

Change in Thinking Demands for Students Across the Phases of a Science Task: An Exploratory Study

Miray Tekkumru-Kisa¹  · Christian Schunn² ·
Mary Kay Stein² · Bertha Reynolds²

© Springer Science+Business Media Dordrecht 2017

Abstract Science education communities around the world have increasingly emphasized engaging students in the disciplinary practices of science as they engage in high levels of reasoning about scientific ideas. Consistently, this is a critical moment in time in the USA as it goes through a new wave of science education reform within the context of Next Generation Science Standards (NGSS). We argue that the placement of high demands on students' thinking (i.e., a high level of thinking) in combination with positioning students to use disciplinary practices as they try to make sense of scientific ideas (i.e., a deep kind of thinking) constitute critical aspects of the reform. The main purpose of this paper is to identify and describe the kinds and levels of thinking in which students engage when they are invited to think and reason as demanded by NGSS-aligned curricular tasks. Our analysis of video records of classrooms in which an NGSS-aligned, cognitively demanding task was used, revealed many ways in which the aspirational level and kind of student thinking will not be met in many science classrooms. We propose a way of characterizing and labeling the differences among these kinds and levels of thinking during the implementation of a reform-based biology curriculum. These categories, which focus on two important features emphasized in the NGSS, can help us to better understand, diagnose, and communicate issues during the implementation of high-level tasks in science classrooms.

Keywords Scientific practices · Cognitive demand · Student thinking · Science tasks

An earlier version of this paper was presented at the National Association for Research in Science Teaching (NARST) International Conference in April 2015.

✉ Miray Tekkumru-Kisa
mtekkumrukisa@fsu.edu

¹ Learning Systems Institute & School of Teacher Education, Florida State University, Tallahassee, FL, USA

² Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA, USA

Science education communities around the world have increasingly been challenged with coordinating goals for students' understanding of scientific knowledge with the development of their scientific reasoning so that they can make educated decisions on a range of contemporary issues (Erduran 2014). For example, a recent Delphi study based on an Australian Science Education Research Association panel on the future of science education highlighted one essential element: Supporting students to become scientific thinkers and reasoners rather than focusing solely on students' accumulation of scientific knowledge (Aubusson et al. 2016). One of the experts in the panel stated:

There must be an opportunity for students to pose questions, design investigations, collect evidence, actively reason with observations and evidence and draw evidence-based conclusions and the evidence-based conclusions have some knowledge claims supported by the evidence base. This requires sophisticated reasoning. (p. 215).

The expert's comment emphasizes engaging students in the disciplinary practices of science as they engage in high levels of reasoning about scientific ideas.

This is a critical moment in time in the USA as well as it goes through a new wave of science education reform within the context of Next Generation Science Standards (NGSS; NGSS Lead States 2013). The Framework for K-12 Science Education (National Research Council 2012) together with the NGSS have established an overarching vision for science teaching and learning that emphasizes supporting students' in-depth development of core explanatory ideas *and* engagement in scientific and engineering practices (Reiser 2013). This new vision is a departure from how students currently learn science in most US classrooms (Banilower et al. 2013; Reiser 2013).

One distinguishing feature of the Framework and the NGSS, as compared to how teaching science as inquiry has been promoted in the past, is the focus on the scientific practices. They emphasize that "Learning science ... involves the *integration* of the knowledge of scientific explanations (i.e., content knowledge) and the practices needed to engage in scientific inquiry ..." (NRC 2012, p. 11). Another important feature is to get students to *think deeply* about the science content and practices. For example, the underlying rationale for focusing on a limited set of core ideas is to allow for "...deep exploration of important concepts, as well as for students to develop meaningful understanding" (NRC 2012, p. 25). Here, we argue that these two features—the placement of high demands on students' thinking (i.e., a high *level* of thinking) *in combination with* positioning students to use disciplinary practices as they try to make sense of scientific ideas (i.e., *kind* of thinking)—constitute the most critical and most ambitious aspects of this new vision for science education. The question to be asked is how these key, ambitious aspects will be adopted in science classrooms. To foreshadow our results, there are, unfortunately, many ways in which the aspirational level and kind of thinking will not be met in many science classrooms; helping more students attain this aspiration will require different supports for different ways in which aspirations are not met.

Curriculum materials have historically been viewed as the main vehicle for infusing new ideas about teaching and learning in order to affect large-scale, instructional reforms (Brown 2009; Brown and Edelson 2003; Powell and Anderson 2002; Stein and Kim 2009; Weiss 1987). Schneider, Krajcik, and Blumenfeld (2005) advocated that well-designed, reform-based curriculum materials in science can be used to anchor discussions about reform efforts and as tools to guide initial attempts in the classrooms. Not surprisingly, with the release of the NGSS, there have been many efforts across the USA to develop science curricula aligned with NGSS vision

(e.g., Amplify Science, FOSS Next Generation, STC 3rd ed.). However, prior research suggests that even though high-quality curricular materials will be a significant step toward the realization of the new vision of the instructional reforms, they do not guarantee rigorous student learning in the classroom (e.g., Harris et al. 2015; Penuel et al. 2011; Schneider et al. 2005). These studies caution us to pay close attention to what happens inside classrooms during the implementation of curriculum materials that are aligned with any ambitious reform vision.

The main purpose of this paper is to describe different forms of student thinking that occurs when students are invited to think and reason as demanded by NGSS-aligned curricular tasks. We do this with reference to the *kind* (*integrated or isolated*) and *level* (*high or low cognitive demand*) of student thinking and reasoning during the implementation of a reform-based biology curriculum.

Theoretical Framework

Characterizing Thinking Demands of Science Tasks

Doyle (1983) suggested viewing curriculum as a collection of *tasks* to which students are exposed in the classroom. He defined tasks as the products students create, the operations that are used to generate the product, and the resources available to students while they are generating the product (p. 161). Tasks not only shape the substance of what students learn but also how students think about and make sense of the subject matter (Doyle 1983; Stein et al. 1996). Therefore, the nature of curricular and instructional tasks in which students are invited to engage in science classrooms shapes the kind of opportunities that they have to think about and engage in science content and practices (Tekkumru-Kisa et al. 2017). We focus on tasks as the basic instructional unit in classrooms (Blumenfeld et al. 1991) to make sense of the opportunities that students have to learn science.

Adapting from the work of Stein and colleagues (1996) in mathematics education and using Doyle's (1983) original definition, we define science instructional tasks as classroom based activities, the purpose of which is to focus students' attention on particular scientific ideas and/or practices (Tekkumru-Kisa et al. 2015). Science tasks vary in terms of their length; while some tasks can last several class periods, some lessons can be divided into several small tasks. Moreover, while some tasks engage students at a surface level, others can engage them at deeper levels through intensive thinking and reasoning. Like others (e.g., Blumenfeld 1992; Boston and Wolf 2004; Doyle 1983; Stein et al. 1996), we believe that a close examination of tasks provides a window into the kinds of disciplinary ideas and practices to which students are exposed and the types of opportunities they have to engage with these ideas and disciplinary practices.

The Task Analysis Guide in Science (TAGS) is a two-dimensional framework to analyze the level (cognitive demand) and kind (integrated or isolated) of student thinking (Tekkumru-Kisa et al. 2015). The integration/isolation dimension identifies whether or not science content and scientific practices are integrated within a task. Isolated tasks focus students' thinking exclusively on scientific concepts such as forces and motion or exclusively on scientific practices such as modeling. Integrated tasks are characterized by the integration of science content and scientific practices, so they require students to develop an understanding of science ideas and concepts within the context of scientific practices, as emphasized in the NGSS.

The second dimension of the TAGS is the *cognitive demand*, which is defined as the level of thinking (i.e., the nature of reasoning) required of students to complete a particular activity (Stein et al. 2000; Doyle 1988), and is related to Webb's "depth of knowledge" (Webb 1997). High-level tasks provide substantive opportunities for student thinking; they require students to make sense of scientific ideas and/or how science works. They often demand self-monitoring and self-regulation of one's own cognitive processes, including deliberate planning for and adjustments to the learning process in order to make progress on the task (Anderson and Nashon 2007; Stein et al. 2000). Low-level tasks provide minimal opportunities for students' thinking by either requiring them to reproduce previously known information or to follow scripted procedures that guarantee arriving at the correct answer. Previous work on tasks (e.g., Blumenfeld et al. 1991; Doyle 1983; Stein et al. 1996) suggests that a dominance of low-level tasks contributes to students' lack of understanding, and often proposes the remedy of including more high-level tasks.

In the original conceptualization of the NRC (2012) Framework for K-12 Science Education, integration of content and practices was likely viewed as embedded in a high-level task by definition. However, students can be given highly detailed scripts to follow that position them to engage in disciplinary practices within the context of a scientific idea (e.g., analyzing results of an experiment), but in a surface-level manner because all of the thinking demands have been removed (Mehalik et al. 2008; Tekkumru-Kisa et al. 2015). Similarly, while a science lab can require students to collect and analyze data or control variables, due to the highly scripted structure of most labs, students may miss what is at the heart of science, which is knowledge building (Duncan and Cavera 2015). Further, tasks, which—on paper—appear to require student thinking can decline in various ways inside the classroom.

Change in Thinking Demands Across the Phases of a Task

Research has revealed that teachers and students face a host of challenges when cognitively complex tasks are used in both science and mathematics classrooms (e.g., Henningsen and Stein 1997; Jones and Eick 2007; Krajcik et al. 1998; Schneider et al. 2005; Winne and Marx 1982). Students often experience novel, cognitively complex tasks as having high levels of ambiguity about what to do and how to do it, necessitating students' decision-making, and having higher possibility of risk that answers will be incorrect (Doyle 1988). Because of the unpredictable nature of the solution process or open-ended nature of developing a causal explanation for a phenomenon, implementing many high-level tasks often involve some level of anxiety and uncertainty, which may interfere with students' productive engagement with the disciplinary ideas and practices embedded in the task. As Blumenfeld et al. (1991) claimed, we think that many inquiry-oriented, project-based activities "are complex and inherently ambiguous and risky (see Doyle 1983)" (p. 380). They raised a caution that as the task gets more demanding, students may focus more on getting the task done rather than engaging in it productively for sense-making. In fact, Krajcik et al. (1998) found that some of the seventh grade students in their study focused on short-term, quick solutions and procedures to complete their work as they were designing and carrying out their investigations. The authors underscored the role of teacher's suggestions and questions in encouraging students to focus on the substantive aspects of their investigations.

Implementing inquiry-based activities was found to be challenging for teachers as well. For example, engaging unmotivated students becomes a challenge for some teachers. One of the teachers in a study by Trautmann, MaKinster, and Avery (2004) commented on her experience

with inquiry-based teaching: “What I find probably even more with the high school kids is that they are kinda like, ‘Just tell us what you want us to know,’ so it’s a struggle” (p.10).

Moreover, many teachers were reluctant to provide autonomy (or needed a lot of support to learn to transfer responsibility to their students, Nashon et al. 2015) because they often believed that their students were not equipped to do the task on their own (Marx et al. 1997). Other challenges associated with using such tasks include allocating enough time for in-depth exploration of ideas, classroom management for maintaining productive independent work, and being able to provide appropriate amounts of scaffolding (e.g., Blumenfeld et al. 1991; Marx et al. 1997). By investigating four middle school teachers’ implementation of an inquiry-based science unit, Schneider et al.’s (2005) findings uncovered the difficulty of giving students opportunities to explore their own ideas through inquiry while ensuring that they are provided with enough support to guide their thinking.

Here, we argue that the challenges experienced by the teachers and students in the above studies will continue to persist in the NGSS Era. In fact, things are even more complex. Crawford (2014) recently pointed out the difference between teaching inquiry in earlier reform initiatives and the focus on scientific practices in the K-12 Framework (NRC 2012) and NGSS. She argued that these new documents (the K-12 Framework and NGSS) underscore a shift from having students simply develop and test hypothesis to “testing and revising theoretically grounded models”. This would require from students more than “experiencing inquiry” because they will be expected to interpret and evaluate data to develop arguments, explanations and models (p. 523). This difference can be interpreted as an even more concerted focus in the current reform documents to place higher demands on students’ thinking processes to make sense of scientific ideas and meaningfully engage in scientific practices.

All of these challenges raised in the literature lead us to conjecture that the high-level thinking demands of many complex, NGSS-aligned curricular tasks will be transformed by students and teachers, often unwittingly, once they are unleashed in science classrooms. Prior research in mathematics education offers a useful analytical lens for considering the nature of change in the cognitive demand of tasks as designed and implemented. One of the key findings from research studies on the implementation of mathematical tasks is that the level of thinking placed on students through the selection of cognitively challenging tasks can decline during their implementation in real classroom settings (Arbaugh and Brown 2005; Henningsen and Stein 1997; Stein et al. 1996). According to the conceptual framework proposed by Stein and colleagues (1996), the level of thinking required of students in order to successfully engage with a task can change across three phases: (1) tasks as they appear in curricular materials, (2) tasks as *set-up* by the teacher in the classroom, and (3) tasks as *enacted* by the teacher and students, and both teacher and student related factors can influence this change such as student self-monitoring, inappropriateness of the task for the students’ background and prior knowledge, amount of scaffolding, and degree of sustained pressure for explanation and meaning (Henningsen and Stein 1997; Stein et al. 1996).

The first box in Fig. 1 represents the demand of tasks as designed (as it appears in the curriculum materials). The second box is the demand in set-up phase, which indicates the potential kind and level of thinking, which is placed on students based on the way the teacher launches the task for students. Jackson and colleagues’ recent research has emphasized the role of the set-up phase in mathematics classrooms. They found that cognitive demand of the tasks was lowered in the set-up in more than half of the lessons that they observed and that the quality of set up appears to be related to students’ opportunities to learn during the whole class discussions (Jackson et al. 2013, 2012). Building on earlier studies (e.g., Kang et al. 2016;

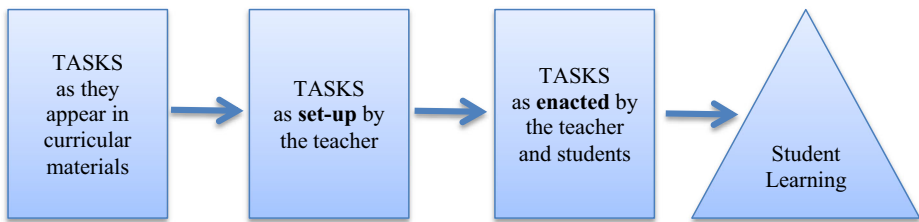


Fig. 1 The Mathematics and Science Tasks Framework (Stein and Smith 1998). Adapted with permission from *Mathematics Teaching in the Middle School*, copyright 1998, by the National Council of Teachers of Mathematics. All rights reserved.

Hammer et al. 2005), we argue that set-up phase can play a critical role in science classrooms for framing the work that students are going to engage in (for example, framing the work as “figuring out” how and why it works rather than “learning about” the scientific idea; Reiser 2015). The third box indicates the demand during the enactment phase, “the manner in which students actually work on the task” (Stein et al. 1996, p.460). This final phase—task enactment—is particularly important because it identifies how students are actually thinking and reasoning during their work in the classroom. There has been strong evidence in the mathematics education literature that students learn best when they are in classrooms in which a high level of cognitive demand is maintained throughout their enactment (Boaler and Staples 2008; Hiebert and Wearne 1993; Stein and Lane 1996; Stigler and Hilbert 2004; Tarr et al. 2008).

The Math Task Framework has recently been taken up by other science education researchers. Kang et al. (2016) analyzed 57 science lessons taught by 19 first year teachers by focusing on the tasks as designed, launched by the teacher and implemented by the teacher and students, along the dimension of cognitive demand. Their findings were consistent with prior research (Stein et al. 1996) in that they identified three trajectories for students’ intellectual demand across the phases of a lesson as designed, launched, and implemented: (i) lessons stayed at high intellectual demand across the phases (approximately 25%), (ii) lessons starting with intellectually demanding tasks but failing to maintain its rigor (approximately 20%), and (iii) lessons starting at a low level and staying at a low intellectual demand (approximately 55%). They did not address the role of integration (e.g., whether task demands sometimes decline in no longer being integrated, in other words, without requiring students to engage in scientific practices) and the factors that can play a role in maintaining or declining demand on students’ thinking.

In this paper, we examine lessons that begin with demanding tasks. We argue that recent research-based instructional reforms require us to attend to more than just cognitive demand, but also to the dimension of integration of science content and scientific practices in students’ engagement with science. For example, a task that asks students to analyze data without any consideration of scientific ideas fundamentally fails to meet the NGSS vision even if it may be cognitively demanding for students. Alternatively, if a teacher replaces a cognitively demanding task involving science content but no science practices (e.g., evaluating the accuracy of science in a movie) with an integrated content and practices task that is not cognitively demanding (e.g., following a teacher’s step-by-step instructions for a science investigation) only superficially meets this vision.

Our study provides a language for characterizing how the kind and level of thinking demanded by cognitively complex tasks can be transformed in science classrooms by focusing on both: (i) the extent to which students are positioned to engage in scientific practices as they try to make sense of disciplinary ideas and (ii) the level of demand on student thinking. This common two-dimensional language for researchers and practitioners (e.g., science teachers,

coaches, etc.) can be used to communicate to one another regarding how tasks unfold in science classrooms. We argue that if we are able to identify the potential kinds of transformations that high-level curricular tasks are susceptible to undergoing in science classrooms, the field can proactively propose detection of and solutions for supporting more effective implementation of these high-level tasks for both teachers and students.

Closer Look into Curriculum Implementation Through a Task-Based Framework

In this study, our guiding question was: What is the level and kind of thinking in science classrooms that are using cognitively demanding curriculum materials that require students to engage in scientific practices to develop a deeper understanding of disciplinary ideas? To answer this question, we use a simplified version of the TAGS (see Fig. 2).

The quadrants within the matrix allowed us to characterize and label the level and kind of thinking demanded of students through science tasks as Integrated-High, Isolated-High, Isolated-Low, and Isolated-Low. Students' thinking in Quadrant-I (*IntHigh*) is characterized by scientific sensemaking. Students are positioned to think like a scientist as they develop "a deeper and broader understanding of what we know and how we know it" (Osborne 2014, p. 183). In other words, while deepening their understanding of scientific theories, principles, and ideas, students can also develop an understanding of how scientific knowledge develops often as they try to explain a phenomenon.

Student thinking in Quadrant-II (*IsoHigh*) is characterized by reasoning about either primarily core scientific concepts (e.g., force and motion) or primarily about the scientific practices (e.g., argumentation or modeling). Students are positioned to engage in high-level cognitive processes such as finding relations, analyzing information, and generalizing to a broader idea, so engage in the tasks for understanding. However, these tasks do *not* position students to understand both the pursuit (i.e., seeking for an explanation for a phenomenon) and the body of knowledge that results from that pursuit (Levin et al. 2013).

Like in Quadrant-II, in Quadrant-III (*IsoLow*) students' thinking focused either primarily on the scientific ideas or the definitions/procedures related to the scientific practices. However, as opposed to Quadrant-II, Quadrant-III is about low-level student thinking which is characterized by students' memorizing information, facts, rules, and definitions, or by following a set of scripted procedures without knowing what they are doing and why.

Finally, students' thinking in Quadrant-IV (*IntLow*) is characterized by superficial engagement about the scientific ideas within the context of some scientific practices. Even though the task involves the integration of content and practices, it fails to engage students in the kind of reasoning processes employed in real scientific inquiry. All in all, these four

Fig. 2 The simplified TAGS matrix for characterizing level and type of student thinking

		Integration of Content & Practices	
		Isolated	Integrated
Cognitive Demand	High	Quadrant-II Iso-High	Quadrant-I Int-High
	Low	Quadrant-III Iso-Low	Quadrant-IV Int-Low

quadrants can help us to categorize the kind and level of thinking that students engage in science classrooms, and also identify the patterns of change across the phases of a task (shown in Fig. 1).

Methods

Background

The study was situated within a National Science Foundation-funded project that focused on the design and implementation of biology units integrating mathematics into big ideas in biology. We focused on the final round of implementation of one of these units.

Curriculum The curriculum was a 4-week unit focusing on the major rules of Mendelian Inheritance that has been found to significantly improve student conceptual understanding and quantitative problem solving (Schuchardt and Schunn 2016). The unit primary storyline involved a design challenge which required students to help an imaginary local zoo to develop a plan to breed rare geckos. The tasks in the unit help students explore how genetic information is inherited and expressed so that they can solve the design challenge.

Focal Task We focused on one task observed across different kinds of biology classrooms with different teachers and population of students, since many teachers are concerned whether all students can actually engage in high-level tasks. In this study, we selected a particular high-level integrated task from the larger unit because it successfully integrates science content and practices at a high cognitive demand level, and because it was focused on science (vs. other tasks in the unit which also focused on mathematics).

In the task (henceforth called the PCR task), students are introduced to a technique called polymerase chain reaction (PCR), which is used to examine what is happening at the DNA level. PCR results depicted in the task show two separate crosses of a male and a female gecko and their offspring. Without being introduced to the rules of Mendelian inheritance or any information about genotype, students are asked to generate the rules of inheritance (i.e., how the genetic information is transferred from parent to offspring) by analyzing the PCR results.

The task is composed of three major parts. In the first part, students are provided with a pictorial representation to situate the levels of organization, highlighting the DNA level as the level of current focus. Then, they are asked to make observations of the PCR results of two separate crosses of the same male with two different female geckos and their offspring. This part of the task positions students to distinguish between observation and inference (part of the more general NGSS practice of analyzing and interpreting data), and at the same time, begin to think about how genes are passed from parents to offspring (the science content). In the second part, students are required to look for patterns in the data and try to make generalized rules (the practice of constructing explanations) about how offspring inherit genetic information from both parents (the science content). Ultimately, from this task, students are expected to derive the rules of Mendelian inheritance: (i) Each organism has two versions (alleles) of each gene; (ii) one version of each gene comes from the male parent and one comes from the female parent; and (iii) offspring can only get what the parents have to give. Once they have determined these general rules, the final part of the task involves mapping the scientific terminology such as alleles, homozygous, heterozygous to their general rules.

The PCR task is categorized as Integrated-High in our 2×2 matrix because it provides the opportunity for students to meaningfully engage in scientific practices to develop new understandings about how genetic information is passed from parents to offspring. The task invites students to observe, analyze and interpret PCR data to uncover patterns and rules that govern inheritance, to collaboratively resolve interpretive disputes about the patterns they see in the data, and also to justify their reasoning and to self-assess whether their argument for how offspring inherit their genetic material from parents is accurate. The final phase asks students to attach scientific terms to ideas that they have already developed, thereby avoiding the common problem in scientific classrooms of memorizing terms for which one has limited, if any, understanding. Considered together, this task has the potential to teach genetic inheritance in a meaningful way that ultimately connects to the scientific canon.

Data Collection Contexts

The main data source associated with this particular study was the video records of implementation of the PCR task in five biology classrooms. In each of these classrooms, the implementation took from 1½ to 3 40-min class periods. In total, we analyzed about 450 min of video-recorded lessons, which is approximately 11 40-min class periods. Table 1 provides details about the five biology classrooms on which we focused. All of the teachers volunteered to participate in the broader study, which incentivized their participation. As part of that study, they were expected to implement the entire unit and they received professional development for doing so.

As shown in Table 1, we focused on a broad range of science classrooms with students that differ in their achievement levels (from honors biology for gifted students to regular track biology for primarily low achievement students) and SES status (from classrooms with very few students qualifying for lunch subsidies to classrooms in which most students qualify for lunch subsidies). Note that there were also classrooms with a more diverse mixture of student abilities. Since cognitive demand of tasks can decline because of teachers' expectations for what students can (or cannot) do, it is useful to observe the same task being implemented across these diverse contexts. Note that, as is common in the USA, student achievement and SES status are highly correlated in this sample. Our goal is not to tease apart the separate effects of the contextual variations, but rather to show robustness of the proposed framework about the level and kinds of student thinking across a diversity of achievement levels. Further, to foreshadow our results, we think it is important to gather supporting evidence that *all*

Table 1 Details about the video-recorded classrooms

Teacher	Classroom type	Grade	Achievement level ^a	Free and reduced lunch ^b
Ms. Ford	Honors/Gifted Bio	9	Mostly high	17%
Mr. Clark	Honors Bio	9	Mostly high	27%
Ms. Allen	Honors Bio	10	Mostly average	53%
Ms. Allen	Regular Bio	10	Mix of levels	53%
Ms. Neal	Regular Bio	9	Mostly low	78%

^a Reported by the teacher by selecting one of the following: (i) Mostly low achievers, (ii) Mostly average achievers, (iii) Mostly high-achievers, (iv) A mixture of levels

^b School-level data

classrooms (not just affluent, high achieving classrooms) can maintain task rigor, and so, deserve science teaching that places high demands on student thinking.

Data Analysis

Two researchers independently analyzed all the video recorded lessons for the PCR task by using a coding protocol based on the TAGS. The protocol required raters to identify the set-up and enactment phases during the entire implementation of the PCR task. Because the PCR task was composed of three parts, which took several class periods, raters observed several set-up and enactment intervals during the entire implementation of the task. Each time the teacher introduced a different part of the PCR task for students to work on, it was marked as set-up. In the protocol, the raters were also asked to place each set-up and enactment into one of the categories within the TAGS framework and justify it by providing evidence from the video.

Next, the two raters discussed their analysis to come to consensus for the set-up and enactment times and ratings. When there was uncertainty and/or no consensus between the two raters, a third researcher watched those sections of the videos. The three researchers discussed their ratings to come to a consensus for the rating.

Results

Our analysis revealed that the NGSS vision was transformed in several ways once the PCR task was launched in the classrooms. Table 2 provides a summary of ratings for demand placed on student thinking as the PCR task was enacted in each classroom. Because the last part of the PCR task mainly required attaching scientific terms to the rules that were derived and the major intellectual work happened in the first two parts of the task, we focused our findings on the first two parts of the task.

In all classrooms, there were two set-up phases each of which was followed by an enactment phase. Each set-up can be mapped onto each part of the PCR task that we described above. As described earlier, the PCR task was categorized as IntHigh because it has the potential of engaging students in higher levels of thinking about genetics and a set of scientific practices. Once the task was unleashed in these classrooms, changes in the level and kind of student thinking were identified. In all cases, the teachers set up the IntHigh task as IntHigh, following the detailed teacher notes they received. However, there were multiple forms of decline observed during enactment. In what follows, we are going to identify these changes in the enactment phase of the PCR task by characterizing the level and kind of student thinking

Table 2 Change in the level and kind of student thinking across the phases of a task

Class	Classroom type	Enactment	
		Task part 1	Task part 2
Ms. Ford	Honors/Gifted Bio	Integrated-High	Integrated-High
Mr. Clark	Honors Bio	Isolated-High	Integrated-High
Ms. Allen	Honors Bio	Integrated-High	Integrated-High
Ms. Allen	Regular Bio	Integrated-High	UNE
Ms. Neal	Regular Bio	NTR	<i>Missing data</i>

into one of the quadrants in Fig. 2. We begin with a description of two classrooms in which the level and kind of thinking was maintained across the set-up and enactment phases to show what is possible with this task. Then, we turn to the forms of decline observed in different phases. Even though our goal in this paper is not to comprehensively identify the factors that caused the maintenance or decline in cognitive demand, in our exploration of the cases, we uncovered some factors that seemed to play a role in the level and kind of thinking observed in these classrooms. These factors are acknowledged as appropriate and available.

Maintaining Thinking Demands of High-Level Science Tasks in Set-up and Enactment

As shown in Table 2, in four out of five classrooms, maintenance in thinking demands of the PCR task was observed both in set-up and enactment phases during at least one part of the task. In those classrooms, the PCR task, which has the potential to engage students in scientific sense making, was used as intended both in the set-up and enactment phases. In other words, students in these classrooms were trying to develop an explanation for how genetic information is transferred from parents to offspring by using the PCR outputs. In what follows, the process of maintaining the kind and level of thinking demanded by the task is presented using two illustrative cases, drawn from diverse classrooms.

The Nature of Student Thinking in Ms. Ford's Classroom

Ms. Ford allocated one full class period (approximately 40 min) for the implementation of the first part of the task during which students were introduced to the PCR data and began to analyze it and make some inferences about how genes are passed from parents to offspring.

Setting Up Tasks as Part of a Larger Unit/Idea Ms. Ford launched the task by making an explicit connection to where they left off in the prior lesson (i.e., the need to look at what is “inside” the geckos that might be playing a role in what the gecko physically looks like). She said, “[today] we are going to learn more about the “hidden factors” (referring to *genes* that they decided to explore in order to understand how genotype might be influencing phenotype of the offspring). This might have helped students to see that there are connections across the ideas in which they engage in different lessons and each lesson does not stand alone around a disconnected fact. She launched the task by asking students to make observations of the PCR outputs; she warned them not to jump to making inferences but instead record what they saw. Students worked in groups for a while as she walked around monitoring students’ conversations about their observations.

She then brought them together for a whole group discussion. As students shared their observations, she typed them on the computer and projected them so that others could also see what she was typing. When the observations were not clearly stated by the students, she asked for clarification; likewise, students were making suggestions to revise the wording of the observations typed by Ms. Ford to better represent what was said. At the end of this first lesson on the PCR task, they had a list of observations that they co-constructed as a whole class. Their list included observations such as: (1) In cross #2, if there is a thick line in a parent, then the offspring have a thin line; (2) Each gecko has two thin lines total-either one thick or two thin; (3) If there is a thick line, then there is no thin lines; if there is a thin line, there is another thin line; (4) None of the offspring has thick line in the bottom row; (5) Five of the 11 offspring

from cross 1 have the same grouping of the lines as F1. Ms. Ford concluded the lesson by telling students that she would organize this list and give it to them the following day and they would continue from there.

Pressing for Evidence-Based Reasoning The next lesson focused on deriving the rules of Mendelian Inheritance based on the list of observations about the PCR data. The way in which Ms. Ford introduced the second part of the PCR task in the beginning of this second lesson was not video-recorded but could be easily envisioned because of the way she ended the previous lesson. When the video-recorded part of this second lesson started, students were talking about the class list of observations that Ms. Ford printed for them. This list appeared to encourage them to support their claims with evidence as they were trying to develop a set of rules based on these observations. Throughout the lesson, Ms. Allen continued to press students to use this list as they were trying to reason about the patterns in the PCR data. Such evidence-based reasoning is expected in the development of a sound argument (Toulmin 1958; Toulmin et al. 1979).

Attending to and Advancing Students' Ideas Students were confidently working in small groups to generate the rules. Ms. Ford circulated around the room *monitoring* students' ideas in each group's conversation to ensure they were on track. Ms. Ford listened to a group of students who were considering three of the observations in the class list and trying to decide whether they could collapse these observations to make one rule. Although they thought that these three observations worked in concert to describe a pattern they saw regarding the number of lines each offspring must have, they were not sure how to word it. Ms. Ford helped them to clarify what they were seeing in the data by asking "Is there some way you can summarize those [pieces of evidence] together into one general rule?" One student struggled to make the rule general but her partner said "there has to be two lines no matter what, if it's thick or thin there has to be two." Ms. Ford said "Ok, I would say that is a general rule, each gecko must have two lines." The students confirmed that was what they meant and wrote the rule. It is important to note that it was only after the students had articulated their thinking that Ms. Ford rephrased what they had said. Ms. Ford continued by asking whether the evidence they had identified supported that as a general rule. The students replied "yeah." This reinforced their evidence-based reasoning. Even though students had yet not used the right terminology, they could identify one of the rules of Mendelian Inheritance: Each organism has two versions (alleles) of each gene. Therefore, while paying close attention to students' ideas during their interpretation of the data, Ms. Ford was at the same time gently moving their ideas through questioning toward the main scientific idea that she wanted them to develop in the PCR task.

Ms. Ford continued to press for students' evidence-based reasoning and to pay close attention to students' ideas as she was visiting other groups. She went to another group and asked, "What kind of inference did you make?" One student responded "Offspring get one line from mom and one line from dad" but then questioned whether that holds for all cases because of the thick line. Ms. Ford reminded them that "Based on the class observations, we said that the thicker line is 2 so I think that is a safe observation at this point." So, while she modeled evidence based reasoning for the students, she also clarified a point for them so that they could make progress in their thinking. Based on this confirmation, the student tested this thinking on the thick bands in one of the crosses to convince himself that his rule worked. Based on the student's test, the group was convinced that "Offspring gets one line from mom and one line

from dad”, which is, in fact, another rule of Mendelian Inheritance that they derived: One version of each gene comes from the male parent and one comes from the female parent.

After visiting several other small groups, Ms. Ford convened the whole group discussion and began typing the rules that students shared and supported with evidence. Once they surfaced the first two rules another student volunteered a rule “if at least one parent has a band in the upper position, its offspring will have a band in the upper position.” Another student volunteered the opposite of this rule for the bottom position. Ms. Ford drew their attention to the two contributions, and asked them to think of a way to combine the two contributions, “What can we say about the how the offspring gets their lines?” The students floundered a bit but said that what the offspring gets depends on their parents. Pleased with this statement, Ms. Ford asked, “Are we saying that offspring can only get what their parents have to give them?” Student replied, “Yes”. Ms. Ford rephrased and wrote “Offspring only get the bands their parents have to give” Ms. Ford informed the students that they had identified all of the rules and asked whether they agree with these three rules based on the data. Students agreed. By the end of this whole class discussion, they had the three rules of Mendelian Inheritance constructed and listed on the board. The rules did not have the right terminology but carried the meaning of how genetic information is passed from parents to offspring. What was missing was attaching the right terms to the rules (i.e., replacing “bands” with “alleles”), which was completed in the final part of the task.

All in all, by interpreting their observations into generalized statements, the students were able to determine for themselves whether they captured rules that met these requirements. Although the teacher provided a number of scaffolds, the teacher did not take away the cognitive demand and the students did the thinking to come up with the rules of inheritance. They were actively trying to make sense of a phenomenon as they were engaging in the practices of the discipline to develop an explanation of how genetic information is passed from parents to offspring as presented in the rules of Mendelian Inheritance. As such, we coded the level and kind of student thinking during this enactment as IntHigh. Students were active scientific sense makers. The teacher was working as hard as the students by carefully monitoring what students were saying and how they were thinking. She was constantly asking for justification and clarification, and helping them make connections between what they were thinking and the larger conceptual ideas embedded in the PCR task.

The Nature of Student Thinking in Ms. Allen's Regular Classroom

Ms. Allen launched the first part of the PCR task in the beginning of a lesson without deviating from the instructions on the worksheet and by maintaining the thinking demand of the task. She first showed the levels of organization from the organism level to the DNA level. She asked what is found inside the nucleus and elicited students' responses, which included genes and DNA. She then reiterated and marked on the picture that in this task, they were going to focus on the genes inside the DNA.

She, then, asked students whether they remembered the difference between observation and inference. Students shared what they remembered and she revoiced what they said and clarified the difference between making an observation and making an inference. She told them to make observations and record what they see even it was obvious. She also told them that they do not need to make any inferences yet. Even if the way she launched the PCR task was not as robust as it was in Ms. Ford's classroom, she did not deviate much from what the task demanded. The task requires “figuring out” how genetic information is

passed from parents to offspring. Since Ms. Allen did not deviate much from the task, it was introduced to the students as figuring out how it works rather than learning about the scientific idea (Reiser 2014).

Attending to Students' Ideas and Engagement Ms. Allen gave about 15 min for students to work in groups to make observations of the data presented in the PCR outputs and record them. The enactment phase involved small group work followed by a whole class discussion. She walked around the groups and encouraged them to make observations. She frequently repeated the same question, "What do you see?" Apparently, she was trying to understand what students were seeing in the PCR data and how they were interpreting it. For example, she sat next to one group and asked students several questions about what they saw to get them started. She asked questions such as "What about the Female?" Students started to share some observations. She accepted the things that they said and left the group by saying, "Ok, good, you got started."

A student from another group said that they came up with four observations; Ms. Allen encouraged them to come up with at least one more. By setting the goal of coming up with five observations, students were motivated to trying to identify many patterns in the PCR data. In one group, as students shared their observations, she encouraged them to compare and contrast the two crosses and the kinds of bands that appeared on the worksheet. By doing that, she was trying to focus their attention on the patterns in the data that could help them to make some inferences about the rules of Mendelian Inheritance, which is what they were going to be asked to do in the second part of the task. Different from Ms. Ford, Ms. Allen was not only attending to students' ideas but also trying to facilitate their engagement. Ms. Allen was working with a group of students, who were not seem to be motivated to engage in any intellectual work on the day of this class observation. While Ms. Allen was pressing on their thinking, she was also using different strategies to facilitate their engagement in the task.

When they gathered as a whole group, Ms. Allen asked students to share the observations that they came up with and recorded them on the smartboard. She accepted all the observations without evaluating any of them. Some of the observations were very superficial in that they could not help students to derive the rules of inheritance, such as "cross-1 has 11 offspring; cross 2 has 12 offspring." Even though Ms. Allen could have prompted students to say a little more about their observations to get them started on making some inferences about the patterns so that they would be able to come up with the rules of inheritance more easily in the second part of the task, she did not chose to do that. This might have caused some of the students to struggle too much when they were asked to come up with rules by using their observation when they moved to the second part of the task.

Toward the end of the whole group discussion, one student had difficulty articulating what he saw. Ms. Allen invited him to come to the board and explain what he wanted to say by showing it on the pictures of the crosses that she projected on the board. What students wanted to say was, when there was a thick band under one of the offspring, there was not a thin band below it. Ms. Allen asked him what he said originally for the thick band. The student did not appear to understand what Ms. Allen was asking. Then, Ms. Allen said, "you said, two thin bands". Then, another student said, "Ohh! Two thin things are put into one" Then, the teacher summarized what both of these students said, "So are we looking at one thick and two thins or are we looking at all thin bands come together look like this? That's something for you to decide as we are moving on" and moved to the second part of the task. When Ms. Allen attended to students' ideas, she seemed to be selective in which student ideas she selected. While selecting certain ideas to be shared in the during the whole group discussion can be a

strategic instructional practice to facilitate productive classroom discourse (Smith and Stein 2011), selective attention to particular student ideas that are closer to the canon or the overarching goal of the lesson, also has the risk of declining the demand on students' thinking by not being responsive to all students' thinking.

This thin line on which Ms. Allen had to walk might be related to the context of her classroom. Ms. Allen has reported that this classroom consisted of students with mix of achievement levels. Majority of the students in this class did not appear to be as engaged as the students in Ms. Ford's classroom. The nature of classroom composition and culture might have added an extra layer to the complexity of implementing cognitively demanding science tasks. She needed to do more to get students started with analyzing the PCR data.

All in all, Ms. Allen started the task by situating students' observations into the DNA level and then let students to closely look at the PCR outputs. During the small group work, since she did not encourage students to make inferences, their observations might have stayed a bit disjointed from the big idea of trying to figure out how genes are passed on from parents to offspring. However, we decided that she left the observation work in this first part of the task to the students and facilitated them to think about the data in more detailed ways. That is why we decided that the thinking demand was maintained at IntHigh during the set-up and enactment phases of the first part of the PCR task, even though there were different ways through which the demand on students' thinking could have been maintained more productively.

Maintaining Thinking Demands of High-Level Science Tasks in Set-up but Decline During Enactment

Our analyses also revealed instances where thinking demands of the PCR task were maintained during the set-up phase but not during the enactment. In all these classrooms, the task was launched in a way that students were invited to work on the task as scientific sense makers. Both Mr. Clark and Ms. Allen in her regular classroom (the one explained above) followed the instructions on the handout to introduce the task to the students. Since the instructions on the handout were written to position students to engage in productive thinking about the PCR data, we decided that by following the instructions on the handout as written—and not adding more direction or modeling how to attack the problem—that the teachers did not decline the thinking demand of the task. These instructions were enough to get students started to work on the task as intended.

Ms. Neal took a different approach as she launched the PCR task, particularly the first part of the task during which she maintained the demand of the task similar to Mr. Clark and Ms. Allen. She provided more scaffolding than what we observed in the other two classrooms, however. For example, she walked students through the key that was provided on the handout—which she referred to as “doing some pre-observations”—to familiarize them with the language used in the task.

After these set-ups where the thinking demand of the task was maintained, we observed a decline in the way students engaged in the ideas once they started to work on the task. The level and kind of thinking was categorized as IsoLow in Mr. Clark's classroom, unsystematic and nonproductive exploration in Ms. Allen's regular classroom and no conceptual thinking occurred in Ms. Neal's classroom. These cases illustrate that selecting and setting up cognitively complex tasks are not enough to engage students in high levels of thinking and reasoning in science. In what follows, we will explain each of these declines in more detail.

Losing Integration: Decline to Isolated-High

As mentioned earlier, Mr. Clark, for most of the time, adhered to the instructions on the handout while setting up the work for the students.

Clarifying Expectations One distinctive feature of Mr. Clark's launch of the task was the way he clarified what students were going to work on in this task. He began by reading the initial passages of the worksheet. He also projected a depiction of levels of organization from the organism level to the DNA level on the SmartBoard. After a brief discussion that situated the day's work at the DNA level, Mr. Clark read the remainder of the worksheet including the task directions for what students were supposed to do: Make observations about the PCR data associated with two different gecko crosses. Just before students started to work, he asked them to pull from their prior discussion in science class what their understanding was of the meaning of observation. Therefore, his launch positioned students to begin to analyze the PCR data of the two separate crosses of a male and a female gecko. By the end of the launch, it was clear that students knew what they were going to work on that day but how they will do it was left to the students, rather than proceduralized by the teacher.

As students were settling in and beginning to share their observations with their partners, Mr. Clark walked around the room, listening in on their conversations and observing what students were writing on their worksheets. Many of the students were trying to make sense of the crosses in productive ways mostly by using what they already knew about genetics. Students' conversations involved comments like, "Can we say that mother's trait is recessive?" "All the offspring have the same trait."

Attending to and Advancing Students' Ideas As the work progressed, Mr. Clark began to comment on students' observations. He was *not* paying close attention to the substance of students' ideas about what they saw in the PCR output. Instead, the majority of his comments centered on whether each observation that students had written down was indeed an observation or inference. It seemed like he *framed* this part of the task as distinguishing between observation and inference (Russ and Luna 2013). He was not paying attention to students' ideas; instead, he was only attending to (or bring students attention to) the distinction between observation and inference. For example, one student said, "We are assuming these are recessive". Mr. Clark said, "If you are assuming something, is that an observation or something else?" The student said that it would be an assumption. The teacher recommended her to use the word "inference", instead. He, however, did not elicit the student's ideas about what she meant by "recessive" and why she thought so.

This pattern repeated itself several times as Mr. Clark monitored and commented on students' work. One pair of students was talking among themselves. One student said, "If this one was dominant, this one was recessive, you will get both because think about the Punnett Square". When Mr. Clark overheard this comment, he interrupted her by saying "No; remember we are making observations. Do you observe dominant and recessive on that chart?" Students accepted that they could not "observe" dominant and recessive on the PCR output and erased what they wrote on the worksheet and turned back to the task of making *only* observations. This indicated that he was not trying to make sense of students' ideas but instead was only evaluating the correctness of students' ideas (Levin and Richards 2011) in terms of whether students knew the distinction between observation and inference.

Emphasis on Correctness and Completeness of the Ideas Mr. Clark's attention to the accuracy of students' ideas continued during the whole class discussion. After approximately 11 min of working on the task in pairs, Mr. Clark asked students to volunteer to share their observations with the whole class. As students shared each observation, he "quizzed" them regarding whether their "observation" was truly an observation. For example, one of the students said that in cross-2 the offspring get a mix of male and female traits. A student at the front of the class wrote the following on a whiteboard (dictated by Mr. Clark): "Cross 2- offspring are a mix of parents." Mr. Clark asked the class whether this was an observation or inference. Through the teacher's prompting, the students conceded that the statement was an inference. Consequently, the teacher instructed the student scribe to cross-out the written statement.

All in all, throughout the enactment phase there was a tension between students' wanting to think and reason in ways unbridled by the constraint that they do not make inferences and Mr. Clark's insistence that they only make observations. Mr. Clark might have been trying to make sure that all the observations were explicitly stated so that in the next segment of the task, students could use these observations to justify the rules they would begin to derive. Because of that, he ended up *not* entertaining all of their contributions. By limiting students to only observations, he ended up obscuring the content and hindering students' efforts to think about genetics patterns that they could infer from the PCR results. Thus, although the students' thinking was clearly not free of content, the focus of their work was not on using scientific practices to make sense of scientific ideas as we see in IntHigh enactments. Too much emphasis on the distinction between observation and inference overshadowed students' trying to make sense of scientific ideas embedded in the task.

Getting Lost: Decline to Unsystematic and Non-productive Exploration

Like Mr. Clark, Ms. Allen, in her regular classroom, launched the task by adhering to the instructions on the handout. They had started the lesson by working on the first part of the PCR task. After they had made some observations and completed the first part of the task in about 25 min, Ms. Allen moved to the second part of the task. She read the instructions and told students that they would use their observations of the PCR data and develop some rules. She told them that the rules should apply to both of the PCR outputs that were provided in the task. By inviting students to look for patterns in the PCR data and to try to make generalized rules about how offspring inherit genetic information from both parents, Ms. Allen maintained the thinking demand of the task in the set-up.

Exploring Without an Overarching Goal The thinking demand that was set up by the task and teacher was not maintained during the enactment phase and changed to *unsystematic and nonproductive exploration*. In Ms. Allen's regular classroom, students were not making progress toward the goal of the task and the teacher was unable to advance students' understanding of ideas. Overall, students appeared to be unwilling to work on the task except for responding to Ms. Allen's questions when they were asked. Ms. Allen continued to ask questions to help the students derive the rules but her questions were framed as looking for one right answer, which then could lead to the defined set of rules that they should derive in that lesson. For example, the following

conversation occurred in one of the groups that Ms. Allen visited during the small group work:

Student: I came up with a rule; you said it's not right.

Teacher: All right, look here. What were your big point that you ...

Student: Two thins

Teacher: You think this one band here is actually one band but two thin bands

Student: Yes, would that be a rule?

Teacher: What would the rule be then?

Student: One thick band equals two thin bands. If it's a thick band, it has to be two thin bands.

Teacher: Here you go. What is it mean? Every single one of these has how many bands?

Student: One

Teacher: You said it –

Student: Two

Teacher: So, the thick ones are

Student: Two

Teacher: And those ones are

Student: Two. So, every cross has two bands

Teacher: Two bands.

As illustrated with this conversation, there was an attempt to come up with the rules of inheritance but students did not seem to be engaging in any sense making. They could identify the pattern (that there were two lines for each organism in the PCR data) but they did not seem to link it with the larger conceptual idea. The teacher made an effort to position students to derive the rules with her questions but her questions were closed ended looking for one right answer. Overall, the main point of the task seemed to be lost in this lesson. The end goal was to come up with rules. At a surface level, that is what they were doing in this lesson. However, students seemed to miss what these rules were for and why they were trying to derive these rules apart from the reason that they were told to do so by their teacher. All in all, students' exploration was around the edges of the main scientific idea and they could not make progress toward an explanation for the phenomenon that they were exploring. That is why we decided to categorize the kind and level of thinking during this enactment as "unsystematic and nonproductive exploration." Henningsen and Stein (1997) defined this pattern as, "students' thinking processes characterized by unsystematic exploration and lack of sustained progress in developing meaning and understanding" (p. 535).

Shutting Down: Decline to No Thinking Happening

Ms. Neal's class had classroom management issues and a lot of unmotivated students, who appeared to be unwilling to engage in the lesson.

Clarifying Expectations Ms. Neal started to introduce the PCR task by reading the instructions on the handout and asked students what they were seeing in the picture at the top of the handout. Students did not respond; the majority was not paying attention. Then, she projected the picture on the board and went over the levels of an organism shown on the picture. Some of the students started to pay attention. She was repeating her questions and directions several

times patiently to make sure everybody followed what she was talking about. Once she clarified for the students that they would be working at the DNA level in this task, she continued to read the rest of the instructions from the handout and invited students to make observations of the PCR output at the bottom of the handout.

Instead of concluding the set-up phase right after she introduced the work, however, she said, “Let’s do some pre-observations here.” She walked them through the key that was provided on the handout, such as where cross 1 and cross 2 were depicted, what was meant by a cross, and what F, M, and O meant. After making sure that students had an idea of what they were looking at and familiarizing them with the language used in the task (which appeared to be a good strategy for the groups of students she was working with), she left the actual thinking work of making analysis of the PCR data to the students.

Classroom Management as a Barrier In spite of her heavy scaffolding during the set-up phase, the thinking demand on students declined to “no thinking” about the disciplinary ideas and/or practices during the enactment phase of the PCR task. The majority of the students were disengaged and appeared to be unwilling to work on the task during its enactment. Their attention was drawn to the task only when Ms. Neal asked what they noticed; students mostly responded by sharing some superficial observations, such as “for every number, there is a letter that represents something”. The majority of the students did not seem to be aware of what they were working on. The class was noisy and the majority of the students were disruptive.

To conclude, these three types of changes showed the possibility that the thinking demand of high-level tasks can decline during the enactment phase even though students are positioned to engage in high level of thinking and reasoning as the teacher launches the task. More importantly, the detailed explanations of these cases illustrate the level and kind of student thinking (presented in the four quadrants in Fig. 2) while students were working on a cognitively demanding task. These cases also revealed the factors associated with maintenance or decline of cognitive demand in science classroom.

Discussion

This study illustrates that developing and selecting NGSS-aligned, high-level tasks can be a significant step toward engaging students in high levels of thinking and reasoning. In some of the classrooms, students appeared to be engaging in high levels of reasoning about genetics while engaging in a set of scientific practices. However, our findings also provide supporting evidence to existing literature that suggests that the thinking demands of cognitively complex tasks can be transformed once they are placed into real classroom settings. In some of the lessons, students did not engage in high level of thinking even though the task as designed and set-up positioned them to engage in higher levels of intellectual work. This study went beyond predicting this change in cognitive demand when high-level tasks are launched in the classroom with the intention of engaging students sensemaking around disciplinary ideas and practices. It helped to *characterize* the kind and level of thinking in science classrooms. It also suggested a set of factors that are associated with maintaining or declining cognitive demand in science classrooms. Therefore, this study has built on but expands earlier studies focusing on the change in cognitive demand across the phases of a task in science classrooms (e.g., Kang et al. 2016).

Our findings point to the importance of the enactment phase, in particular, the interaction among teacher, students and the task (i.e., instructional triangle; Cohen and Ball 1999, 2000)

as students actually go about working on the task. They do not, however, disregard the cognitive demand of the task as presented in written materials since tasks constituted one component of the instructional triangle. Students' motivation, prior knowledge, experiences, and interests, undoubtedly, play a role in the way they interact with the materials and respond to their thinking demands. Thus, we observed classroom management and students' motivation to learn as important factors associated with maintenance or decline of students' thinking. More importantly, however, *teachers' moves and practices* were observed to play a role in shaping the interaction between the task and the students. Teachers' high-level questions, pressing students for engaging in evidence-based reasoning, close attention to students' ideas appeared to play a role in students' high levels of thinking. These student and teacher factors are consistent with the factors that Stein and colleagues (Henningesen and Stein 1997; Stein and Smith 1998) identified as being related to maintaining cognitive demand in mathematics classrooms such as sustained press for justifications, or shift in emphasis from meaning, concepts, or understanding to the correctness and completeness of answers. Blumenfeld (1992) has seconded the importance of these factors saying, "high-level tasks do not exist separately from the way teachers implement them" (p. 109).

One observation did not transpire as expected. Based on earlier studies of science classrooms, one might expect IntHigh tasks to decline to IntLow kinds of thinking and reasoning during the enactment phase. This is because teachers often chose to walk students through challenging activities in a very detailed step-by-step fashion (e.g., scripting the work for the students so that by following the script they can complete the work). Prior research has provided evidence for this kind of thinking during the implementation of "cookbook labs" during which students are often asked to follow a procedure without needing to understand what they are doing and why (Chinn and Malhotra 2002; Germann et al. 1996). We did not observe this type of decline in any of the five observed classes. One reason could be that the PCR task—despite being an IntHigh task—does not lend itself to a decline to IntLow. If the task involved an investigation, this kind of decline could have been more likely because lab investigations often involve some set of procedures which could easily be turned into a set of scripts for students to follow tediously. Interestingly, we did observe a decline into IntLow in one of the classrooms that participated in the larger project but was not a part of the data analyzed for this paper. In that classroom, the teacher showed students "how to read" PCR data by developing a color-coding strategy and asking them to apply that strategy to read the PCR data. Students' thinking was focused on making sense of the strategy. Instead of trying to make sense of how genes are passed from parent to offspring, the students were trying to make sense of what blue and red lines represented and how they could apply that strategy to the second cross. Even though this could be a less likely observation, it shows that trying to script the task for the students could be one factor associated with decline the thinking demand for the students.

Conclusion and Implications

This was an exploratory study conducted with a small sample size to gain insight about the nature of change in thinking demands of cognitively complex science tasks from how they are designed to how they are set-up by the teacher in the classroom to how they are enacted by the teacher and the students. Our goal is not to make generalized claims but propose a way to investigate the level and kind of thinking in which students engage in science classrooms during the implementation of high-level tasks aligned with the NGSS vision. Future research

should investigate the change in the thinking demands of similar tasks in larger numbers of classrooms, which could allow for the identification of more generalizable patterns. Our analyses suggest that the TAGS in conjunction with the quadrant system that we proposed (Fig. 2) can be used as a lens to identify the nature of change across the phases of a task as presented in the Math and Science Framework.

In this study, we have also proposed a language to characterize the level and kind of student thinking in which students engage when they are invited to think and reason as demanded by NGSS-aligned curricular tasks. Specifically, the terms of the TAGS framework have been used to characterize the thinking demand of the tasks at three phases: as they appear in curricular materials, as they are set up by the teacher, and as students actually go about working on them. By providing a way to describe the kind and level of student thinking in science classrooms, the matrix allows us to identify and look for what might be typical changes in the thinking demands of cognitively complex curricular materials across different classrooms. We believe that these categories, which focus on two important features emphasized in the Framework and NGSS, can help us to better understand, diagnose, and communicate issues during the implementation of high-level tasks in science classrooms.

Identifying patterns of change in the thinking demands at the set up and/or task enactment phases is important for several reasons. Developing curriculum materials will be one of the main methods of infusing the NGSS vision across large number of science classrooms. This study supports the existing knowledge base that developing high-level tasks is essential *but not sufficient* for engaging students in high levels of thinking and reasoning about disciplinary ideas (Stein et al. 1996; Kang et al. 2016). Identifying the nature of change in thinking demands could help to develop tools and resources to support teachers' maintenance of high-level thinking demands. For example, the shift from integrated to isolated suggests that it would be more difficult to juggle both pressing students (a) to engage in deep reasoning about scientific ideas and (b) to productively engage in scientific practices which could allow them to understand how scientific knowledge develops. Building some scaffolds into the tasks and/or curriculum guides designed for teachers could help to support maintaining cognitive complexity of high-level science tasks.

This study brings attention to two phases of science instruction that has not been a focus of research in science education but has recently been taken up by others (e.g., Kang et al. 2016; Tekkumru-Kisa 2013; Tekkumru-Kisa and Stein 2014, 2015). These are set-up and enactment phases of a task. Examining these phases separately helped us to identify the breakpoints in the way students are positioned to think and reason in science classrooms. As discussed by Stein and colleagues (2000), the set-up phase of instruction shows the teacher's "communication to students regarding what they are expected to do, how they are expected to do it, and with what resources" (p.25). So, set-up is important to help frame the work that students are invited to do. Guided by the studies in mathematics education (e.g., Jackson et al. 2013; Stein et al. 1996), future research can shed light on how teachers' *set up* of a science task impacts students' opportunities for learning science. The Four Quadrants that we propose could provide a lens to identify the patterns, which then can be elaborated further in future research shedding light on the factors associated with the maintenance and decline in thinking demands of cognitively complex tasks during the set-up as well enactment phases.

Last but not least, this study began to uncover the *factors* that appeared to play a role in the level and kind of thinking observed in these classrooms. These included (1) setting up tasks as part of a larger unit/idea, (2) Clarifying expectations, (3) Pressing for evidence-based reasoning, (4) Attending to and advancing students' ideas, (5) Attending to students' ideas and

engagement, (6) An emphasis on correctness and completeness of the ideas, and (7) Classroom management as a barrier. By identifying these factors that has been discussed in the literature separately to describe the features of quality teaching, in this study, we combined them under a coherent theme: factors associated with maintenance (or decline) of high-cognitive demands on students' thinking. For example, there has been growing research on attention and responsiveness to students' ideas (Jacobs et al. 2011; Levin and Richards 2011; Levin et al. 2013, 2012; Sherin and Han 2004). The cases that we discussed in this paper illustrated that attention to students' ideas plays a critical role to maintain cognitive demand on students' thinking. Similarly, consistent with the prior literature, which emphasized that procedures and correct answers at the expense of meaning and understanding (Davis 1997; Levin 2008; Stein et al. 2000) could limit students' engaging in sensemaking. The cases that we discussed revealed that an emphasis on correctness and completeness of the ideas may decline demand on students' thinking. Even though we cannot make causal claims about the effect of these factors on students' thinking, these cases provided a concrete depiction of these factors in relation to the kind and level of thinking that students engaged in these classrooms. Future research should systematically analyze classrooms based on the teacher and student-related factors that could be associated with maintaining (or declining) cognitive demand of tasks on students' thinking.

Acknowledgements Work on this paper was supported by a grant DRL-1027629 from the National Science Foundation to the Learning Research and Development Center of the University of Pittsburgh. All opinions and conclusions in this paper are those of the authors and do not necessarily reflect the views of the funding agency. We would like to thank Louis Alferi for his contribution to the data analysis.

References

- Anderson, D., & Nashon, S. (2007). Predators of knowledge construction: Interpreting students' metacognition in an amusement park physics program. *Science Education*, 91(2), 298–320.
- Arbaugh, F., & Brown, C. A. (2005). Analyzing mathematical tasks: A catalyst for change. *Journal of Mathematics Teacher Education*, 8(6), 499–536.
- Aubusson, P., Panizzon, D., & Corrigan, D. (2016). Science education futures: “great potential. Could do better. Needs to try harder”. *Research in Science Education*, 46(2), 203–221.
- Banilower, E., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weiss, A. M. (2013). *Report of the 2012 national survey of science and mathematics education*. Chapel Hill: Horizon Research Inc..
- Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist*, 26(3&4), 369–398.
- Blumenfeld, P. C. (1992). The task and the teacher: Enhancing student thoughtfulness in science. In J. Brophy (Ed.), *Advances in research on teaching: Planning and managing learning tasks and activities, Vol. 3*. Greenwich: JAI.
- Boaler, J., & Staples, M. (2008). Creating mathematical futures through an equitable teaching approach: The case of Railside school. *Teachers' College Record*, 110(3), 608–645.
- Boston, M., & Wolf, M. K. (2004). *Using the Instructional Quality Assessment (IQA) toolkit to assess academic rigor in mathematics lessons and assignments*. San Diego: Paper presented at the Annual Meeting of the American Educational Research Association Meeting.
- Brown, M., & Edelson, D. (2003). Teaching as design: Can we better understand the ways in which teachers use materials so we can better design materials to support their changes in practice. In *Design Brief*. Evanston: Center for Learning Technologies in Urban Schools.
- Brown, M. W. (2009). The teacher-tool relationship: Theorizing the design and use of curriculum materials. *Mathematics Teachers at Work: Connecting Curriculum Materials and Classroom Instruction*, 17–36.

- Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, 86(2), 175–218.
- Cohen, D. K., & Ball, D. L. (1999). *Instruction, capacity, and improvement* (CPRE Research Report No. RR-043). Philadelphia: University of Pennsylvania, Consortium for Policy Research in Education.
- Cohen, D. K., & Ball, D. L. (2000). *Instructional innovation: Reconsidering the story*. New Orleans: Paper presented at the meeting of the American Educational Research Association.
- Crawford, B. A. (2014). From inquiry to scientific practices in the science classroom. In N. Leder-man & S. Abell (Eds.), *Handbook of research on science education, Vol II*. New York: Routledge.
- Davis, B. (1997). Listening for differences: An evolving conception of mathematics teaching. *Journal for Research in Mathematics Education*, 355–376.
- Doyle, W. (1983). Academic work. *Review of Educational Research*, 53(2), 159.
- Doyle, W. (1988). Work in mathematics classes: The context of students' thinking during instruction. *Educational Psychologist*, 23(2), 167–180.
- Duncan, R., & Cavera, V. (2015). DCIs, SEPs, and CCs, Oh My! Understanding the three dimensions of the NGSS. *Science and Children*, 52(2), 16–20.
- Erduran, S. (2014). Beyond nature of science: The case for Reconceptualising “Science” for science education. *Science Education International*, 25(1), 93–111.
- Germann, P. J., Haskins, S., & Auls, S. (1996). Analysis of nine high school biology laboratory manuals: Promoting scientific inquiry. *Journal of Research in Science Teaching*, 33(5), 475–499.
- 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: Information Age Publishing.
- Harris, C. J., Penuel, W. R., D'Angelo, C. M., DeBarger, A. H., Gallagher, L. P., Kennedy, C. A., et al. (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.
- Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: classroom based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education*, 28(5), 524–549.
- Hiebert, J., & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. *American Educational Research Journal*, 30(2), 393–425.
- Jacobs, V. R., Lamb, L. L. C., Philipp, R. A., & Schappelle, B. P. (2011). Deciding how to respond on the basis of children's understandings. In M. G. Sherin, V. R. Jacobs, & R. A. Philipp (Eds.), *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 17–34). New York: Routledge.
- Jackson, K. J., Shahan, E. C., Gibbons, L. K., & Cobb, P. A. (2012). Launching complex tasks. *Mathematics Teaching in the Middle School*, 18(1), 24–29.
- 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.
- Jones, M. T., & Eick, C. J. (2007). Implementing inquiry kit curriculum: Obstacles, adaptations, and practical knowledge development in two middle school science teachers. *Science Education*, 91(3), 492–513.
- Kang, H., Windschitl, M., Stroupe, D., & Thompson, J. (2016). Designing, launching, and implementing high quality learning opportunities for students that advance scientific thinking. *Journal of Research in Science Teaching*.
- Krajcik, J., Blumenfeld, P. C., Marx, R. W., Bass, K. M., Fredricks, J., & Soloway, E. (1998). Inquiry in project-based science classrooms: Initial attempts by middle school students. *Journal of the Learning Sciences*, 7(3–4), 313–350.
- Levin, D. M. (2008). *What secondary science teachers pay attention to in the classroom: Situating teaching in institutional and social systems*. University of Maryland at College Park: Unpublished Doctoral Thesis.
- Levin, D. M., & Richards, J. (2011). Learning to attend to the substance of students' thinking in science. *Science Educator*, 20(2), 1–11.
- Levin, D. M., Grant, T., & Hammer, D. (2012). Attending and responding to student thinking in science. *The American Biology Teacher*, 74(3), 158–162.
- Levin, D., Hammer, D., Elby, A., & Coffey, J. (2013). *Becoming a responsive science teacher: Focusing on student thinking in secondary science*. Arlington: NSTA Press.
- Marx, R., Blumenfeld, P., Krajcik, J., & Soloway, E. (1997). Enacting project-based science. *The Elementary School Journal*, 97(4), 341–358.
- Mehalik, M. M., Doppelt, Y., & Schunn, C. D. (2008). Middle-school science through design-based learning versus scripted inquiry: better overall science concept learning and equity gap reduction. *Journal of Engineering Education*, 97(1), 71–85.
- Nashon, S. M., Anderson, D., Okenywa, P., Kelonye, F., Ooko, S., & Ombogo, P. (2015). Student learning impact on science teachers' teaching: The case of a form 3 science case in Kenya. *Journal of Technology & Socio-Economic Development*, 4(6), 32–38.

- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and Core ideas*. Washington: The National Academies Press.
- NGSS Lead States. (2013). *Next Generation Science Standards: For States, by States*. Washington: The National Academies Press.
- Osborne, J. (2014). Teaching scientific practices: meeting the challenge of change. *Journal of Science Teacher Education*, 25(2), 177–196.
- 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.
- Powell, J. C., & Anderson, R. D. (2002). Changing teachers' practice: Curriculum materials and science education reform in the USA. *Studies in Science Education*, 37(1), 107–135.
- Reiser, B. J. (2013). *What Professional Development Strategies Are Needed for Successful Implementation of the Next Generation Science Standards?* White paper presented to the Invitational Research Symposium on Science Assessment, K-12 Center at ETS.
- Reiser, B. J. (2014, April). *Designing coherent storylines aligned with NGSS for the K-12 classroom*. Boston: National Science Education Leadership Association Meeting.
- Reiser, B. J. (2015). Designing coherent storylines aligned with NGSS for K-12 classroom Received on April 1, 2017 from https://www.academia.edu/6884962/Designing_Coherent_Storylines_Aligned_with_NGSS_for_the_K-12_Classroom
- Russ, R. S., & Luna, M. J. (2013). Inferring teacher epistemological framing from local patterns in teacher noticing. *Journal of Research in Science Teaching*, 50(3), 284–314.
- 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.
- Schuchardt, A. M., & Schunn, C. D. (2016). Modeling scientific processes with mathematics equations enhances student qualitative conceptual understanding and quantitative problem solving. *Science Education*, 100(2), 290–320.
- Sherin, M. G., & Han, S. Y. (2004). Teacher learning in the context of a video club. *Teaching and Teacher Education*, 20(2), 163–183.
- Smith, M. S., & Stein, M. K. (2011). *Five practices for orchestrating productive mathematics discussions*. Reston: National Council of Teachers of Mathematics.
- Stein, M. K., & Kim, G. (2009). The role of mathematics curriculum materials in large-scale urban reform. In G. M. Lloyd, J. T. Remillard, & B. A. Herbel-Eisenmann (Eds.), *Mathematics teachers at work: Connecting curriculum materials and classroom instruction* (pp. 37–55). New York, NY: Routledge.
- Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: an analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation*, 2(1), 50–80.
- Stein, M. K., & Smith, M. S. (1998). Mathematical tasks as a framework for reflection: from research to practice. *Mathematics Teaching in the Middle School*, 3(4), 268–275.
- Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: an analysis of mathematical tasks used in reform classroom. *American Educational Research Journal*, 33(2), 455–488.
- Stein, M. K., Smith, M. S., Henningsen, M. A., & Silver, E. A. (2000). *Implementing standards-based mathematics instruction: A casebook for professional development* (First Edition). New York: Teachers College Press.
- Stigler, J. W., & Hiebert, J. (2004). Improving mathematics teaching. *Educational Leadership*, 61(5), 12.
- Tarr, J. E., Reys, R. E., Reys, B. J., Chavez, O., Shih, J., & Osterlind, S. (2008). The impact of middle grades mathematics curricula and the classroom learning environments on student achievement. *Journal for Research in Mathematics Education*, 39, 247–280.
- Tekumru-Kisa, M. (2013). Science teachers' learning to notice from video cases of the enactment of cognitively demanding instructional tasks. Unpublished doctoral dissertation, University of Pittsburgh.
- Tekumru-Kisa, M., & Stein, M. K. (2014). Using contrasting video cases of enactment of cognitively demanding science tasks in professional development. In J. L. Polman, E. A. Kyza, D. K. O'Neill, I. Tabak, W. R. Penuel, A. S. Jurow, K. O'Connor, T. Lee, & L. D'Amico (Eds.), *11th International Conference of the Learning Sciences* (pp. 808–815). Boulder, CO: International Society of the Learning Sciences, Inc.
- Tekumru-Kisa, M., & Stein, M. K. (2015). Learning to see teaching in new ways a foundation for maintaining cognitive demand. *American Educational Research Journal*, 52(1), 105–136.
- Tekumru-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.

- Tekumru-Kisa, M., Hiester, H., & Kisa, Z. (2017). *Nature of science tasks as a lens to understand students' opportunities to learn*. San Antonio: Paper presented at 2017 Annual Meeting of the American Educational Research Association.
- Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Toulmin, S. E., Rieke, R. D., & Janik, A. (1979). *An introduction to reasoning*. New York: Macmillan.
- Trautmann, N., MaKinster, J., & Avery, L. (2004). What makes inquiry so hard? (and why is it worth it?). In *Annual meeting of the National Association for Research in Science Teaching*, Vancouver, BC.
- Webb, N. L. (1997). *Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education*. Research Monograph No. 6.
- Weiss, I. R. (1987). *Report of the 1985–86 national survey of science and mathematics education*. Research Triangle Park: Research Triangle Institute.
- Winne, P. H., & Marx, R. W. (1982). Students' and teachers' views of thinking processes for classroom learning. *The Elementary School Journal*, 82(5), 493–518.