Title

Designing, launching, and implementing high quality learning opportunities for students that advance scientific thinking

Permalink

https://escholarship.org/uc/item/5qw3675d

Journal

Journal of Research in Science Teaching, 53(9)

ISSN

0022-4308

Authors

Kang, Hosun
Windschitl, Mark
Stroupe, David
et al.

Publication Date

2016-11-01

DOI

10.1002/tea.21329

Designing, Launching and Implementing High Quality Learning Opportunities for Students that Advance Scientific Thinking

This manuscript was accepted by Journal of Research in Science Teaching on April 13, 2016

Author Information

[1]Hosun Kang (corresponding author): University of California—Irvine; 3200 Education, Irvine, CA 92697-5500; Hosunk@uci.edu; fax: (949) 824-2965; Phone (949) 824-6914

[2]Mark Windschitl: University of Washington, Seattle; 115 B Millder, Box 353600 Seattle, WA 98195-3600, USA; mwind@u.washington.edu; Phone (206) 221-4736

[3]David Stroupe: Michigan State University; Erickson Hall, 620 Farm Lane, East Lansing, MI 48824; dstroupe@msu.edu; fax (517) 432-5092; Phone (517) 353-0664

[4]Jessica Thompson: University of Washington, Seattle; 122M Miller, Box 353600 Seattle, WA 98195-3600, USA; jjthomps@u.washington.edu; Phone (206) 221-4630

Disclaimer

This material is based upon work supported by the National Science Foundation under Grant No. 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.

Designing, Launching and Implementing High Quality Learning Opportunities for Students that Advance Scientific Thinking

Abstract

Instructional tasks are key features of classroom practice, but little is known about how different components of tasks—such as selecting or designing tasks for a lesson, launching, and implementing them with students—shape the conditions for students' intellectual engagement in science classrooms. Employing a qualitative multiple case study approach, we analyzed 57 science lessons taught by 19 first-year teachers. We examined the potential for students' intellectual work built into the tasks across the phases of instruction, and how the demand of the unfolding task deepened (or failed to deepen) students' engagement in science. The findings suggest the importance of beginning a lesson with high quality instructional tasks—complex tasks that bear appropriate levels of epistemic uncertainty for a particular group of students in a particular moment. Beginning a lesson with high quality tasks however was insufficient by itself to ensure rigorous learning opportunities. With the use of complex tasks, higher quality opportunities to learn were observed in lessons in which (a) the tasks were framed as a process of understanding contextualized phenomena, (b) the specific disciplinary concepts in the task were related to big science ideas that transcended the activities themselves, and (c) students' implementation of these tasks were structured using tools that supported changes in thinking.

Key words: opportunities to learn, instructional tasks, intellectual demand, instructional practices

Introduction

Despite the use of various hands-on, practical, and experiential activities in science instruction, research indicates that classroom learning for students largely remains undemanding, procedural, and often disconnected from the development of substantive science ideas (Banilower et al., 2012; Consortium for Policy Research in Education, 2011; Roth & Garnier, 2007; Roth et al., 2011). For example, in a study using a nationally representative sample of 31 middle schools in the U.S., Weiss et al. (2013) found that few science teachers help students link activity to substantive science ideas. Weiss et al. also found that appropriate sense-making opportunities for students coupled with activities were rare (about 16 percent of observations), and classroom discourse that could support reasoning by students was one of the weakest dimensions of practice. Roth and Garnier (2007) analyzed a random sample of 500 8th grade classroom videos across five countries and concluded, "U.S. teachers did not typically use the various activities to support the development of content ideas in ways that were coherent and challenging for students" (p. 20). This trend of "activity without understanding" is also reported by other studies outside the U.S. (e.g. Goodrum, Hackling, & Rennie, 2000; Osborne & Dillon, 2008).

Science educators struggle to help students make sense of science activities, connect observation and everyday experiences with science ideas, and deepen students' understanding about the natural world. It is unclear, however, whether this problem originates in the design of the activity, how the activity is framed by the teacher, how the students enact the work, or how the teacher supports sense-making in subsequent discussions. Researchers tend to foreground either the curriculum or teachers' practices to understand this problem. There is a significant body of literature that reports on expert-developed curriculum and its impact on student achievement (e.g., Harris et al, 2015; Lynch, Pyke, & Grafton, 2012), on teachers' implementation of reform-oriented curricula (e.g., Fogleman, McNeill, & Krajcik, 2011; Forbes, 2011; McNeil & Krajcik, 2008; Schneider, Krajcik, & Blumenfeld, 2005; Zangori, Forbes, & Biggers, 2013), and on effective instructional practices such as the facilitation of classroom discussion (e.g., McNeill & Pimental, 2010; Puntambekar, Stylianou, & Goldstein, 2007). Some studies examine multiple characteristics of teaching that are linked with enhanced students' performances. For example, McNeill, Pimentel and Strauss (2013) found that when implementing an inquiry-oriented curriculum, significant learning gains were associated with students spending a larger percentage of their time on group work, the sharing of ideas within groups, and students' engagement in argumentation. Despite the insights provided by this body of scholarship, it is still unclear *how* both features of the curriculum and practices by the teacher play out *together* to sustain rigorous learning opportunities over the arc of a lesson trajectory in classrooms.

We contend that students' opportunities to learn depend upon the integrity of a planning-implementation process, in which thoughtfully designed lessons are framed for students as intellectually challenging and are enacted in ways that maintain high rigor. While such studies are rare in science education, parallel studies in the field of mathematics education offer a useful analytical lens. For example, Stein and colleagues have conceptualized the process in which student learning opportunities are created by attending to the mediating role of instructional tasks (Stein et al., 1996; Stein & Lane, 1996). In their mathematical tasks framework, instruction begins with selecting or modifying tasks from curriculum materials during the planning stage (referred to as "tasks appearing in curriculum materials"). Two stages of enactment follow: teachers setting up the tasks at the beginning of the lesson ("tasks-as-set up") and then teachers and students implementing the tasks (“tasks-as-implemented”). Stein et al. traced the changes in the cognitive demand of 144 mathematical tasks used during reform-oriented middle grades mathematics instruction from planning through implementation (Stein et al., 1996; Stein & Lane, 1996) and found that the rigor of the tasks tended to be lowered from planning to implementation stages, depending on how the tasks were set up by the teacher and implemented during the enactment. Student performance gains were greater in instances where tasks were *both* set up *and* implemented in a way that maintained high cognitive demand.

Drawing upon Stein et al.’s mathematical tasks framework, Jackson and colleagues (Jackson, Garrison, Wilson, Gibbons, & Shahan, 2013) examined how teachers, required to implement reform-oriented curriculum materials, set up tasks at the beginning of their lessons (what they called the “launch of the tasks”). Jackson et al. found that in approximately 64% of analyzed lessons, the cognitive demand of the tasks was lowered during this setup phase of instruction. Based on the analyses of 165 middle grades mathematics lessons taught by experienced teachers, they specified two features of productive launches that were significantly associated with high quality learning opportunities observed during the concluding discussions: (a) maintaining cognitive demand during the launch, and (b) developing common language with students to describe task-relevant contextual features and mathematical relationships.

These studies in the field of mathematics education provide two important insights that may inform how learning opportunities are created in science classrooms. First, the studies point to the critical role of instructional tasks as a mediator for students’ opportunities to learn. Stein and Lane (1996) theorize that instructional tasks influence learning by: (a) affecting the way students *think* as opposed to influencing learning directly, (b) configuring the kind and level of classroom discourse that is possible, (c) configuring the possibilities for the occurrence of other forms of recommended instructional practices (p.55). Second, these studies suggest that the kinds of decisions that teachers make in order to support deeper learning require them to go *beyond what is prescribed in curriculum materials*.

Building upon the prior research in mathematics education, this study explores how different components of science tasks, such as selecting or designing tasks for a lesson, introducing, and implementing them with students, come to shape the conditions for students' intellectual engagement in science classrooms. We think that it is critically important to look at both tasks-as-planned and tasks-as-enacted *together* because their influence on students' learning opportunities is inter-related. Examining either the quality of instructional tasks *or* how teachers enact the tasks limits our understanding of how we can create high quality learning opportunities in classrooms. As numerous curriculum studies have demonstrated, high quality tasks themselves do not ensure high quality learning opportunities for students in classrooms (Harris et al, 2015; Lynch, Pyke & Grafton, 2012; Penuel, Gallagher, & Moorthy, 2011). Examining only teachers' enactment of tasks is also a narrow analytical lens because the nature of learning opportunities is likely influenced by the initial quality of the tasks selected/modified/designed by teachers during planning. We hypothesize that it is a *sequence of coordinated events* that constitute opportunities to learn in the classroom. Therefore, as researchers, we intend to characterize the conditions that *lead* up to students' opportunities to learn. In other words, we aim to understand how planned and enacted tasks "set the stage" for student learning.

In this study we examine what science teachers decide to do with curriculum tasks (selected or adapted), how they launch the tasks, and what happens once the teacher implements the tasks with students. Our analysis focuses on the built-in intellectual demand of instructional tasks in the phases of designing, launching, and implementing, and how it is related to students' intellectual engagement observed in 57 science lessons. The following questions guide our analysis:

1. What is the potential for intellectual demand built into common instructional tasks at the onset of a science lesson (task as designed)?
2. How does the intellectual demand shift as the tasks are launched, then implemented during a lesson (task as launched and implemented)?
3. How does the quality of instructional tasks as designed, launched and implemented, relate to student learning opportunities?

In this article, we specify the conditions under which students come to engage in meaningful intellectual work and the kinds of pedagogical decisions entailed for creating those conditions. This study is significant because the findings illuminate how intellectually challenging learning opportunities in science classrooms are developed and sustained across the duration of a lesson. In the following section, we first discuss how we conceptualize opportunities to learn and the nature of instructional tasks in science classrooms—the focus of this investigation. Next, we present the science tasks framework that represents the sequence of instructional events from planning, launching, and enactment of instructional tasks. This framework guides our research design and analyses.

Situative Perspective, Opportunities to Learn, and Instructional Tasks

We draw upon a situative perspective (Greeno, 2006; Greeno & Gresalfi, 2008) to study the relationship between instructional tasks and student learning opportunities. Rather than focusing on the processes of acquiring new mental structures, a situative perspective attends to a process in which individuals participate more proficiently in practices through social and intellectual interactions within *an activity system*. From this perspective, a learner's participation is dependent upon features of the settings (i.e., elements of an activity system), such as the nature of tasks, resources and tools available, and facilitation by a more knowledgeable other (Greeno, 2006; Sohmer, Michaels, O'Connor, & Resnick, 2009). *Opportunities to learn (OTL)* refer to the affordances of a setting for changing learners' participation and practices (Greeno & Gresalfi, 2008). High quality learning opportunities created in science classrooms make resources available that facilitate students' more proficient participation in the practice and language (i.e., tool) of science over time.

Instructional tasks are one key feature of a setting that can influence on students' opportunities to learn in science classrooms. Expanding upon the scholarship on instructional tasks in the field of mathematics education (Doyle, 1983; Hiebert & Wearne, 1993; Stein et al., 1996; Stein & Lane, 1996), we conceptualize an instructional task as a form of work assigned to students (e.g., an assignment, an activity) that is defined by the teacher for the purpose of developing students' understandings of concepts, skills, and scientific practices. There are two key features that can be useful in recognizing and characterizing instructional tasks in science instruction. First, an instructional task typically results in some documentation of the outcomes of intellectual work (i.e. the main products that students are expected to produce). Second, an instructional task includes procedures and resources that students are expected to use to complete the work (i.e., specifications for how are students expected to produce outcomes and with what resources) (Doyle & Carter, 1984; Stein et al., 1996). The nature of the task students are asked to do, although prompted by the curriculum materials, can vary depending on the ways in which a teacher modifies procedures in response to students' needs and their understanding of the discipline. For example, the three physics teachers in this study coincidently used the same expert-developed curriculum materials to teach the topic of forces in springs. The instructional tasks of each of the three observed lessons were: (a) students as a group of 2-4 produced a graphical representation of collected data on a small whiteboard that showed the relationship between force and the stretched length of two springs, (b) students as a small group generated a data table to produce evidence for arguing which of the springs with different stiffness would work better for an exercise machine, (c) students individually solved problems on the provided worksheet using formulas after confirming Hooke's law with data.

One feature of instructional tasks in science that relates to intellectual demand is the way in which observable and unobservable (theoretical) aspects of natural phenomena are addressed in relation to students' sense-making. Tasks that involve collecting and evaluating data, or identifying evidence, serve the implicit purpose of drawing students' attention to observable elements of natural phenomena which is an important first step to make sense of how and why natural phenomena occur (Kang, Thompson, & Windschitl, 2014; Osborne & Patterson, 2011; Windschitl, Thompson, & Braaten, 2008). Other kinds of tasks prompt students to use such observations to theorize about abstract, microscopic, or otherwise inaccessible events and processes that influence the world we see. For example, in a physics class, a task can prompt students to use the idea of unbalanced forces to account for why a roller coaster car moves the fastest at the lowest point in the track. The process of scientific sense-making involves developing multiple and coherent relationships between/among observable and unobservable elements in a system (Ohlsson, 2002; Thagard, 2008) through discursive interactions with others and in the context of disciplinary practices such as argument and explanation. Students' interaction in these situations is significantly mediated by the nature of instructional tasks (Greeno & Gresalfi, 2008).

Science Tasks Framework: Linking Instruction to Student Learning Opportunities

Building upon the scholarship on instructional tasks in mathematics (Stein et al., 1996), we propose the science tasks framework (Figure 1) that guides the systematic examination of how teachers' planning and enactment of instructional tasks are related to student learning opportunities. It should be noted that even though the proposed framework presents planning and implementing instructional tasks and students' learning opportunities around single lessons in a linear manner, teachers often reason about these features in relation to one another both during the design phase and during the act of implementation. Pedagogical decisions at multiple levels accumulate over time and have consequences for learning (see Tiberghien, Cross, & Sensevy, 2014). The following section conceptualizes the phases of instructional tasks as designed, launched, and implemented, and its connection to students' learning opportunities.

--Figure 1 about here--

Tasks as designed. As instructional designers, teachers have access to various teaching resources. By *resources* for instruction, we mean any assets that are used for designing learning opportunities in classrooms. They include conventional or material resources (e.g., curriculum, textbook, laptop, and equipment), human resources (e.g., knowledge, teaching strategies), and social resources (e.g., relationships with knowledgeable others who can assist in the design of instruction). Material resources can either be modified or left unchanged from standard curricula. By *task as designed*, we mean the status of the selected or designed activity at the onset of a lesson before being communicated to students. It is the outcome of teachers' work in the "design arena" (Remillard, 1999)—teachers' activities to appropriate or invent particular tasks for students.

Task as launched. Teachers introduce and communicate the planned instructional task to students at certain point during instruction. Similar to mathematics education, we refer to this phase as the "*launch of task*" (Jackson et al., 2013). Drawing upon the research on framing (Hammer, Elby, Scherr, & Redish, 2005; Hand, Penuel, & Gutierrez, 2012; van de Sande & Greeno, 2012), we posit that the launch is the time when a teacher frames a science situation for students, specific to the task, through discursive interactions. During the launch, specific features of the task are represented in a particular way, which gives rise to the meaning of a situation for participants interacting within it (Greeno, 2009; Hand et al., 2012). Teachers may draw students' attention to certain aspects of the tasks, project the goals that the teacher are designated to address the task, name and draw upon resources to organize the activity, delineate the rules for "knowing" that are needed to succeed in their activity, and assign the roles and expectations for participants. Frames, however, are not simply "given" for students (Hand et al., 2012; see Maskiewicz & Winters, 2012); instead frames are negotiated and co-constructed between a teacher and students, each of whom comes in with their own framing (e.g., what it means to do activities in science classroom), and find the "fit" between existing and new framing to figure out "what is it that is going on here?" (Goffman, 1974). Through this process the tasks come to be collectively *framed* as figuring things out or perhaps completing the work for credits. Research on framing in classrooms problematizes a prevalent key school frame, that of "doing school" or "completing the worksheet" frame (Hammer et al., 2005; Hand et al., 2012). This frame represents the structuring of and engagement in rote and shallow learning performances (Jimenez-Aleixandre, Bugallo Rodriguez, & Duschl, 2000). Given that science classrooms are likely populated with students (and teachers) coming in with varied framings, including the students who assume a "doing school" as the default for participation from past experiences, it is important to attend to how tasks are framed as meaningful work (or not) during the launch, and how the ways in which tasks are framed are related to students' later intellectual engagement.

Task as implemented. The final part of the enactment is implementing the tasks with students. The phase of "*Implementation of tasks*" refers to the time when students engage in tasks with the support of the teacher. From an instructional standpoint of view, a key of this phase is how students' social and intellectual interactions, that are framed and set up through the launch of tasks, are supported to achieve instructional goals. It is largely unknown what forms of instructional support should be provided for whom, when and in what ways, in order to assist students' intellectual engagement mediated by high quality instructional tasks in science classrooms.

In this study, we attend to one kind of scaffold that teachers can provide—the use of instructional tools. We attend to tools because they can provide students with scaffolding to engage in challenging tasks. Stein and her colleagues (1996) found that in 58% of lessons, the maintenance of high-level cognitive activity was associated with provision of scaffolding. By instructional tools, we refer to material artifacts provided for students to mediate social and intellectual interaction during task implementation. Instructional tools are designed to achieve particular goals by improving conditions of student work, especially when they engage in a complex activity. Therefore, tools embody some of the designer's knowledge about the object of work (i.e., goals) and effective ways to achieve the goals (i.e., knowledge-embedded function). Different from curriculum materials such as a worksheet, tools can often be used across different science topics, units, and even across grade levels. One prevalent tool in current science instruction is a "Claim-Evidence-Reasoning" framework (McNeill & Krajcik, 2006; McNeill & Krajcik, 2008). This tool, that incorporates knowledge about the structure of argumentation, is designed to assists students' engagement in core scientific practices. This study focuses on how the use of instructional tools, if there is any, affects the quality of students' interactions (i.e., students' talk and performance) during the implementation of tasks.

Methods

Research Contexts and Participants

This study is part of a larger project that explored beginning science teachers' learning trajectories from preparation to first year of teaching. In this particular investigation, we focus on the quality of instructional tasks and student engagement in science classroom. The data came from 19 first year science teachers who received their teaching certificates from 13 different college-recommending teacher preparation programs in the Northwest United States. Even though all lessons were observed in 1st year teachers' classrooms, the quality of the instruction observed across the lessons varied widely, which allowed us to draw reasonable inferences about the relationship between the quality of instructional tasks and opportunities to learn in classrooms in this sample. All participants had BA in science or equivalent degree and certified to teach science at the secondary level (6th to 12th grades). The participants' backgrounds and their school contexts are representative of various populations of new science teachers in the United States (see the participants' profiles and school contexts in Table S1).

Data Sources

Observations. Participating teachers' instruction was observed at least three times from three different units upon their invitation between December 2011 and June 2012. Our preference was to observe the most discourse rich lessons in a unit selected by teachers. We assumed that those lessons likely provided best opportunities for us to examine how the different quality of instructional tasks afforded or constrained opportunities for students to advance their thinking by mediating classroom talk (Stein & Lane, 1996). Selecting the three lessons from three different units helped us reflect the various topics in the sciences and reduce the likelihood of the observational bias caused by teachers' differential content knowledge. We specifically requested that participating teachers invite us to observe lessons in which students were engaged in a "sense-making conversation," such as 'having a discussion after finishing lab activities' toward the end of unit or project. In most cases teachers invited us when they did a lab, project, or hands-on activity. When an observed lesson did not include any sense-making conversation, we made an additional observation on the following day with the teacher's permission to document any follow-up conversation coupled with previous activity. Ten out of 19 teachers were visited more than three times. A total of 83 lessons were observed. For the final analysis, we selected three most discourse rich lessons per each teacher. Therefore a total of 57 lessons were coded. Each visit generated a set of data that consisted of detailed field notes, audio-recorded classroom dialogues, a 30 to 50 minute long post-observation interview with a teacher, 12 samples of produced student work, various teaching artifacts, and photographs.

Interviews. Post-observation interviews were conducted for 30-50 minutes. The teachers debriefed their instruction in response to the observer's questions. We specifically asked about the context of the observed lesson, the instructional tasks, including the curricular sources, the intention behind the task design, their perception about the success and challenges in the ways the tasks unfolded during instruction. Some examples of questions were: Since I am not here for the entire inquiry, could you tell me about the lessons that came before this lesson and the lessons that will follow it? For this particular lesson, where did you get the idea for this lesson? What kinds of resources were available for you as you planned this lesson? (see the Interview Protocol S1).

Teaching and learning artifacts. We collected or took photographs of any inscribed artifacts that showed the instructional tasks as designed, launched or implemented. They included lesson plans, slides, task guidance sheets, student worksheet, rubrics, or any instructional tools used to assist students' engagement during task implementation. In addition, we collected learning artifacts produced by students either individually or collectively. Ten to 12 samples of produced student work were collected per lesson.

Data analysis

This study takes a qualitative multiple case study approach (Merriam, 2009) to examine the quality of instructional tasks and student engagement in 57 science lessons. The first phase of the data analyses focused on developing a conceptual framework and coding scheme. Both were initially informed by the literature and later by emergent themes from the data. Specifically, the first author presented the initial coding schemes and framework to six members of the research team. The coding scheme was debated and iteratively revised in weekly research meetings until the group reached a consensus. The first and second author coded a sample of data using the final coding scheme (Table S2), identified areas of disagreement and refined the coding process over time. Based on the results of coding, cross-case analyses were conducted to theorize how and under which conditions students were provided high quality learning opportunities that advanced their scientific thinking. The following describes the process in detail, including data sources and coding scheme for the quality of tasks at each stage.

Phase I: Coding the intellectual demand of instructional tasks as designed, launched, and implemented.

Tasks as designed. We first examined each lesson holistically to identify instructional tasks by determining: (a) what students are asked to produce, and (b) how and with what resources. Then tasks as *designed* are identified by cross referencing lessons or unit plans, instructional sheets for the day's activity, student worksheets, or slides prepared for the instruction from planning. Two sets of sub-codes emerged with respect to high and low cognitive tasks through open and analytical coding (see Table S2). Building upon the prior studies on instructional tasks (Kang & Anderson, 2015; Stein et al., 1996; Stein & Lane, 1996; Tekkumru-Kisa, Stein, & Schunn, 2015), cognitive science (Mayer, 2005), students' scientific reasoning (Ohlsson, 2002; Thagard, 2008) and informed by recent influential documents (NGSS Lead States, 2013; NRC, 2007, 2012), *high intellectual demand tasks* were defined as the ones that have potential for advancing students' thinking by inviting them to link observable phenomena and unobservable ideas. Our data suggested that high intellectual demanding tasks prompted students to do one or more of the following: (a) reason with science ideas to explain observable phenomena, (b) reason through data and observation to construct or evaluate explanatory models, or (c) develop arguments with use of evidence. In contrast, *low intellectual demand tasks* prompt students to (a) remember, recall, confirm, describe, or reproduce known scientific ideas, (b) practice skills procedurally, or (c) solve generic problems without connecting to existing knowledge (see the details of the coding in Table S2).

Launching the tasks: framing the tasks. The sources of data for analyzing the tasks *as launched* were field notes, audio-recorded conversation along with hand-out worksheets. The intellectual demand of instructional tasks at the stage of launch was analyzed using the same coding scheme presented above. In addition, we coded two other features of launches that revealed how the tasks were *framed*. This coding was informed by previous studies on effective facilitation of classroom discussion for sense-making (see for example Puntambekar, Stylianou, & Goldstein, 2007). One feature has to do with the ways in which the tasks are related to learners' experiences (code=contextualized vs. generic tasks). For example, in one of Hooke's law lesson, the teacher brought a bicep expander and had students to try it out, and then introduced the task, "figuring out which of the springs with two different stiffness will work better for this bicep expander" (code=contextualized). In contrast, the tasks in the other lessons were introduced mainly centering on science ideas, such as "collecting data about Hooke's law using two springs" (code=generic). The other feature of launch characterizes the ways in which the introduced tasks were connected to big ideas (code=connected vs. stand-alone). In some lessons, the launch included explicit discussion about the big science ideas that transcended the activity itself (code=connected). In contrast, in other lessons the tasks were launched as 'a stand-alone activity' with no or little discussion about how the task is related to big ideas beyond the ideas specific to the activity itself (see the details in Table S2).

Tasks as implemented—student learning opportunities. The sources of data for analyzing the tasks *as implemented* were classroom field notes, audio-recorded conversation, and produced student work. Students generated various artifacts from task implementation, either individually or as a group. Using these data sources, student learning experiences mediated by instructional tasks were coded in three ways (see Table S2). First, we examined *what students actually do and talk about* during the phase of task implementation. The level of intellectual work observed in students' *talk and performance* were coded as high or low (see the details of coding scheme in Table S2). Second, we examined *produced student work*. The produced work was coded as 'high' if the artifacts showed evidence of generating multiple relationships among elements of a natural phenomenon through intensive cognitive processing. It was coded as 'low' if the artifacts reflected students' simple or passive intellectual engagement, such as documenting, recording, reproducing information, repeating already demonstrated procedures, or solving generic problems. Finally, we analyzed *the range of participation* during task implementation by quantifying how many students engaged in the tasks (i.e., talk and/or work as prompted by teacher) to assess the influence of task implementation on students' learning (high: more than 80% (participation of students during task implementation), medium: 40~80%, low: less than 40%). Only the lessons that were coded as 'high' in all three sub-dimensions were considered as high quality learning opportunities ('Implementation of task' was coded as HHH in Table S3).

Phase II: Cross-analyses—Analyzing the relationship between the quality of instructional tasks and student learning opportunities. Using a lesson as a unit of analysis, we first categorized the 57 lessons into two groups, depending on the quality of tasks *as designed*: one with high (n=26 lessons) and the other with low intellectual demands (n=31 lessons). In each group of lessons, we examined the changes in intellectual demand and the quality of student learning opportunities, reflected in students' engagement with tasks, as the lessons unfolded during launch and implementation. Specifically, with the 26 lessons that were coded as high quality instructional tasks *as designed*, the lessons were categorized once again as two sub-groups: 15 lessons that maintained their intellectual rigor all the way through student learning opportunities (H-H-H), and the other 11 lessons that failed to maintain their rigor throughout the lesson (H-L-L or H-H-L). We then further analyzed qualitatively using field notes, post-observation interviews, and other teaching artifacts, focusing on when, how, and why these breakdowns happened. The 31 lessons that had low-level intellectual tasks as designed were also examined, focusing on whether there were any cases when the intellectual rigor increased as the tasks were launched or implemented (e.g., L-H-H or L-L-H).

Next we identified the lessons that were successful in providing rigorous student learning opportunities with respect to *all three sub-dimensions* ('Implementation of task' was coded as HHH in Table S3). Within the 15 lessons that met the criteria, we focused on the design of the tasks and how the intellectual rigor was maintained as tasks unfolded. This analysis helped us to identify the conditions under which students were likely to engage in robust forms of learning.

Findings

The analyses reveal three trajectories for students' intellectual demand as a lesson is designed, launched, and implemented: (a) staying at high intellectual demand (H-H-H), (b) beginning with intellectually demanding tasks but failing to maintain its rigor (H-H-L or H-L-L), and (c) staying at low intellectual demand (L-L-L) (see the coding results and the profiles of 57 lessons in Table S3). There was *no* case when a lesson designed with low intellectual demand was then framed and implemented with high rigor (e.g., L-H-H), nor were there cases of lessons that had low intellectual demand characterizing the framing (or launch) but "recovered" their rigor during the implementation, even if the initial quality of the task was high (e.g. H-L-H or L-L-H). Among the 57 lessons, about a half of the instructional tasks *as designed* appeared to have the potential of engaging students in higher level of intellectual work (n=26 out of 57, 45.6%). Among those 26 lessons, students' rigorous intellectual engagement was observed in only about a half of them (n=15 out of 26), and in the other half (n=11) students failed to engage in rigorous forms of science learning (see Figure 2).

--Insert Figure 2 about here—

We now present three illustrative cases (lessons) that describe different ways in which intellectual demand of tasks alter (H-H-H, H-L-L, and L-L-L). Each case provides a detailed description about the quality of tasks in each phase and student engagement mediated by the tasks. All together these cases demonstrate how the opportunities for students' intellectual engagement hinge not only upon the quality of planned instructional tasks (tasks as designed), but *how they are launched and implemented*.

Staying High: Designing Intellectually Challenging Tasks and Maintaining Intellectual Rigor (H-H-H)

In approximately one quarter of lesson (n=15 out of 57, 26.3%), intellectually demanding tasks were designed and implemented. Among the 15 lessons, 9 lessons were observed from three teachers, Ann, Bob, and Lynn. These teachers consistently designed and enacted intellectually challenging tasks across all three observed lessons. The other six lessons were observed in six different teachers' classrooms in one of three observations at the end of year. Notably, the tasks of the lessons in this *staying high* group were either designed or significantly modified by the teachers (as opposed to being taken "as is" from curricula) with two exceptions. We present one case—Ann's transfer of energy lesson.

Ann's 7th grade physical science: Transfer of energy. Ann is a Caucasian female who majored in biology. Upon her completion of teacher preparation program, she began to teach a 7th grade physical science at a highly under-resourced, ethnically and linguistically diverse urban middle school (free/reduced lunch rate: 86 %). While the school did not mandate the use of a particular curriculum, a district-level team provided a pacing guide for several curricular options. In addition, Ann's science department faced pressure from the administration to help increase students' standardized test scores. In particular, the administration requested that the science department focus on building students' scientific literacy skills. Therefore, Ann and her colleagues had to justify any activity or request for resources through the perspective of scientific literacy.

The observed lesson described here was part of a unit on types and transfers of energy. In this unit, students had been working on one overarching question: "How can a roller coaster car go through the same loop twice (once forward, once backward)?" Students were presented with a model of the event, drawn on poster paper by Ann, on the first day of this unit (see a snapshot in Figure S1). This model showed a roller coaster car making one loop. Students were asked to speculate what happened at one of three parts of the model: before, during, and after the roller coaster goes through a loop (i.e., start and going down, making a loop, and going up the ramp). Students who picked the same segment worked together to develop their models in small groups while discussing how potential energy was transferred to kinetic energy or vice versa after the rollercoaster car had left the starting point.

Task as designed: Revising the whole class model based on evidence. The focal question appearing on the teacher's projected slides was, "Which design gives the rollercoaster the right amount of energy to make it through the loop twice?" During the interview, Ann told us that she decided to do this activity because in the previous day's lesson several students raised questions about the whole class model that Ann developed. Based on these conversations with students, Ann made four different loop designs of a roller coaster car for this observed lesson. The task was designed to evaluate the four alternative models of loop design (A to D) using evidence from prior activities in the process of answering the focal question (see Figure S2). In this lesson, the task *as designed* was coded as "high" because this task, as constructed, invited students to engage in intellectually challenging work. That is, reasoning with evidence to evaluate and revise initial models about a complex real world phenomenon (codes HT-a & d in Table S2).

Launch of task: contextualizing & connecting a task to the big ideas. Ann began the lesson with a question, “Why is Ms. K’s model incorrect based on evidence from our activities?” During the introduction of the day’s activity, Ann drew students’ attention to several key elements relevant to explaining this complex real world phenomenon, such as the focal observable event (i.e., the car makes it through the loop twice—once forward, once backward), the roller coaster design, and unobservable science idea (i.e., energy). Ann gave the detailed instruction of the task, the rule of ‘knowing’ and expectations, and stated goals that she intended to achieve:

> Ok, now we are going to revisit our whole class model of the roller coaster. Just like when we did the Skater Girl[1], we are going to look at different parts of the model in addition to looking at the whole model. We are trying to understand, *as a class*, which roller coaster design gives the car enough energy to make it through the loop twice – once forward, once backward. We will do this by using ‘red light’, ‘green light’, and ‘yellow light’ sticker. Generally, we know that scientists can’t just say ‘yes’ or ‘no’ to a question. They have to back up their claims with evidence. Red, green, yellow stickers will help us do that as scientists...[*explain that red, green, yellow colors signal how well evidence supports a model*]…We are using your evidence from your activities to evaluate your model, just like scientists…[*explain the procedure in detail while modeling one example for herself*]…The idea is that you tell me both the activity and the evidence you got from the activity to support a model or to show how a model is faulty. We want to have one per class, so that we can start to use evidence across classes to build a stronger case for 1-2 models. The goal of this is to come to a class understanding of what works and what

---

[1] The anchoring phenomenon of the previous unit

> doesn't, before we move onto final explanations. Remember that we are scientists and we use evidence.

The task as launched was still intellectually challenging given that the completion of this task as set up required students' intensive reasoning such as evaluating evidences from activities to figure out which model works (codes=HT-a & d in Table S2). The focal phenomenon of this task was contextualized in a specific situation that most students of this class had experience (code=contextualized). In addition, the task was conceptually related to the big ideas of this unit—making sense of a complex real world phenomenon (code=connected). During this launch, students were repeatedly positioned as "scientists" and a member of the classroom learning community who contributed to collective construction of knowledge. Overall, during the launch the task was framed as "come to a class understanding" about this complex phenomenon "as a scientist."

Implementation of task & student engagement.

Student talk and performance. Ann, and the other two teachers who were successful in designing and maintaining intellectual demand across all three lessons, commonly used some forms of instructional tools, such as Ann's red/yellow/green light notecards (see Figure S2). In this focal lesson, students smoothly transitioned into small group conversation with their peers. They talked about whether and how their evidence from previous lessons supported the initial whole class model. They filled out either red-, green-, and yellow-dotted notecards, signaling whether they have convincing evidence (green), needed more evidence (yellow), or have contradictory evidence (red). As guided during the launch, students identified evidence from previous in activities that they had engaged and then began to wrestle for explaining their reasoning—"How the evidence supports a model or to show how a model is faulty." Ann visited each small group and had conversation with students for 5-6 minutes. As shown in the following, Ann first read and listened to students' current ideas. Then she asked a series of questions. Her questions either drew students' attention to some key elements or press students to further elaborate their ideas:

Ann: How's it going?

Student 1: Good. We are talking about the speed the car needs for going through the loop.

[Ann reads student cards]

Student 2: I think that the ramp activity proved that we need more speed on the car than the ramp in coaster A can give.

Student 3: Yep, the car doesn't have enough speed to make it up the ramp. When the marble started lower in the ramp activity, it didn't go up enough. It was too slow.

Ann: Good job using evidence from an activity, not just the activity itself. But I want to press on you to go further. Relate speed to energy for me. How did we use the ramp activity to talk about relationships between speed and energy? How can evidence from the ramp activity about speed and energy help us with the roller coaster?

Student 1: So we talked about in the ramp activity how the more speed something has, the more energy it has.

Student 3: Potential or kinetic [energy]…?

Student 2: Good question.

Student 1: Kinetic, since we're talking about a moving car.

Student 2: Ok, so if the car has more speed, like the marble, it has more kinetic energy and can get through the loop.

Ann: Good start. Write that down on the card.

Since the structure of the task required students to write down their ideas on the notecards (set up by task design), Ann provided herself with easy access to students' ideas during the visits. As demonstrated, students initially made connections at the observational level, drawing upon the ramp activity—i.e., the relationship between speed of the car and going through the loop. Ann pressed students to further reason about the focal phenomena (a roller coaster car making its way through the loop) and justify their claim by pointing out an unobservable science idea—energy. In response to Ann's question, the students collectively engaged in reasoning with this idea to build coherent relationships with the existing set of ideas. Students drew upon the scientific terminology (i.e., potential and kinetic energy), and used this to explain what they saw (the speed of the car, the height of the model, making a loop) (code=high intellectual engagement).

Produced student work and the range of student participation. In the index card, students explained why they thought some part of the model should be kept or revised referring to the lessons from the activities. For example, one student stated on the "red-light" note: "The equal height activity; It tells you, you can't go through it both sides aren't equal. Design D wouldn't work because the big loop makes it impossible to go through since it doesn't have equal energy." After students wrote down their ideas on the index card individually, they put the cards on the wall. They collectively produced the evidence poster on the classroom wall (see Figure S2). As demonstrated in the red/green-light cards on the wall, all students participated in the tasks (codes=HS-a, c & d; high level of participation).

Break Down: Beginning with Intellectually Challenging Tasks as Designed, but Failing to Maintain Intellectual Rigor (H-L-L or H-H-L)

The 'break down' of maintaining intellectual rigor was observed both in the phases of launch and implementation of tasks. Six out of 11 lessons that began with a high cognitive task were lowered in intellectual rigor during the launch (H-L-L). The others were during implementation (H-H-L). In the following, the process of lowering intellectual rigor in the phases of launch and implementation as well as its projected impact on students' learning is presented using one illustrative case—Nick's Momentum Lab lesson.

Nick's Momentum Lab. Nick was a Caucasian male who had BA in physics. His school was located in a suburban community serving diverse population of students with a wide range of families' socioeconomic backgrounds. The primary resources that Nick drew heavily on to plan his instruction was a set of curriculum materials developed by a university-based research team. Nick was one of the three physics teachers who used this curriculum as major resources for instruction in this project.

The focal lesson came from a 12th grade unit on momentum. This followed a unit on energy, and in the previous lesson students were introduced the concept of momentum through the conversation about two hypothetical scenarios: (a) being hit by two different balls that have different masses but the same speed (i.e., throwing a "bouncy ball" vs. a heavier baseball), and (b) being in a vehicle that collides with a light car moving at 60 miles per hour vs. with a heavy car moving at 5 miles per hour. The idea of and formula for momentum ($F\Delta t = P$, $P = mv$) were presented to students in the prior lesson.

Tasks as designed. The main task *planned* for this focal lesson, that was observed in the slides and outlined on students' worksheets, was solving a problem in an interesting scenario by applying the ideas of conversation of momentum. Students were framed as an actor in the movie of Jurassic Park who was chased by a dinosaur, a velociraptor, and they had to decide which ball between bouncy and clay ones is more likely to close the door. The question stated on the slide was, “Would you rather have a super bouncy ball or a sticky lump of clay (assuming that the mass and size of the two balls are equal)? Explain your reasoning” (see the snapshot of the slide & worksheet in Case S1). As designed, this task has a potential of engaging students in a higher level of intellectual work. The key question stated in the worksheet was, “Let’s use the online collision simulation to help you find the answer to this question” (see the worksheet in Case S1). Then the task prompted students to collect and analyze data about any changes of momentum before and after collisions using a computer simulation. The task was designed to guide students in coming up with the law of conservation of momentum, both conceptually and mathematically, by working with data as a group. The task *as designed* was coded as “high” because solving the focal problem by engaging in this set of work required students’ intensive reasoning with the science ideas, moving back and forth among mathematical symbolism, conceptual definitions, and imagined phenomena (code: HT-a and b). Completing this task requires a complex reasoning that involves taking into account of multiple variables, such as elasticity, mass, and velocity.

Tasks as launched. The intellectual demand of the task, however, lowered in the phase of launch by framing the tasks mainly as ‘filling out the data spreadsheet.’ Nick began the lesson by reading out-loud the two learning targets on the screen, which was scripted by Teacher Guideline of the curriculum materials. Nick spent about 10 minutes stating the instructions of what individual students had to do while going over the worksheet procedure (see the full transcripts of the launch in Case S1). He explained the steps to follow, including where to record the data, and how to input the data. Each student was assigned to fill out a designated part of the data spreadsheet. At the end of the launch, Nick said, “After you do some data analysis, you are just gonna work through a series of questions to get you to think about the law of conservation of momentum” (codes=LT-e; generic & stand-alone). Nick pointed out the dinosaur problem at the end of the worksheet but there was no discussion of the connection with this data collection activity. This form of a launch that lowered the intellectual demand was typical across all five observed lessons in Nick’s classroom, as well as other teachers in this group. After all, the tasks requested relatively lower level of intellectual work because students mostly work to confirm or illustrate known scientific ideas using the data and automatically calculated numbers in the provided spreadsheet. Notably, the situation was mostly framed as ‘completing the worksheet’ in those lessons. Despite the intellectually challenging tasks as designed, the intellectual work to solve the problem was neither guided nor communicated in the phase of launch.

Task as implemented & student engagement.

Student talk and performance: The intellectual demand manifested in both student talk and performance was relatively low in the phase of task implementation. After the launch, students moved to their lab table, logged into computers, and started their work. Nick circulated the groups to respond to students’ questions during the rest of the period. Students collected data as a small group using a computer simulation following the procedures in the worksheet. The simulation was designed to measure the amount of momentum when two moving balls collide in a closed system with the changes of key variables (mass, velocity, and elasticity). However, there were a great deal of confusions among students, such as which button on the screen they should click in the simulation, what numbers they should put for each variable, what the positive and negative means, and what they ought to record in the spreadsheet, etc. Most small group conversation was focused on clarifying the procedures, ‘what to do’ but there was little sense-making conversation (see the samples of small group conversation in Case S1). For example, in one group, two male students were unsure what number they should put in for mass and velocity in the simulation. One student called and asked, “Can I put just like any numbers?” As responding to the students, Nick revealed the key patterns (i.e., the law of conservation of momentum) for himself that students were supposed to find from the data. Nick said, “Whatever number are, any numbers provided will satisfy the law.” This conversation shifted the nature of the work from students’ active reasoning with data (i.e., finding the relationship between initial and final momentum) into simply confirming or illustrating the known scientific law by plugging numbers into the simulation. During another small group conversation with two female students, Nick again lowered the intellectual demand of the task. These two female students, and many other students in this classroom, were confused about the negative sign of the velocity. These two female students called on Nick and asked, “Is velocity of the two [balls] supposed to be negative all the time?” Nick gave them the answer right away, “Not necessarily”:

> Student 1: (ask to her peer) Does velocity have to be different all the time?
>
> Student 2: I honestly don’t know what is going on.
>
> Student 1: [Call on Nick] Mr. C, is velocity of two supposed to be negative all the time?
>
> Nick: uhmmm…okay. Not necessarily. It can be if you always start like two on one side together. Does it make sense? So let say you always start two on the right hand side, then two will always be negative. But somehow the two gets on from this side, so if you remember that two always needs to be on, on the right then it will always be negative. Easy enough, uhmm?
>
> Student 1: Thank you.

This form of interaction during the small group work was fairly typical across observations in Nick's classrooms. During the small group work, Nick worked hard to help students get the tasks done by clarifying procedures and answering their questions. By doing so, the intellectual demand built in the task design became lowered because either the intellectual space for students to struggle and reason for themselves was removed or the nature of the task was transformed.

Produced student work and the range of participation: Produced work reflected the deep confusion students experienced during the task implementations. The reported data and the calculations appearing on the data table were inaccurate in most provided samples of student responses. In all the nine samples of student work, students' explanations about the bouncy and clay balls were neither matched with nor connected to the data that they collected. For example, one student, who were recognized as a high achieving student by Nick, provided evidence of engaging in reasoning with everyday knowledge with evidence in her response to the pre-assessment: "Um, I think that the clay would be more effective because it is solid and I feel that it will be ok if you miss, you could hit the velociraptor and temporarily blind it, so that way you could grab a T-Rex to kill the velociraptor." This student collected inaccurate data from the simulation, miscalculated the initial and final momentum. The final explanation was changed to the correct answer, like all the other students. But her response did neither show any progress in her thinking nor in-depth understanding despite the use of academic language. This student stated, "The bouncy ball [is more likely to close the door] because it takes momentum to close the door and send the ball back" (see this students' pre and post responses in Case S1). During the implementation, most students logged in the simulation and click the button while plugging the numbers (codes=LS-b & c; high level of participation).

Staying Low: Low intellectually demanding tasks were designed, launched, and implemented (L-L-L)

Thirty-one out of 57 lessons (54.4%) began with a low intellectual demand task as designed, and maintained a low level of intellectual rigor throughout the lessons. About half of the lessons in this group (n=15) were observed in five teachers who consistently designed and enacted lower level of intellectual work across all three observed lessons. For this paper, we present one illustrative case—Sarah's Battery Lab. This case is particularly interesting because of the natural comparison established with similar context, the same grade level and the same topic of the lesson with the case of Ann in stay high group. Just like the case of Ann, Sarah taught 7th grade physical science at a highly under-resourced urban middle school that had ethnically and linguistically diverse populations of students. The topic of this focal lesson was the same as Ann's lesson—types and transfer of energy. This case demonstrates how instructional tasks taught in similar contexts can be designed and enacted in a different way, and how it affects student learning opportunities.

Sarah's Battery Lab: types and transfer of energy. In the previous day students were given a lecture about energy transformation in battery and light bulb. They took a note about the 1st transfer from "chemical to electrical energy" (from liquid to wire), and the 2nd transfer from "electrical to light energy" (from wire to light bulb). Sarah invited us on the lab day when students built a battery following the lecture day.

Task as designed—Build a battery to light a bulb. The planned tasks of this focal lesson were building a battery and testing it using a light bulb as a pair to answer the four questions: "(a) How can a light bulb test whether you made a battery? (b) Draw a diagram of your battery, (c) What did you observe as you made the battery? (d) What transfer of energy did you observe?" Despite the hands-on engagement with the task, the task *as designed* required lower levels of intellectual work for three reasons. First, the task prompted students to observe and describe what they saw, without further pressing for deeper reasoning such as "how" or "why." Second, most questions had a single and discrete correct answer expected by the teacher, such as chemical energy is transferred into light energy. Third, the task was designed to illustrate or confirm the known science ideas presented in the prior lecture. Therefore, this task *as designed* was coded as low (codes=LT-a & -c).

Task as launched—covering standards and giving directions. The task as launched was coded as "low" because the task showed the least amount of complexity, and the completion of this task only requires students to procedurally follow the directions, and recall the process of energy transfer in battery that they took a note in the previous day. As the bell rang, Sarah called on three students from volunteers who raised their hands. Each student read out-loud the Standards—"Energy Forms," Learning Target—"I will build a battery to light a light bulb," and Do Now—"Make a list of everything you already know about batteries." After students wrote down what they knew about batteries, Sarah asked students to share their lists with class. When Sarah explained the procedure of the lab, a student, Tom raised his hands and asked a question:

Teacher: Tom, questions for me?

Tom: You know the [battery] that you put your finger on it and does it shock you?

Teacher: Yes, those are, the square battery that you are talking about, yeah?

[After Tom's comment, students are getting excited and starting to talk about it]

Michelle: [talk to her friend] I used to do that. That hurts.

Teacher: Listen, five, four, three two, one! We are going to read through all the procedures today. After reading through all 10 steps, then we are going to set up lab work. (see the full transcripts of Sarah's launch in the Case S2)

It appeared that Sarah framed the task as 'complete the worksheet following directions' to cover one standard ("types of energy"). The 7th grade student, Tom (and other students), who had yet to develop "doing school" frame, seemed to bid for his frame, that is inquiring his question about every experiences. However, Tom's question did not make any change in the flow of the instruction as Sarah decided to stick to her plan (codes=LT-a & c; generic and stand-alone).

Task as implemented & student engagement.

Student talk and performance: The intellectual demand manifested in both student talk and performance was low in the phase of task implementation. Students built a battery following the 10 steps described in the textbook and responded to the discussion questions as prompted by the task. After building batteries, Sarah reminded students of drawing their batteries in their notebooks and answering the question of "what transfer of energy did you observe?"

As students finished building their batteries and filled out the answers to the questions on a worksheet, Sarah circulated and posed the same questions as on the worksheet and students provided one or two words answers. Sarah's particular press was on naming energy transfers with the talk pattern of I-R-E (initiating-responding-evaluating) (see the samples of the small group conversation in Case S2). Students mostly engaged in procedural talk when they worked together, and responded to the teacher's question with one or two short words.

Produced student work and the range of participation: Students produced their notebook from this task implementation (see a sample of student work in Case S2). The notebook consisted of Standards, Learning target, Do Now that they copied down. The notebook also included a drawing of their battery and responses to three questions. Students recorded observation and reproduced the known scientific ideas to produce this learning artifact. The intellectual work manifested in this artifact was low although most students participate in building batteries (codes=LS-a & b; high participation).

Discussion

Our analyses suggest that in order to create opportunities for advancing scientific thinking, a lesson must begin with intellectually challenging tasks *and* maintain its rigor during the enactment. In this discussion we first conjecture about how and why beginning a lesson with the high quality instructional tasks is essential. We then present three definable characteristics of the launch and the implementation observed from the lessons that maintained intellectual rigor theorizing how it may come to create conditions for students' rigorous engagement.

Getting off on the right foot: Creating learning opportunities for students that advance scientific thinking with intellectually challenging instructional tasks

The findings highlight the importance of beginning lessons with high quality tasks—meaning intellectually challenging and complex tasks—for providing robust learning opportunities in classrooms. Across the 57 observed lessons, there was no indication that students engaged in meaningful sense-making conversation when the planned tasks were of low demand (tasks *as designed*).

The analyses suggest that the quality of tasks *as designed* sets up the initial potential for students' engagement, which could then be either maintained or lowered as the lesson unfolds. We theorize that high quality tasks are not just arbitrarily difficult, rather they create the appropriate level of epistemic and procedural uncertainty for a particular group of students learning a particular topic in a particular context and moment in time (see Manz, 2014; Tiberghien et al., 2014). Researchers who attend to the joint construction of the classroom knowledge by the teacher and students note, “uncertainty is an essential component of the growing of knowledge, in the scientist’s activity as well as in the science classroom activity” (Tiberghien et al., 2014, p. 934). The uncertainty of knowing and the students’ awareness of it allow them to question claims, argue about ideas, and in the process gain a better understanding of the science through participation in the authentic socio-intellectual processes by which knowledge is refined over time. Researchers also note that wrestling with epistemic certainty/uncertainty is strongly associated with the development of an intellectual autonomy of students (Engle & Conant, 2002; Manz, 2014). We speculate that the appropriate level of uncertainty, set up with the carefully designed instructional tasks, affords a space for students’ meaningful intellectual work. Examples here are the two 7th grade physical science lessons that addressed the same science ideas in similar classroom contexts. In Ann’s roller coaster lesson, the task offered the basic structure for productive social and intellectual interactions, such as reviewing ideas from the prior activities, and making connections among and between the ideas through collective conversation as a small group. The task was puzzling and complex enough for this group of middle school students in that moment who had been engaged in resolving one major science question over the past several days. The level of uncertainty and complexity appeared to be appropriate and sufficient given that the students could develop evidence-based explanations after exchanging and reconstructing ideas. It should be noted that the task of evaluating four-loop designs for a roller coaster car emerged from the class conversation of the prior lesson. In contrast, the planned task of Sarah’s battery lesson showed little uncertainty either in terms of the knowledge or procedure necessary to complete the work. Even though most students participated in this hands-on activity, the procedure to complete this task was instructed clearly through the textbook, and there was one correct answer communicated through the prior day's lecture. In other words, there was little or no uncertainty either in relation to knowledge construction work or the procedures for developing new forms of evidence or information.

Our analysis suggests that a lesson likely begins with high quality instructional tasks that create appropriate level of uncertainty when the tasks are designed or selected *responsively* with careful attention to students' everyday experiences and questions that emerge in the flow of instruction, instead of simply following recommended curriculum materials. The majority of intellectually demanding tasks observed in classrooms were either designed by the teacher, or substantially modified from existing curricula (12 out of 15 lessons, 80%). Ann and other two teachers, who consistently provided intellectually challenging learning opportunities, were the people who designed their instructional tasks by substantially modifying various existing resources (see Table S3). The designed tasks of these three teachers typically addressed a puzzling question or a phenomenon that were directly related to students experience in their daily life, as demonstrated with Ann's case. The potential of responsively designing instructional tasks to afford students' intellectual engagement is also implicated by the Sarah's battery lesson. Recall one student, Tom and his question during the launch of the tasks—this query activated his everyday knowledge relevant to the topic ("You know the ones that you put your finger on it and does it shock you?"). We believe that taking up Tom's question, that was deeply rooted in his everyday experiences as an aspect of task, boosted the intellectual rigor of this lesson by increasing the epistemic uncertainty for that particular group of students. It is highly likely that the tasks that are responsively designed, based on students' ideas or questions, often bear the appropriate level of uncertainty for a particular group of students in a particular moment.

Maintaining intellectual rigor during launch and implementation of tasks

Our further analysis shows that beginning a lesson with a high quality tasks is essential but not sufficient in creating robust student learning opportunities in classroom. Among the 26 lessons that began with high demand tasks, only 15 provided evidence of students' active engagement for sense-making (see Figure 2). The cross-analysis of the 26 lessons revealed three definable characteristics of launch and implementation of tasks. Two characteristics have to do with the ways in which the tasks were framed during the launch. The third one involves teachers' deliberate use of instructional tools or scaffolds that assist students' implementation of tasks.

The first characteristic observed during the launch was that the tasks were communicated with students as a way to understand contextualized phenomena (13 out of 15 lessons; see Table S3), rather than completing a generic task. We conjecture that framing tasks as a way of understanding complex phenomena or solving a question that matters to students may help the teacher and students to *co-generate* the meaningfulness of the work. It may help students formulate an epistemic goal for the knowledge construction work entailed in the proposed tasks, which is essential for later engagement in meaningful scientific practices during task implementation (Berland et al., 2015). Recall how the task of Ann's roller coaster lesson was framed as the work for building "class understanding" while positioning students as scientists who were expected to work together in order to solve *their* puzzle. Explicit discussion on the rule of knowing (e.g., "scientists can't just say 'yes' or 'no' to a question. They have to back up their claims with evidence") seemed to assist students to engage in the procedure meaningfully beyond rote performance. Although cognitively demanding intellectual work can be built into the carefully designed tasks, we conjecture that it is through this process of *framing* in which intellectual demand of the tasks is co-generated as students formulate their sense of "what is it that is going on here?" (Goffman, 1974) and constructing the meaningfulness of the work.

Another distinctive characteristic of launches observed from the 15 lessons was the richness of the language and ideas that surfaced on the social plane of the classroom through conversation between teachers and students. In those lessons, the connection to big ideas was explicitly discussed during the launch (n=14 out of 15 lessons coded as "connected"; see Table S3). Consistent with Jackson and her colleague's findings (Jackson et al., 2013), students more likely engaged in in-depth sense-making conversation during the task implementation when the intellectual demand of tasks *as launched* was coded as high (see intellectual demand of task as launch and intellectual level of student talk in Table S3). We conjecture that connecting the tasks to big ideas during the launch activated conceptual resources that students could draw upon to complete cognitively demanding work. Intellectually demanding science tasks, such as the one from Ann's roller coaster lesson, addressed complex real world phenomena. When lesson launches began with intellectually demanding tasks, students were set up to work on several ideas, including their own existing ideas, to generate multiple and coherent relations among them. As shown in Ann's case, when teachers and students discussed the connection of the task to the big idea of the unit or prior activities, the key elements of observable phenomena and unobservable science ideas from the prior lessons became readily available and accessible as cognitive resources for students' sense-making conversations. As reported by Puntambekar et al's study (2007), teachers' deliberate talk moves to connect the tasks to ideas that transcended the activity at hand seemed to help students to successfully engage in rich sense-making conversation by activating and enriching available cognitive resources and seeding ideas for how to interact with one another during the implementation.

With respect to the implementation of tasks, distinctive use of scaffolding tools was observed in lessons that maintained intellectual rigor (13 out of 15 lessons, see Table S3) whereas lessons that failed to maintain rigor typically did not utilize any scaffolding tools. Tools in the lessons that maintained intellectual rigor publically represented students' ideas, such as Ann's green/yellow/red light cards; teachers used these tools to structure and problematize the work (Reiser, 2004). On one hand, the tool in Ann's class structured the task by guiding students through key components and supporting their performance. On the other hand, the tool forced students to incorporate key disciplinary ideas and practices in their work by requiring that they evaluate four different loop designs, to review and connect evidence from class activities and to explain their thinking at length. As such, disciplinary ideas, such as energy and speed, were problematized in the process of solving a complex problem. It should be noted that the teachers who were successful in beginning a lesson with intellectually challenging tasks *and* maintaining intellectual rigor were the ones who actively utilized various instructional tools during task implementation (see Table S3). Reiser (2004) states that "learners, tools, and teachers work together as a system", and "scaffolded tools can create opportunities, but whether learners capitalize on these opportunities depends on the expectations and practices established in the classroom" (p. 298). We speculate that the deliberate use of tools, as a part of a carefully constructed and coordinated system of instruction, make it possible for students to have interactions effectively as set up by tasks, therefore advancing their thinking.

Conclusion & Implications

We began our inquiry with the goal of better understanding how to create opportunities for students to advance their scientific thinking through the informed use of instructional tasks. Building upon the studies in the field of mathematics education, we analyzed the potential for students' intellectual work built into the tasks across the three phases of instruction (tasks as designed, launched, and implemented), and documented how the intellectual demand of the unfolding task deepened (or failed to deepen) students' engagement in science. It should be noted that the lessons analyzed in this study were taught by 1st year science teachers who faced various challenges in selecting, modifying or enacting curriculum in new contexts. Even though our findings are consistent with the analyses of mathematics lessons taught by experienced teachers, the outcomes of our study should be interpreted with caution, specifically considering that novices may use curricular materials and enact lessons differently than educators who have been in the classroom for several years.

This study does allow us to theorize how students' learning opportunities are the product of both design and goal-directed pedagogical support throughout the duration of a lesson. First, the findings strongly suggest that it is a *sequence of coordinated events* that constitute opportunities to learn in the classroom. There are at least three kinds of task-related mediation, facilitated by teachers, that help translate curriculum into learning opportunities for students: (a) selecting and/or modifying tasks for a lesson, (b) launching high-quality tasks by framing the work in light of learning goals that transcended the task itself, and (b) providing deliberate and appropriate support during the implementation. In our sample, students' meaningful intellectual engagement was observed when: (a) a lesson began with high quality instructional tasks—complex tasks that bore appropriate levels of epistemic and procedural uncertainty for students, (b) the task was framed in a way that helped learners connect themselves to the situations specific to the tasks, as well as seeing the relationship to disciplinary ideas during the launch, and (c) students' engagement with these complex tasks was guided with use of scaffolding tools that structured the work and problematized disciplinary content. These findings support the argument that teachers need to be re-positioned as instructional designers who work to create particular forms of learning opportunities, rather than as passive consumers of pre-developed curricula (Brown, 2009; Clandinin & Connelly, 1992; Lloyd, Remillard, & Herbel-Eisenmann, 2009; Remillard, 2005). This study begins to shed light on the kinds of pedagogical decisions and practices during various parts of lessons that influence students' intellectual engagement.

This study also highlights a phase of science instruction that is not often a focus of attention in research or in classroom practice—the launch of tasks (see the launch of mathematics instruction in Jackson's et al, 2013). Instead of conceptually conflating launch and implementation as same activity, this study examined these phases separately, with attention to their unique potential to influence students' opportunities to learn. We argue that this approach increases the analytical power applied to what is impacting students' intellectual engagement with complex tasks.

One important point emerging from the analyses is the critical role that students themselves play in setting up rigorous learning opportunities. From a situated perspective (Greeno, 2006; Greeno & Gresalfi, 2008), learning occurs in and through systems of activity that are jointly constructed in classrooms. Within this frame, students do not merely respond to instruction, rather, their particular forms of participation—asking questions, sharing ideas, creating artifacts—constitute opportunities for the intellectual engagement of their peers and for the class as a collective (see Maskiewicz & Winters, 2012). Our findings suggest that curriculum developers and teachers can only engineer *potential* intellectual demand into the design of tasks that mediate students' interactions, and that rigorous learning opportunities cannot be created if students fail to see either the meaningfulness of the work or do not exert the intellectual agency to engage in the work. A corollary question then, is whether it is possible to create rigorous learning opportunities for *all* students without gaining greater access to their diverse experiences and existing repertoires of sense-making practice. Future studies that investigate the roles students play in curriculum design and enactment will be critical in addressing issues of equity and well as rigor.

Acknowledgement

We are grateful to the teachers who made this research possible. We also thank to Lynsey Gibbons who helped to develop ideas for analyses and provided thoughtful comments. This article benefitted from the critical feedback of three anonymous reviewers and the editors of the journal.

References

Banilower, E.R., Smith, P.S, Weiss, I. R., Malizahn, K.A., Campbell, K.M., & Weis, A.M. (2012). *The report of the 2012 national survey of science and mathematics education*. Chapel Hill, NC: Horizon Research, Inc.

Berland, L.K., Schwarz, C.V., Krist, C., Kenyon, L., Lo, A.S., & Reiser, B.J. (2015). Epistemologies in practice: Making scientific practices meaningful for students. *Journal of Research in Science Teaching*, n/a-n/a.

Brown, M.W. (2009). *Toward a theory of curriculum design and use: Understanding the teacher-tool relationship*. In J. T. Remillard, B. A. Herbel-Eisenmann & G. M. Lloyd (Eds.), Mathematics teachers at work: Connecting curriculum materials and classroom instruction (pp. 17-36). New York: Routledge.

Clandinin, D.J., & Connelly, F.M. (1992). *Teacher as curriculum maker*. New York: Macmillan.

Consortium for Policy Research in Education. (2011). *Science instruction in Newark public schools*. New York, NY: Corcoran, T.B., & Gerry, G.B.

Doyle, W. (1983). Academic Work. *Review of Educational Research, 53*(2), 159-199.

Doyle, W., & Carter, K. (1984). Academic tasks in classrooms. *Curriculum Inquiry, 14*(2), 129-149. doi: 10.2307/3202177

Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a Community of Learners classroom. *Cognition and Instruction, 20*, 399-483.

Fogleman, J., McNeill, K.L., & Krajcik, J. (2011). Examining the effect of teachers' adaptations of a middle school science inquiry-oriented curriculum unit on student learning. *Journal of Research in Science Teaching, 48*(2), 149-169.

Forbes, C. T. (2011). Preservice elementary teachers' adaptation of science curriculum materials for inquiry-based elementary science. *Science Education, 95*(5), 927-955.

Goffman, E. (1974). *Framing analysis: An essay on the organization of experience*. New York: Harper & Row.

Goodrum, D., Hackling, M., & Rennie, L. (2000). *The status and quality of teaching and learning of science in Australian school*. Research report prepared for the Department of Education, Training, and Youth Affairs (August, 2000).

Greeno, J.G. (2006). Learning in activity. In K. R. Sawyer (Ed.), *The cambridge handbook of the Leanring Sciences* (pp. 79-96). New York, NY: Cambridge University Press.

Greeno, J.G. (2009). A Theory bite on contextualizing, framing, and positioning: A companion to Son and Goldstone. *Cognition and Instruction, 27*(3), 269-275.

Greeno, J.G., & Gresalfi, M.S. (2008). Opportunities to learn in practice and identity. In P.A. Moss, D. Pullin, J.P. Gee, G. Haertel & L.J. Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 170-199). New York: Cambridge University Press.

Hammer, D, Elby, A., Scherr, R, & Redish, E. (2005). Resources, framing, and transfer. In J. Mestre (Ed.), *Transfer of learning: Research and perspectives* (pp. 89-120). Greenwich, CT: Information Age Publishing.

Hand, V., Penuel, W.R., & Gutierrez, K.D. (2012). (Re)Framing educational possibility: Attending to power and equity in shaping access to and within learning opportunities. *Human Developmenet, 55*, 250-268.

Harris, C. J., Penuel, W. R., D'Angelo, C. M., DeBarger, A. H., Gallagher, L. P., Kennedy, C. A., Cheng, B. H. and Krajcik, J. S. (2015). Impact of project-based curriculum materials on student learning in science: Results of a randomized controlled trial. *Journal of Research in Science Teaching, 52*: 1362–1385.

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

Jackson, K., Garrison, A., Wilson, J., Gibbons, L., & Shahan, E. (2013). Exploring relationships between setting up complex tasks and opportunities to learn in concluding whole-class discussions in middle-grades mathematics instruction. *Journal for Research in Mathematics Education, 44*(4), 646-682.

Jimenez-Aleixandre, M.P., Bugallo Rodriguez, A., & Duschl, R.A. (2000). "Doing the lesson" or "doing science": Argument in high school genetics. *Science Education, 84*(6), 757-792.

Kang, H., & Anderson, C. W. (2015). Supporting preservice science teachers’ ability to attend and respond to student thinking by design. *Science education, 99*(5), 863-895.

Kang, H, Thompson, J, & Windschitl, M. (2014). Creating opportunities for students to show what they know: the role of scaffolding in assessment tasks. *Science Education, 98*(4), 549-742.

Lloyd, G.M., Remillard, J.T., & Herbel-Eisenmann, B.A. (2009). Teachers' use of curriculum materials. In J. T. Remillard, B. A. Herbel-Eisenmann & G. M. Lloyd (Eds.), *Mathematics teachers at work: Connecting curriculum materials and classroom instruction*. New York, NY: Routledge.

Lynch, S.J., Pyke, C., & Grafton, B.H. (2012). A retrospective view of a study of middle school science curriculum materials: Implementation, scale-up, and sustainability in a changing policy environment. *Journal of Research in Science Teaching, 49*(3), 305-332.

Manz, E. (2014). Representing student argumentation as funtionally emergent: From scientific activity. *Review of Educational Research.*

Maskiewicz, A.C., & Winters, V.A. (2012). Understanding the co-construction of inquiry practices: A case study of a responsive teaching environment. *Journal of Research in Science Teaching, 49*(4), 429-464.

Mayer, R.E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 31-48). New York, NY: Cambridge University Press.

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

McNeill, 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.

McNeill, K.L., & Pimentel, D.S. (2010). Scientific discourse in three urban classrooms: The role of the teacher in engaging high school students in argumentation. *Science Education, 94*(2), 203-229. doi: 10.1002/sce.20364

McNeill, K.L., Pimentel, D.S., & Strauss, E.G. (2013). The impact of high school science teachers' beliefs, curricular enactments and experience on student learning during an inquiry-based urban ecology curriculum. *International Journal of Science Education, 35(15),* 2608-2644. doi: 10.1080/09500693.2011.618193

Merriam, S.B. (2009). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Jossey-Bass.

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states. Washington*, DC: The National Academies Press.

NRC. (2007). *Taking science to school: Learning and teaching science in grade K-8*. Washington, DC: National Academies Press.

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

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. & Dillon, J. (2008). *Science Education in Europe: Critical Reflections*. A Report to the Nuffield Foundation. Kings College London.

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

Penuel, W.R., Gallagher, L.P., & Moorthy, S. (2011). Preparing teachers to design sequences of instruction in earth systems science: A comparison of three professional development programs. *American Educational Research Journal, 48*(4), 996-1025.

Puntambekar, S, Stylianou, A, & Goldstein, J. (2007). Comparing classroom enactments of an inquiry curriculum: Lessons learned from two teachers. *Journal of the Learning Sciences, 16*(1), 81-130. doi: 10.1080/10508400709336943

Reiser, B.J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences, 13*(3), 273-304.

Remillard, J.T. (1999). Curriculum materials in mathematics education reform: A framework for examining teachers' curriculum development. *Curriculum Inquiry, 29*(3), 315-342.

Remillard, J.T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research, 75*(2), 211-246.

Roth, K., & Garnier, H.E (2007). What science teaching looks like: An international perspective. *Educational Leadership, 64*(4), 16-23.

Roth, K., Garnier, H.E., Chen, C., Lemmens, M., Schwille, K., & Wickler, N.I.Z. (2011). Videobased lesson analysis: Effective science PD for teacher and student learning. *Journal of Research in Science Teaching, 48*(2), 117-148.

Schneider, R.M., Krajcik, J., & Blumenfeld, P. (2005). Enacting reform-based science materials: The range of teacher enactments in reform classrooms. *Journal of Research in Science Teaching, 42*(3), 283-312. doi: 10.1002/tea.20055

Sohmer, R., Michaels, S., O'Connor, K.M., & Resnick, L.B. (2009). Guided construction of knowledge in the classroom: The troika of talk, tasks and tools. In B. Schwarz, T. Dreyfus & R. Hershkowitz (Eds.), *Transformation of knowledge through classroom interaction*. New York, NY: Routledge.

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

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

Tekkumru-Kisa, M., Stein, M.K., & Schunn, C. (2015). A framework for analyzing cognitive demand and content-practices integration: Task analysis guide in science. *Journal of Research in Science Teaching, 52*(5), 659-685.

Thagard, P. (2008). Explanatory coherence. In J. E. Adler & L. J. Rips (Eds.), *Reasoning* (pp. 471-513). Cambridge, England: Cambridge University Press.

Tiberghien, A., Cross, D., & Sensevy, G. (2014). The evolution of classroom physics knowledge in relation to certainty and uncertainty. *Journal of Research in Science Teaching, 51*(7), 930-961.

van de Sande, C. C., & Greeno, J.G. (2012). Achieving alignment of perspectival framings in problem-solving discourse. *Journal of the Learning Sciences, 21*(1), 1-44.

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

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.

Zangori, L., Forbes, C.T., & Biggers, M. (2013). Fostering student sense making in elementary science learning environments: Elementary teachers' use of science curriculum materials to promote explanation construction. *Journal of Research in Science Teaching, 50*(8), 989-1017.

Supplementary Information linked to the online version of this paper at Wiley-Blackwell

(a) Table S1.pdf (b) Interview protocol S1.pdf (c) Table S2.pdf (d) Table S3.pdf
(e) Figure S1.tif (f) Figure S2.tif (g) Case S1.pdf (h) Case S2.pdf