and three student dummy variables and entered them in the first and second block of the hierarchical multiple regression analysis, respectively. The five forms of scaffolding were entered in the third block as predictors for quality of student explanations. The following equation describes the model for this hierarchical multiple regression:

$$Y_{\text{quality of explanation}} = \alpha + \beta_0 \times X_t + \beta_1 \times X_s + \beta_2 \times X_1 + \beta_3 \times X_2 + \beta_4 \times X_3 + \beta_5 \times X_4 + \beta_6 \times X_5 + \varepsilon$$

where Y: quality of student explanation, α: an intercept, $\underline{\beta_0}$: coefficients of 33 teacher dummy variables, $\underline{\beta_1}$: coefficients of three student group dummy variables, $\beta_{2\sim6}$: coefficients of scaffolding, X_t: 33 teacher dummy variables, X_s: three student dummy variables, $X_{1\sim5}$: five types of scaffolding, and ε: errors.

This first model helped us examine the general effect of the five types of scaffolding on the quality of student explanation. However, this model did not tell us the effect of each form of scaffolding that was coded as different levels within one type. Four of the five scaffolding types included two different levels of sophistication, such as posterizing (Level 1) versus modeling (Level 2) being levels of the variable "drawing in combination with writing." Furthermore, we were also interested in examining which *combinations* of scaffolding would best predict the quality of student explanation and the interaction effect between different types of scaffold. Accordingly, we created an additional eight dummy variables for the four scaffoldings that had two different levels and then ran a series of exploratory hierarchical multiple regression analyses. In these analyses, each form (level) of scaffolding used the predictors. A total of 11 models that included two to four types of scaffolding were created. Model 2 shows one example developed to examine the influence of two types of scaffolding combined—drawing and contextualized phenomena.

Model 2: Combinations of two scaffoldings, drawing and contextualized phenomena:

$$Y_{\text{quality of explanation}} = \alpha + \beta_0 \times X_t + \beta_1 \times X_s + \beta_2 \times X_{D1} + \beta_3 \times X_{D2} + \beta_4 \times X_P + \beta_5 \times X_{D1}X_P + \beta_6 \times X_{D2}X_P + \varepsilon$$

where Y: quality of student explanation, α: intercept, $\underline{\beta_0}$: coefficients of 33 teacher dummy variables, $\underline{\beta_1}$: coefficients of three student group dummy variables, $\beta_{2\sim6}$: coefficients of scaffolding, X_t: 33 teacher dummy variables, X_s: three student dummy variables, X_{D1}:

Drawing Level 1 (generic drawing or posterizing), X_{D2}: Drawing Level 2 (modeling), X_P: contextualized phenomena, and ε: errors.

Instead of considering all possible combinations, we developed statistical models using the combinations of scaffolding that teachers actually used. This allowed us to examine the effect of actual combinations of scaffolding used in assessments, and those models could be statistically interrogated with empirical data. Interactions among each form of scaffolding were examined in those models, but in the cases of combining more than three types of scaffolding in one model, we put only the interaction term of drawing with other scaffoldings. We made this decision because we were interested in the effect of drawing in combination with other types of scaffolding. Despite its most frequent use, drawing initially appeared as a negative predictor in the result of the overall multiple regression analysis (i.e., Model 1), which puzzled us with regard to the effect of drawing as a form of scaffolding. We also made this decision because of the natural tendency to increase the adjusted R^2 by adding more variables. From these exploratory multiple regression analyses, we first examined the impact of specific categories of scaffolding by comparing the standardized coefficients. Second, we examined which combinations of scaffolding best predict the quality of student explanation by looking at the *patterns* of the adjusted R^2, keeping it mind the natural increase of R^2 value with the increasing number of variables.

FINDINGS

What Types and Combinations of Scaffolding Do Teachers Use When Designing Explanation-Based Assessment Tasks?

Types of Scaffolding. To address the first research question, we analyzed 76 assessment tasks and found that the first-year science teachers used five types of scaffolding (see Table 3). The most frequently used were (a) encouraging drawing in combination with writing ($n = 42$, 55.3% of the 76 assessment tasks). Following this in frequency of use came (b) contextualizing a phenomenon ($n = 25$, 32.9%), (c) providing a checklist ($n = 21$, 27.6%), (d) providing a rubric ($n = 19$, 25.0%), and (e) providing sentence frames ($n = 10$, 13.1%). As shown in Table 3, only a few assessment tasks provided highly sophisticated forms of scaffolding (i.e., Level 2 scaffolding), such as encouraging drawing a model ($n = 17$, 22.4% of the 76 assessment tasks), providing explanation checklist ($n = 3$, 3.9%), or providing sentence frames that prompt students to make connections ($n = 2$, 2.6%).

Combinations of Scaffolding. The analysis showed that about 20% of assessment tasks did not provide any form of scaffolding ($n = 15$, 19.7%). About 40% of the assessments only provided one type of scaffolding ($n = 29$, 38.2 %), and one third provided two or three types of scaffolding ($n = 25$, 32.9%). Less than 8% of assessments provided four types of scaffolding ($n = 6$, 7.9%), and one assessment included all five types of scaffolding ($n = 1$, 1.3%). As shown in Table 4 and indicated by the total frequency of each type of scaffolding in Table 3, providing a space to draw explanations was the most popular type of scaffolding across all the assessments.

Relationship Between Each Type of Scaffolding and Quality of Student Explanation

The results from computing Spearman's correlations showed that all five scaffolding types were statistically significantly correlated with the quality of student explanations. The

TABLE 3
Frequency of the Five Types of Scaffolding Embedded in Assessment Tasks Depending on the Levels of Sophistication

Type of Scaffolding Number of Assessments (%)	Level 0: No Scaffolding Number (%)	Level 1: Moderate Number (%)	Level 2: More Sophisticated Number (%)
Drawing in combination with writing $N = 42$ (55.3)	No drawing $N = 34$ (44.7)	Generic drawing/posterizing $N = 25$ (32.9)	Modeling $N = 17$ (22.4)
Contextualizing phenomena $N = 25$ (32.9)	Generic phenomena $N = 51$ (77.1)	Contextualizing $N = 25$ (32.9)	NA
Checklist $N = 21$ (27.6)	No checklist $N = 55$ (72.4)	Simple words checklist $N = 18$ (23.7)	Explanation Checklist $N = 3$ (3.9)
Rubric $N = 19$ (25.0)	No rubric $N = 57$ (75.0)	Simple guideline $N = 12$ (15.8)	Comprehensive table rubric $N = 7$ (9.2)
Sentence frame $N = 10$ (13.1)	No sentence frame $N = 66$ (86.8)	Focusing $N = 8$ (10.5)	Connecting $N = 2$ (2.6)

range of the computed Spearman's rho (ρ) was from .46 to .11. Contextualized phenomena showed the strongest degree of association ($p < .01$, $\rho = .46$) following by the rubric ($p < .01$, $\rho = .41$), checklist ($p < .01$, $\rho = .26$), drawing ($p < .01$, $\rho = .18$), and sentence frame scaffolds ($p < .01$, $\rho = .11$).

The box plots suggested that overall some linear relationships exist between levels of sophistication at each type of scaffolding and the quality of student explanation, with the exception of the rubric (see Figure 4). With respect to the rubric, the quality of student explanation dramatically increased with the Level 1 rubric; in other words, providing simple guidelines as prompts were enough to support students in providing richer scientific explanations. With respect to drawing, engaging students in a generic form of drawing or posterizing (i.e., Level 1) was not significantly related to the quality of student explanation.

What Forms and Combinations of Scaffolding Predict the Quality of Student Explanation?

We wanted to further understand the roles that scaffolding played in the quality of student explanation, controlling for both teacher effect and student academic background. This section presents the results of hierarchical multiple regression analyses using scaffolding as a predictor. The roles of type, combination, and amount of scaffolding are examined both statistically and qualitatively.

The Overall Quality of Student Explanation. The overall average percentage score for quality of the 707 copies of student explanations was 2.1 points of 7, or 30.1% (see Table 5). Among the four dimensions of scientific explanation, the average score in the area of "use of evidence" was particularly low (14.5%). Less than 6% of student explanations (41 of 707) showed strong use of evidence to support students' explanations.

TABLE 4
Frequency of the Combinations of Scaffolding Used in Assessments

Number of Scaffoldings Used in One Assessment Tasks	Number of Assessments (%)	Combinations of Scaffolding (Number of Assessments)
Five types of scaffolding	1 (1.3)	• Drawing + contextualizing + rubric + checklist + sentence frame (1)
Four types of scaffolding	6 (7.9)	• Drawing + contextualizing + rubric + checklist (4) • Drawing + contextualizing + sentence frame + checklist (1) • Drawing + contextualizing + sentence frame + rubric (1)
Three types of scaffolding	8 (10.5)	• Drawing + contextualizing + rubric (3) • Drawing + rubric + checklist (3) • Drawing + contextualizing + checklist (2)
Two types of scaffolding	17 (22.4)	• Drawing + contextualizing (6) • Drawing + rubric (5) • Drawing + checklist (2) • Drawing + sentence frame (2) • Contextualizing + checklist (1) • Sentence frame + checklist (1)
One scaffold	29 (38.2)	• Drawing (13) • Contextualizing phenomena (6) • Checklist (6) • Rubric (2) • Sentence frame (2)
No use of scaffolding	15 (19.7)	NA

Using the Five Types of Scaffolding as Predictors for Overall Quality of Student Explanations. We ran hierarchical multiple regression analyses to determine the predictors that explain the quality of student explanations while controlling for teacher and student variances. As described in the preceding section, we developed Model 1 using the five types of scaffolding as the predictors for quality of student explanations, assuming a general linear relationship. The adjusted R^2 changes analysis suggested that about 58% of variance in quality of student explanation were explained by teacher variance, student academic background, and use of scaffolding. Specifically, teacher and student academic background explained about 38% and 9%, respectively. The change of the adjusted R^2 indicated that the use of scaffolding explained about 11% of the variance (see ${}_{adj.}R^2$s in Table 6).

Table 6 shows coefficients of the five scaffolding types in Model 1. Three scaffoldings (i.e., contextualized phenomena, providing a checklist, and using rubrics) were positively associated with the quality of student explanation ($p < .05$). Drawing was negatively associated ($p < .05$), and using sentence frames was positive but not statistically significant. The standardized coefficients suggested that contextualizing phenomenon is the strongest predictor of the quality of student explanation ($\beta = .39$), following by rubric ($\beta = .26$) and checklist ($\beta = .19$).

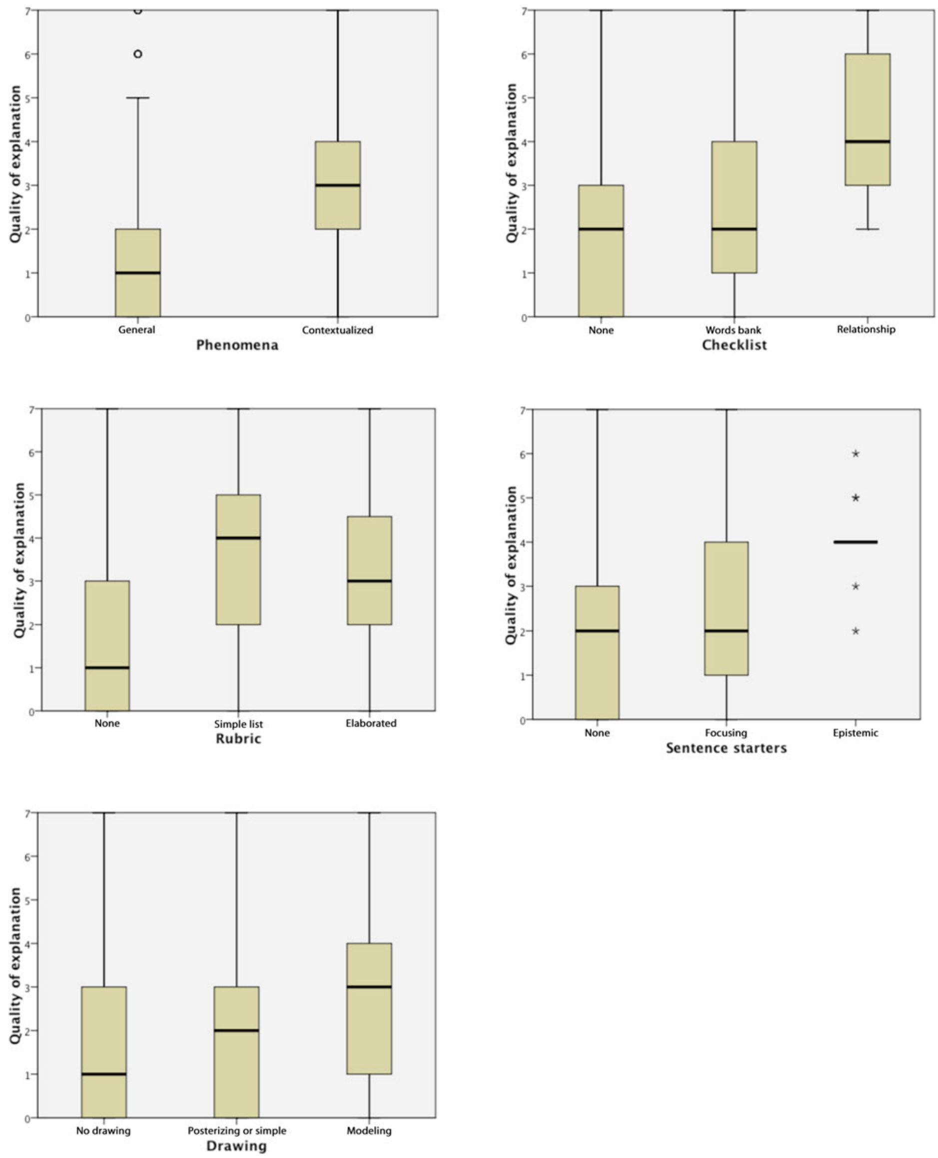


Figure 4. Relationship between each type of scaffolding and quality of student explanation.

TABLE 5
The Average Percentage Scores of Student Explanation

Parameter	Mean Percentage (*SD*)
Nature of explanation	37.0 (0.48)
Use of evidence	14.5 (0.57)
Depth of explanation	37.5 (0.72)
Coherence of reasoning	35.0 (0.61)
Total	30.1 (1.87)

TABLE 6
Hierarchical Multiple Regression Results for the Impact of Five Types of Scaffolding on Quality of Student Explanation

	B	*SE*	*β*	*T*
First block				
33 Teachers				
$F(30, 604) = 14.17$, $p < .001$, adj.$R^2 = .38$				
Second block				
Students who are typical	0.66	0.12	.16	5.47***
Students who learn easily	1.34	0.12	.33	11.20***
F_{change} $(2, 602) = 52.93$, $p < .001$, $R^2_{\text{change}} = .09$				
Third block				
Contextualizing phenomena	1.55	0.17	.39	9.31***
Rubric	0.75	0.12	.26	6.35***
Checklist	0.65	0.13	.19	5.17***
Drawing	−0.49	0.13	−.20	−3.75***
Sentence frame	0.13	0.17	.03	.77
F_{change} $(5, 597) = 30.78$, $p < .001$, $R^2_{\text{change}} = .10$				
The overall model: $F(37, 597) = 24.52$, $p < .001$, adj.$R^2 = .58$				

Notes: [a]A total of 36 dummy variables were created with respect to 33 teachers and three groups of students (students who learn easily, are typical, and underserved). Underserved students are reference group (coded as 0). The coefficients for 33 teachers are not reported in this table for its brevity.
***$p < .001$.

Using Each Form of Scaffolding as a Predictor for Quality of Student Explanations. The first purpose of these analyses was to examine the effect of different levels of *sophistication* within a type of scaffolding (e.g., posterizing vs. modeling, simple checklist vs. explanation checklist). The second purpose was to compare the effect of different *combinations* of scaffolding (i.e., combination of two or more forms of scaffolding). Finally, we intended to examine the interaction effect among scaffolding types, particularly with drawing.

Examination of the standardized coefficients of each form of scaffolding across 11 regression models suggested that using contextualized phenomena is the strongest single predictor of the quality of student explanation. This result is consistent with the result from the previous regression analysis (Model 1). Both rubric and checklist were also significant predictors, as suggested in Model 1. A Level 2 checklist (i.e., explanation checklist) is a stronger predictor than a Level 1 checklist (i.e., simple word list). Interestingly, however, a Level 1 rubric (i.e., simple rubric) is a stronger predictor than a Level 2 rubric (i.e., elaborated rubric).

In terms of the effect of combinations of scaffolding, the most powerful combination was the one that had three or more upper level scaffolding types that *included contextualized phenomena*. Most statistical models from the combinations of three or four scaffoldings show consistently higher values of the adjusted R^2 (.56–.57). In contrast, using two scaffoldings showed some spectrum of the adjusted R^2 from .50 to .55, indicating that the selection of forms of scaffolding is critical. The following section features a case of a teacher using a task with three upper level scaffolds (*contextualizing a phenomena*, *drawing as modeling*, and a *comprehensive rubric*) and one low level scaffold (a “focusing” sentence frame). The case illustrates the nature of opportunities created with use of scaffolding for student

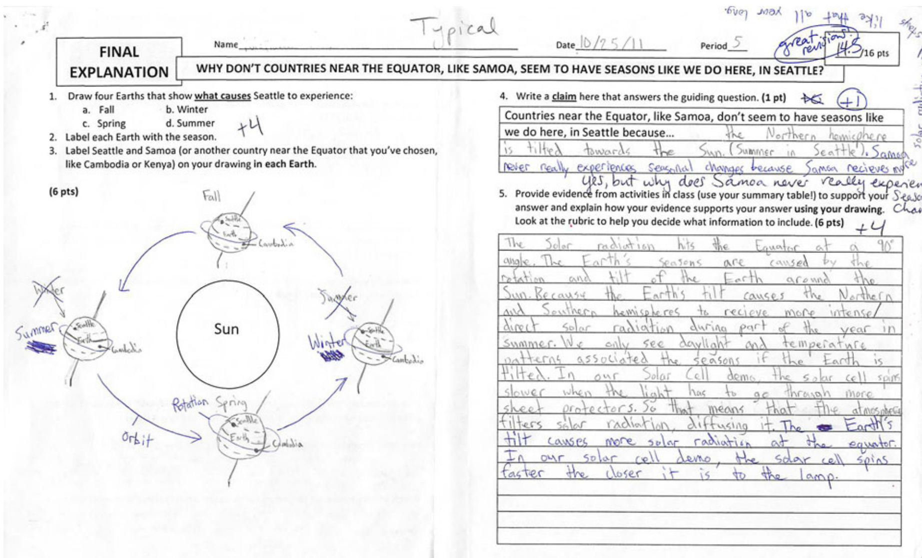


Figure 5. Seasonal change assessment: Using a combination of three high- and one low-level scaffolding in a task.

responses, and how the teacher used this assessment to further support and enhance student learning.

Seasonal Change Assessment: Using a Combination of Three High- and One Low-Level Scaffolding in a Task. This focal assessment was about seasonal changes in a seventh-grade earth science unit. Students were prompted to explain: "Why don't countries near the equator, like Samoa, seem to have seasons like we do here, in Seattle?" As shown in Figure 5, students drew the positions of the earth around the sun in the four seasons and then labeled Seattle and Samoa on their drawing. The teacher gave students options to choose another country near the equator like Cambodia or Kenya. It should be noted that this school has a large population of students of immigrant families from these regions of the world. Next, students were prompted to write a claim in a few sentences guided by the sentence frame: "Countries near the equator, like Samoa, don't seem to have seasons like we do here, in Seattle because . . . " It follows the prompt of "Providing evidence from activities in class (use Summary Table[2]) to support your answer and explain how your evidence supports your answer using your drawing. Look at the rubric to help you decide what information to include." By combining modeling with contextualization in this assessment, students were invited to engage in a high level of intellectual work that involved (a) locating geographic positions for two different countries on the earth, (b) identifying the relative positions of the sun and earth during the orbit of the earth at different seasons, and (c) simulating the seasonal changes at two different locations in terms of exposure to the sun's light. The *focusing sentence frame* seems to help students to get right into the heart of the work, that is, writing a claim about the focal phenomena. In this assessment task, the combination of this high-level scaffolding enabled students not only to draw on their everyday reasoning resources, such as the seasonal differences noted by themselves

[2]A Summary Table is a form of public representation that lists activities (labs, readings, demonstrations) and key ideas addressed with each of the activities, in the form of a table.

and their relatives and travel experiences to visit their relatives, but also to express their ideas in a modality other than writing.

Most students demonstrated significant progresses in their explanation about seasonal changes. For example, one student, Nick, a student as a typical category, initially had a "distance theory" about seasonal changes—"the summer is hot because the earth is closer to the sun, and winter is cold because the earth is far." This was Nick's initial idea about seasonal changes that were elicited on the first day of this unit. Nick produced the following explanation 2 days before the final day of this unit as response to the assessment task:

> [Sentence frame] Countries near the Equator, like Samoa, don't seem to have seasons like we do here, in Seattle because ... [Nick's response] the Northern hemisphere is tilted towards the sun (Summer in Seattle). The solar radiation hits the equator at a 90 angel [sic]. The Earth's seasons are caused by the rotation and tilt of the earth around the sun. Because the Earth's tilt causes the Northern and Southern Hemispheres to receive more intense/direct solar radiation during part of the year in summer. We only see daylight and temperature patterns associated the seasons if the Earth is tilted. In our Solar Cell demo, the solar cell spins slower when the light has to go through more sheet protectors. So that means that the atmosphere filtered solar radiation, diffusing it. (score: 6 of 7, a constructed, evidence-based, in-depth why explanation with partial coherence & reasoning)

The solar cell investigations being referred to were designed to show that not only is the sun's light more concentrated in northern latitudes during the months of June through August (amount of radiant energy per unit area), but that when the sun's light is at a less oblique angle to the earth's surface, it does not have to pass through as many miles of atmosphere before reaching the surface. As shown in Figure 5, the teacher provided feedback to this question in the form of question to press students to further elaborate his idea: "Yes, but why does Samoa never really experience seasonal changes?" Nick revised his model using a different color pen, stating, "Samoa never really experiences seasonal changes because Samoa receives more solar radiation and stays like that all the year long." He also added his evidence about it: "The Earth's tilt causes more solar radiation at the equator. In our solar cell demo, the solar cell spins faster the closer it is to the lamp."

The overall average score of student explanations across 12 samples of student work from this assessment was 4.4 of 7 (63.1%). Of note, the four students who were identified as the category of "learns with difficulty" did as well as the four students in the typical group (the average score of students in easily learn group: 6.3, typical 3.5, and learn with difficulty: 3.5).

DISCUSSION

We make two claims from this study. First, with effective use of scaffolding, teachers create better opportunities for students to demonstrate disciplinary proficiency. Some types of scaffolding, such as using explanation checklists and contextualizing phenomena, clearly define the level of rigor expected of students and appear to make the task more accessible for learners. Second, the *quality and combination* of scaffolding types matters more than the number of scaffolds embedded in the assessments. Quality combinations include a contextualized phenomenon in addition to one or more of the other types of upper-level scaffolds. To unpack these claims, we first focus on how each type of scaffolding supports students' construction of evidence-based explanations. Then, we discuss how and why combinations of high-quality scaffolding are particularly effective. We finish this section with a discussion of the relationships among scaffolding, teachers' instruction, and student learning.

Role of Each Type of Scaffolding in Supporting Evidence-Based Explanation

Contextualized Phenomena: Making the Task Intellectually Challenging But Accessible. We hypothesize that contextualization helps students engage in deeper forms of reasoning and demonstrate in-depth explanations in four ways. First, contextualization problematizes a generic set of conditions. For example, in the assessment about the seasons, contextualizing the generic phenomena of seasonal changes in two different geographic regions (i.e., Samoa and Seattle) generates multiple variables that must be taken into account, such as relative distances of different parts of the earth from the sun (negligible), changes in the angles of the sunlight on the earth, penetration of light through layers of the atmosphere, etc. This contextualization prompts students to recognize the general model of seasonal change and then reason how this model plays out under different conditions. Second, a contextualized phenomenon supports students in moving beyond reproductions of textbook explanations about general phenomena. The fact that there is no authoritative "answer" from a textbook helps a teacher and students reposition themselves as coinvestigators in the process of actually making sense of the phenomena. Third, situating the phenomena in the everyday experiences of students and their families helps them draw in additional intellectual resources from observations and relevant accounts of others (Nordine, Krajcik, & Fortus, 2010; Shwartz, Weizman, Fortus, Krajcik, & Reiser, 2008). For example, in the skater girl assessment (see Figure 3), students used scientific ideas, such as momentum, friction, and acceleration, but also drew upon their everyday reasoning resources, such as "she shouldn't drag her feet/bottom because of the friction from the coarse cement would hurt her" or "When you roll on the floor there is friction so it slows you down if you have a lot of ways." In this way, students drew on prior experiences and observations about slowing down as their bodies interacted with different surfaces and made sense of science ideas simultaneously. Finally, contextualizing a phenomenon in a particular local community helps students relate to the problem, which allows them to become engaged in the work actively and emotionally. Walqui (2006) describes the importance of "bridging" as establishing a personal link between the students and the subject matter by showing the relevance of new materials to the students' life "here and now" (p. 172). The problems that produced richer student explanations—richer meaning longer, more detail, and support for a claim—tended to describe an event/phenomenon that was relevant to the students' everyday life, such as the skater girl's story. In short, contextualized phenomena makes the task more cognitively challenging by problematizing a generic set of conditions and by inviting students to engage in complex reasoning, but at the same time its situatedness in a set of recognizable conditions makes the task more accessible to students.

Providing a Rubric: Priming Disciplinary Ways of Thinking and Talking. Rubrics are frequently used to engage students in monitoring their current level of thinking and determining next steps toward learning goals (Shepard, 2005; Walqui, 2006). In this study, rubric scaffolds alone were significantly associated with the quality of student explanations. Many rubrics explicitly encouraged epistemic features of disciplinary thinking and talking. Even prompts at a Level 1 revealed the tacit structure of evidence-based explanations to students. For example, in the skater girl assessment, each prompt guided students to describe what happened (i.e., a skater girl who is about to hit the wall), use scientific concepts as part of the explanation (i.e., force, friction, momentum) and then to describe the reasoning behind a claim (i.e., explain why this is better than another choice she has). Shepard (2000) reminds us that the importance of transparency is to make the assessment an activity for learning. A rubric that makes the expectations for achievement transparent seemed to support students in successfully constructing evidence-based explanations.

Providing Checklists: Inviting Complex Reasoning With Multiple Relationships. Checklists, in particular an explanation checklist that consisted of statements about important relationships relevant to the focal phenomena, significantly explained the quality of student explanations. We hypothesize that even a simple word checklist reduces the cognitive load for locating terminology, and refocuses students' intellectual resources toward synthesizing information, examining relationships, and evaluating evidence. In contrast, more elaborated forms of checklists, such as explanation checklists, highlight particular relationships or dimensions of process that students need to consider. For example, in the cell membrane assessment (see Figure 2), the explanation checklist helped students attend to key relationships that define a system, such as paramecium's in-take of water in relation to the amount of salt. Also, it invites students to reason about how the relationship fits into a larger activity system (e.g., paramecium's survival). Without explanation checklists, students were able to produce explanations, but with this form of checklist the task challenged students to consider specific dimensions of the phenomena that they may not have otherwise attended to in written explanations.

Drawing in Conjunction With Writing: Prompting to Attend to Relationships and Underlying Mechanisms. The combination of drawing with other forms of scaffolding showed substantially different and positive impacts of two different forms of drawing—namely, generic drawing or posterizing (i.e., Level 1) in contrast with modeling (Level 2)—as predictors. When students engaged in posterizing, they illustrated known facts or information from authoritative sources, typically from textbooks. For example, in the rock cycle assessment task in an earth science unit (see Figure 1), most students produced almost the same drawings that simply illustrated the scientific model of the rock cycle. This kind of drawing may direct students to reproduce the canonical scientific models rather than actively engaging in sensemaking processes. In contrast, when students were prompted to construct models that were products of their own sense making through drawing, they produced diverse types of inscriptions that revealed a wider variety of developing ideas, partial understandings, and different ways of reasoning. For example, in the assessment about buoyancy (see Figure 1), students were prompted to first draw a person floating in salt water, then in fresh water, and to describe all the forces acting on the person, and then to write their explanation of how the density of a fluid affects the buoyant forces. Engaging in this assessment task, students used their drawings to explain the focal phenomenon (i.e., why people float higher in salt water than in fresh water) by highlighting the underlying mechanism (i.e., how the density of a fluid affects buoyant forces in relation to the force of gravity).

We theorize that "drawing as modeling" provides support for student reasoning in the process of making sense of relationships among events, structures, properties, and concepts. Modeling prompts students to identify unobservable events/processes and then connect them causally to patterns of observation (Windschitl, Thompson, & Braaten, 2008). The prompts to illustrate changes either over time or between conditions guide students to express the relationship between unobservable events and changes in the state of the systems across conditions. All of these reasoning tasks are challenging for learners new to reasoning with scientific practices.

Previous studies have suggested that scaffolding different modes of students' expression of ideas, such as drawing, open up opportunities for students who may have difficulty in using scientific language. This is one of the major barriers in science learning (Moje et al., 2001; Rosebery et al., 2010). Our analyses suggest that drawing allows students to show more of what they know, but it needs to be prompted in particular ways to support students' representations of evidence-based explanations.

Sentence Frames: Guiding Disciplinarily Valid Ways of Thinking and Talking. We conjecture that well-designed sentence frames support disciplinary and epistemic reasoning. They also help students express their thinking semantically. "Focusing" sentence frames (i.e., Level 1) usually took the form of "things happened because____" and provided a semantic structure to construct causal explanations about the focal event. Previous studies have shown that students often have difficulty addressing the main question or issue when prompted for explanations (e.g., Ruiz-Primo et al., 2010). By providing a "focusing" sentence frame, students are guided into the problem space. Furthermore, more sophisticated sentence frames, specifically "connecting" sentence frames, provided additional support for using epistemic structures for writing scientific explanations. As demonstrated in the cell membrane assessment (Figure 2), sentence frames guided students to construct their explanation while examining the relationship between evidence and reasoning. Providing clear examples of appropriate language use for the performance of specific academic functions, such as the use of evidence, is a critical form of scaffolding (Walqui, 2006). It seems that well-designed sentence frames have potential to support the epistemic and semantic challenges of constructing evidence-based explanations. In our study, however, teachers used sentence frames infrequently ($n = 10$, 13.1%), and upper level sentence frame were rarely used. More evidence is needed to understand the roles of different forms of sentence frames in supporting students' productive disciplinary participation.

Quality and Combinations of Scaffolding Matter

The results of hierarchical multiple regression analyses show no linear relationship between the number of scaffolding types used in assessments and the quality of student explanations. In other words, simply adding more forms of scaffolding did not increase the quality of student explanations. However, close examination of both quantitative and qualitative analyses suggests that when several forms of *higher level scaffolding types* were used in *strategic combinations*, it creates opportunities for more students across different academic achievement levels to show a spectrum of ideas and reasoning through written explanation.

Constructing evidence-based explanations is a highly complex task that poses multidimensional challenges in understanding the task itself, planning a response, and producing representations of one's thinking. Different learners encounter different kinds of challenges in the process of constructing evidence-based explanations. Students, for example, vary in their reading and writing proficiencies as well as their self-perceptions as science learners. As discussed in the preceding section, each form of scaffolding serves its own unique function in supporting the work of constructing evidence-based explanations while addressing different challenges for such performances. Therefore, a combination of several types of high-level scaffolding can create opportunities for more students to succeed in engaging in constructing evidence-based explanations.

Scaffolding, Teachers' Classroom Instruction, and Supporting Deeper Learning

The assessment tasks in this study were designed to be used as a part of teaching. We noticed that the most sophisticated forms of assessment were developed by the teachers who were known to enact similarly sophisticated classroom practices. We conjecture that the teachers who created quality opportunities for students to demonstrate their ideas, thinking, and ways of reasoning with the effective use of scaffolding come to be in a better position to modify their teaching practices because the scaffold-rich assessments produce highly descriptive information about students' strengths and difficulties. This study focused on one critical part of assessment design—providing opportunities for students to show what they are capable of.

Our findings implicate the intertwined relationship between the performative aspects of teaching practices and the use of scaffolding. Students' written explanation produced from their engagement in assessment tasks is an outcome of activity that is always situated in larger instructional contexts. It is reasonable to ask to what degree the increased quality of student explanations is the effect of the embedded scaffolding within assessments. From hierarchical multiple regression analyses, the teacher—as a variable—explained the greatest amount of variance in student performance (about 38%), followed by use of scaffolding (about 11%) and students' academic backgrounds (about 9%).

The findings strongly indicate a clear relationship between students' explanatory performance and the use of scaffolding. In our data set of 707 copies of student work, the overall score for quality of student explanation was 2.1 of 7 (30.1%). This result indicates that most students are currently failing to meet the expectations grounded in ambitious visions of twenty-first century learning. In a recent study that examined the quality of students' written explanations from science inquiry-based middle school classrooms in five states, Ruiz-Primo et al. (2010) also found that a low percentage of students (18%) provided explanations meeting the criteria for evidence-based explanation. A close examination of teachers' use of scaffolding and student performance in the present study suggests that the lower level of student performance has to do with teachers' less frequent use of scaffolding. In this study, about one quarter of assessment tasks did not provide any form of scaffolding (15 of 76 assessment tasks, 19.7%). In those assessments, the average quality of student explanation was about half ($n = 121$, average score = 1.12 of 7; *SD*: 1.10) compared to student responses from those with *any form* of scaffolding ($n = 586$, average score = 2.32 of 7; $SD = 1.94$). Even when the assessment tasks provided some scaffolding, those were generally one or two types of scaffolding with *lower levels* of sophistication.

Most assessment tasks did not scaffold students' substantial use of evidence for scientific explanations. It should be noted that new forms of learning that we are envisioning, such as constructing evidence-based explanations, involve students' engagement in particular ways of thinking and talking that are unfamiliar to most students. Ruiz-Primo and her colleagues (2010) noted that

> [w]hat students are missing and lacking is learning experiences and guidance in the fundamental activities of constructing explanations. Unless or until such experiences and guidance are adequately offered, it is not surprising to find that constructing explanations is challenging for students. (p. 605)

It would be unrealistic for our students to meet the increased expectations without providing appropriate and sufficient support. Put another way: Designing for rigor in curriculum in student learning experiences or in assessments may not necessarily lead to higher student performances. A narrow focus on rigor may in fact obscure what students are capable of, if given strategic (and temporary) forms of scaffolding.

This leads to the issue of fading. The question of fading supports while simultaneously maintaining necessary support of all students' efforts at science reasoning and practice remains understudied. The big concern is "when" and "under which conditions" will students be able to construct evidence-based explanations without scaffolds? We hypothesize that the absence of scaffolds of any kind will be most challenging when students are presented with a novel phenomenon that is neither directly relevant to students' everyday lives nor locally contextualized. We also believe that, across units of instruction, some scaffolds may gradually be removed as classroom communities become more practiced in talking and writing about models and explanations. In terms of further research, it will be fruitful to take a situative perspective (Greeno, 2006) to understand the varied ways students learn to appropriate the habits of thinking behind each form of scaffolding and what kinds of learning environments are most influential for students to take up particular kinds of writing and discourse. We recognize that there cannot be a one-size-fits-all approach to fading. But we may be better able to understand the process of fading by showing how scaffolds for varied learners are released over time and observing these effects in the contexts of the development of social and scientific communities in K-12 classrooms.

CONCLUSION AND IMPLICATIONS

We conclude that providing effective scaffolding is necessary, not optional, when trying to support students in meeting twenty-first century standards (NRC, 2012a; Resnick, 2010). For teachers and science teacher educators who are interested in designing assessment tasks for supporting and enhancing student learning, we recommend using a combination of two or more high-quality types of scaffolding including the use of contextualized phenomena. This use of contextualized phenomena has implications, of course, for instruction as well as assessment. The use of "anchoring" phenomena for units of instruction has been strongly suggested in the Framework document for the Next Generation Science Standards (NRC, 2013).

With the new reform emphases on science as practice, we see an emerging challenge for teachers—that of scaffolding both instruction around these practices and the assessment of students' abilities to take up the conceptual, social, and epistemic dimensions of disciplinary work. We believe that assessment and instruction must be interwoven more than they currently are and not treated as separate events. If classroom teachers can "see" more of what their students are capable of through well-designed assessments, there will be new opportunities to study how they use these rich forms of data to make instructional decisions that attend to the needs of different groups of learners. With scaffolding, the quality of information to base decisions on is much greater than without. We see this as a lever to promote both rigor and equity in classroom teaching.

We wish to thank the teachers who provided their student work and were willing to share their stories with us. We also thank to the research group at the University of Washington for their help in developing this paper. This article benefitted from the critical feedback of three anonymous reviewers and the editor of the journal.

REFERENCES

Braaten, M., & Windschitl, M. (2011). Working toward a stronger conceptualization of scientific explanation for science education. Science Education, 95(4), 639–669.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). How people learn: Brain, mind, experience, and school. Washington, DC: National Academies Press.

Brown, A. L., & Campione, J. C. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), Classroom lessons: Integrating cognitive theory and classroom practice. Cambridge, MA: MIT Press/Bradford Books.

Bruner, J. (1985). Narrative and paradigmatic modes of thought. In E. Eisner (Ed.), Learning and teaching: The ways of knowing. Chicago: National Society for the Study of Education.

Creswell, J. W., & Plano Clark, V. L. (2011). Designing and conducting mixed methods research (2nd ed.). Thousand Oaks, CA: Sage.

Ford, M. J., & Wargo, B. M. (2012). Dialogic framing of scientific content for conceptual and epistemic understanding. Science Education, 96(3), 369–391.

Furtak, E. M., & Ruiz-Primo, M. A. (2008). Making students' thinking explicit in writing and discussion: An analysis of formative assessment prompts. Science Education, 92(5), 799–824.

Greeno, J. G. (2006). Learning in activity. In K. R. Sawyer (Ed.), The Cambridge handbook of the learning sciences (pp. 79–96). New York: Cambridge University Press.

Herman, J. L. (1992). What research tells us about good assessment. Educational Leadership, 49(8), 74–78.

Kenyon, L., & Reiser, B. J. (2006). A functional approach to nature of science: Using epistemological understandings to construct and evaluate explanation. Paper presented at the AERA annual meeting, San Francisco, CA.

Knorr-Cetina, K. (1999). Epistemic cultures: How sciences make knowledge. Cambridge, MA: Harvard University Press.

Latour, B., & Woolgar, S. (1979). Laboratory life: The social construction of scientific facts. Los Angeles: Sage.

McNeil, K. L., & Krajcik, J. (2006). Supporting students' constructions of scientific explanation through generic versus context-specific written scaffolds. Paper presented at the AERA Annual Meeting, San Francisco, CA.

McNeil, K. L., & Krajcik, J. (2008). Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. Journal of Research in Science Teaching, 45(1), 53–78.

Moje, E. B., Ciechanowski, K. M., Kramer, K., Ellis, L., Carrillo, R., & Collazo, T. (2004). Working toward third space in content area literacy: An examination of everyday funds of knowledge and discourse. Reading Research Quarterly, 39, 38–72.

Moje, E. B., Collazo, T., Carrillo, R., & Marx, R. W. (2001). "Mastro, what is 'quality'"?: Language, literacy, and discourse in project-based science. Journal of Research in Science Teaching, 38(4), 469–498.

Nasir, N. S., Rosebery, A. S., Warren, B., & Lee, C. (2006). Learning as a cultural process: Achieving equity through diversity. In K. R. Sawyer (Ed.), The Cambridge handbook of the learning sciences (pp. 489–504). New York: Cambridge University Press.

National Research Council. (2000). Inquiry and the National Science Education Standards. Washington, DC: National Academy Press.

National Research Council. (2007). Taking science to school: Learning and teaching science in grade K-8. Washington, DC: National Academy Press.

National Research Council. (2012a). Education for life and work: Developing transferable knowledge and skills in the 21st century. Washington, DC: The National Academies Press.

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

National Research Council. (2013). Next generation science standards: For states, by states. Washington, DC: The National Academies Press.

Nersessian, N. (2005). Interpreting scientific and engineering practices: Integrating the cognitive, social, and cultural dimensions. In M. Gorman, R. D. Tweney, D. Gooding, & A. Kincannon (Eds.), Scientific and technological thinking (pp. 17–56). Hillsdale, NJ: Erlbaum.

Nordine, J., Krajcik, J., & Fortus, D. (2010). Transforming energy instruction in middle school to support integrated understanding and future learning. Science Education, 95(4), 670–699.

Ohlsson, S. (2002). Generating and understanding qualitative explanations. In J. Otero, A. Leon, & A. C. Graesser (Eds.), The psychology of science text comprehension. Mahwah, NJ: Erlbaum.

Osborne, J., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? Science Education, 95(4), 627–638.

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. Journal of the Learning Sciences, 13(3), 423–451.

Pickering, A. (1995). The mangle of practice: Time, agency, and science. Chicago: The University of Chicago Press.

Resnick, L. B. (2010). Nested learning systems for the thinking curriculum. Educational Researcher, 39(3), 183–197.

Rosebery, A. S., Ogonowski, M., DiSchino, M., & Warren, B. (2010). "The coat traps all your body heat": Heterogeneity as fundamental to learning. Journal of the Learning Sciences, 19(3), 322 – 357.

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

Ruiz-Primo, M. A., Li, M., Tsai, S.-P., & Schneider, J. (2010). Testing one premise of scientific inquiry in science classrooms: Examining students' scientific explanations and student learning. Journal of Research in Science Teaching, 47(5), 583 – 608.

Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. Journal of the Learning Sciences, 12(1), 5 – 51.

Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. Cognition and Instruction, 23(1), 23 – 55.

Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. Science Education, 88(3), 345 – 372.

Sawyer, K. R. (2006). The new science of learning. In K. R. Sawyer (Ed.), The Cambridge handbook of the learning sciences (pp. 1 – 16). New York: Cambridge University Press.

Shepard, L. A. (2000). The role of assessment in a learning culture. Educational Researcher, 29(7), 4 – 14.

Shepard, L. A. (2005). Linking formative assessment to scaffolding. Educational Leadership, 63(3), 67 – 70.

Shwartz, Y., Weizman, A., Fortus, D., Krajcik, J., & Reiser, B. (2008). The IQWST experience: Using Coherence as a Design Principle for a Middle School Science Curriculum. The Elementary School Journal, 109(2), 199 – 219.

Smith, C. L., Maclin, D., Houghton, C., & Hennessey, M. G. (2000). Sixth-grade students' epistemologies of science: The impact of school science experiences on epistemological development. Cognition and Instruction, 18(3), 349 – 422.

Songer, N. B., & Gotwals, A. W. (2012). Guiding explanation construction by children at the entry points of learning progressions. Journal of Research in Science Teaching, 49(2), 141 – 165.

Supovitz, J. (2012). Getting at student understanding—The key to teachers' use of test data. Teachers College Record, 114, 1 – 29.

Thompson, J., Braaten, M., Windschitl, M., Sjoberg, B., Jones, M., & Martinez, K. (2009). Collaborative inquiry into students' evidence-based explanations: How groups of science teachers can improve teaching and learning. The Science Teacher, November, 48 – 52.

Toulmin, S. (1958). The uses of arguments. Cambridge, England: Cambridge University Press.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Cambridge, MA: Harvard University Press.

Walqui, A. (2006). Scaffolding instruction for English language learners: A conceptual framework. The International Journal of Bilingual Education and Bilingualism, 9(2), 159 – 180.

Warren, B., Ballenger, C., Ogonowski, M., Rosebery, A. S., & Hudicourt-Barnes, J. (2001). Rethinking diversity in learning science: The logic of everyday sense-making. Journal of Research in Science Teaching, 38(5), 529 – 552.

Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. Science Education, 92(5), 941 – 967.

Windschitl, M., Thompson, J., & Braaten, M. (2011). Ambitious pedagogy by novice teachers: Who benefits from tool-supported collaborative inquiry into practice and why? Teachers College Record, 113(7), 1311 – 1360.

Windschitl, M., Thompson, J., Braaten, M., & Stroupe, D. (2012). Proposing a core set of instructional practices and tools for teachers of science. Science Education, 96(5), 878 – 903.