Research Article

Analysis of inquiry materials to explain complexity of chemical reasoning in physical chemistry students' argumentation[1]

Moon 0000-0003-0379-294X, Alena; Stanford, Courtney; Cole, Renee; Towns, Marcy

Alena Moon: Department of Chemistry, University of Michigan, Ann Arbor, MI, 48109
Courtney Stanford : Department of Chemistry, Virginia Commonwealth University, Richmond, VA, 23284
Renee Cole : Department of Chemistry, University of Iowa, Iowa City, IA 52242
Marcy Towns : Department of Chemistry, Purdue University, West Lafayette, IN, 47907

[1] This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi:10.1002/tea.21407

Abstract
One aim of inquiry activities in science education is to promote students' participation in the practices used to build scientific knowledge by providing opportunities to engage in scientific discourse. However, many factors influence the actual outcomes and effect on students' learning when using inquiry materials. In this study, discourse from two physical chemistry classrooms using the Process-Oriented Guided Inquiry Learning (POGIL) approach was analyzed using a lens of scientific argumentation. Analysis of the complexity of reasoning in students' arguments using a learning progression on chemical thinking indicated that students did not employ very complex reasoning to construct arguments. To explain the distribution of reasoning observed, a separate analysis of the curricular materials was performed using the Task Analysis Guide for Science (TAGS). Results indicate a relationship between the task's targeted scientific practice and how students used evidence in their arguments as well as between the task's cognitive demand and the complexity of reasoning employed in arguments. Examples illustrating these relationships can be used to inform implications for design of inquiry materials, facilitation of classroom discourse, and future research.

Keywords: Science argumentation, inquiry, classroom discourse, physical chemistry, task analysis

With the widespread recognition that social, cultural, and discursive components of a learning environment impact student outcomes, discourse analysis has emerged as a means of evaluating the quality of the teaching and learning that takes place in classes (O'Loughlin, 1992; Mercer, 2007; Authors, 2014). There is experimental evidence that suggests that when children participate in sustained quality discussions, their problem-solving skills and individual learning improve (Mercer & Howe, 2012). This warrants investigation of the quality of classroom discourse when evaluating the learning that takes place, though few studies have explored the relationship between discourse quality and learning outcomes (Mercer, 2007). Consideration of collective reasoning is especially valuable for its capacity to reveal how students participate in disciplinary practices in the classroom (Cobb, Stephan, McClain, & Gravemeijer, 2001). In light of national and disciplinary calls for explicit emphasis on scientific practices, discourse analysis becomes a necessary tool for researchers and instructors to ensure the development of these practices in science classrooms (NRC, 2012; NRC, 2012).

Models of argumentation, a discursive practice central to the construction of scientific knowledge, have provided useful methodological frameworks for investigating student reasoning (Authors, 2014; Erduran, 2007; Erduran, Simon, & Osborne, 2004; Cole et al., 2012; Bricker & Bell, 2008; Sampson & Clark, 2011; McNeill & Krajcik, 2012; Kulatunga & Lewis, 2013; Kulatunga, Moog, & Lewis, 2013; Kulatunga, Moog, & Lewis, 2014). Among the multiple models of argumentation, Toulmin's model consisting of claim, data, and warrant is perhaps the most widely used (Erduran, 2007). Using this model, researchers have investigated the relationships between argumentation and chemistry reasoning (Moon et al., 2016; Becker et al., 2013), conceptual learning (Asterhan & Schwarz, 2007; Cetin, 2014), students' questioning (Chin & Osborne, 2010), problem solving (Cho & Jonassen, 2002), students' understanding of the nature of science (Khifshe, 2014), task goals (Garcia-Mila, Gilabert, Erduran, & Felton, 2012), and curricular materials (Kulatunga, et al., 2014).

The relationships between argumentation and instructional context provide important insights that can be used to help design curricula that target the outcomes of interest. Duschl and Osborne (2002) speak broadly about classroom conditions for effective argumentation. To promote argumentation in the classroom, the tasks must be group activities that require collaboration in order to complete. This means that they should promote discourse. Finally, they must include questions that scaffold the construction of arguments (Duschl & Osborne, 2002). It can also be helpful to provide students with a model of argumentation. Using a cognitive apprenticeship model, Jimenez-Aleixandre (2007) adds curricular considerations for an inquiry curriculum to promote and scaffold argumentation: diversity of outcomes, problem-solving, depth, epistemic practices, discursive practices, and authenticity. In their learning progression on scientific argumentation, Berland and McNeill (2010) relate complex instructional contexts (i.e. diversity of outcomes, student-defined data set, and no scaffolding) to complex argumentative products and processes. There is quite a bit of variability allowed within this progression as a task can have a complex student-defined data set and still have scaffolding, for example. Further, they found that high complexity on the instructional context does not necessitate high complexity on the process or product. Berland and McNeil's (2010) results highlight the importance of using students' discursive products to evaluate the instructional context.

Garcia-Mila et al. (2012) considered the impact of two argumentative task goals on argumentation: persuasion and consensus. In the persuasion condition, the students were tasked with convincing each other of their position; while in the consensus condition, the students were tasked with reaching a collaborative conclusion. There was greater variety in the structure of arguments constructed in the consensus condition than in the persuasion condition. Specifically, the consensus condition included more two-sided argument structures, with more rebuttals. These findings are important as they reveal that the objective assigned to students impacts the amount of student reasoning that will be exposed during the argument task, confirming Mercer's (2000) claim that not all classroom conversation tasks promote student reasoning equally.

In chemistry, Kulatunga and colleagues (2014) also explored the relationship between the curricular materials (guided inquiry materials for general chemistry) and students' argumentation. They distinguished between directed questions, prompts that can be directly answered with previous information or knowledge, and convergent questions, prompts that require multiple pieces of information and some level of synthesis to answer. They found that directed questions resulted in more 'claim' and 'claim and data' utterances, while convergent questions resulted in more basic and high level arguments. The authors also considered how questions in the learning cycle of exploration, concept development, and application impacted resulting argumentation. It was evident that application questions elicited more arguments (claim, data, and warrant) than non-arguments (claim or claim, data). The authors propose having a blend of different question types, explicit calls for explanation, and the presence of a scaffolded learning style to support argumentation (Kulatunga et al., 2014).

Though these studies provide an important start in investigating the relationship between curricular materials and resulting argumentation, the results were limited to consideration of the number of arguments of varying structures. We believe that in addition to a quantitative picture of argumentation, we must build a qualitative understanding of the nature of arguments resulting from certain types of curricular prompts. Better understanding of the nature of the resulting discursive products serves to answer the question of what kind of student reasoning is revealed through argumentation.

There is one primary reason for considering the relationship between the quality of argumentation and task goals and curricular materials. Namely, the enacted curriculum is one component of the classroom that we, as instructors and curriculum designers, have a significant capacity to influence (Berland & McNeill, 2010). To this end, the study presented herein uses classroom discursive products (arguments) to evaluate the enacted inquiry curricular materials (Process-Oriented Guided Inquiry Learning [POGIL]).

Process-Oriented Guided Inquiry Learning

The POGIL pedagogy was grounded in constructivism, focusing on the notion that learning is enhanced when students are actively engaged in class, constructing knowledge and drawing conclusions by analyzing data and discussing ideas. (Farrell, Moog, & Spencer, 1999; Moog, 2008) A key aspect of a POGIL implementation is that students spend class time working in small groups while the instructor serves as a facilitator, monitoring progress and intervening when necessary. Argumentation serves as an appropriate methodological framework for revealing how students use the information provided to them in the materials to make claims and reason about the content (Cole et al., 2012).

Our choice to investigate the POGIL approach in a thermodynamics classroom was motivated by two reasons. With increasing instructor buy-in to this approach, it is crucial that student work be used to evaluate how the curriculum meets desired outcomes in order to support instructors in effective implementation (Daubenmire, Bunce, Draus, Frazier, Gessell, & van Opstal, 2015). To this end, a few studies have investigated classroom discourse norms and patterns resulting from implementing POGIL activities in introductory and advanced undergraduate chemistry classes (Kulatunga et al., 2013; Kulatunga & Lewis, 2013; Becker et al., 2013; Becker et al., 2015). The widespread and expanding use of POGIL within the undergraduate science classroom warrants further investigation of its impact on discursive practices, which have been understudied in the post-secondary science classroom, particularly in upper-level courses.

Chemical thermodynamics is a worthwhile content area to investigate for multiple reasons. Chemical thermodynamics is the study of energy changes associated with chemical processes, specifically targeting the crosscutting concept of ‘Energy and matter: Flows, cycles, and conservation’ in the context of chemistry (NRC, 2012). This subject has shown to be particularly difficult for students for a variety of reasons. Students can bring strongly held prior conceptions to this material, built through personal experiences as ideas of heat, work, and energy can be used ubiquitously in daily speech (van Roon, van Sprang, & Verdonk, 1994). Finally, chemical thermodynamics explicitly involves multiple scientific practices and cross-cutting concepts (i.e. use of mathematics and computational thinking, engaging in argument from evidence, constructing explanations, cause and effect, and stability and change). As this subject houses so many important components of scientific inquiry, investigating students’ development of conceptual understanding and scientific practice in thermodynamics has implications for broader science education. Additionally, this course is often taken by senior-level students, which means that investigation at this level reveals where attention needs to be paid earlier in the curriculum to support students achieving desired program outcomes.

Mechanistic or causal reasoning in science education

One specific aspect of discourse that was of interest in this study was students’ use of causal reasoning. The Next Generation Science Standards highlight cause and effect reasoning as a crosscutting concept in science (NRC, 2012), but less work has been done to characterize the causal models that students, especially postsecondary students, use to explore phenomena. This is especially important because understanding chemical thermodynamics requires complex causal models. Expert-like causal models have been shown to be complex, dynamic, and integrated (Perkins & Grotzer, 2005; Brown et al., 2010). In contrast, novice-like causal models tend to consider a single salient feature and assign it total causal agency in explaining an outcome (Smith, Carey, & Wiser, 1985; Perkins & Grotzer, 2005; Sevian & Talanquer, 2014). An intervention involving explicitly teaching different causal models to primary students resulted in students' use of more complex causal models (Perkins & Grotzer, 2005). Perkins and Grotzer (2005) argue that access to more complex causal models increases access to a broader set of scientific concepts. Situating upper-level postsecondary students in this spectrum according to their use of causal reasoning is immensely important, then, for understanding how this skill might progress (Moon et al., 2016). Beyond understanding how the skill might progress, it is essential to understand how complex causal reasoning can be scaffolded.

The Sociocultural Perspective

The use of classroom discourse analysis was justified by the theoretical position that to evaluate the quality of a feature of instructional context, it is fruitful to evaluate the discursive products of that curriculum (Mercer, 2007; Mercer & Howe, 2012). This view is informed by Vygotskian assertions that knowledge is constructed socially through mediating tools, which includes curricular materials, and then internalized by individual students in their own knowledge development (Wertsch, 1991; John-Steiner & Mahn, 1996; Mercer & Howe, 2012). The sociocultural perspective frames this study by directing us to consider two factors in understanding student development of knowledge. First, it places emphasis on students' interactions with mediating tools. In this case, how do students interact with the information provided to them? Second, it places emphasis on evaluating the quality of the social activity. As individuals will internalize this social activity, ensuring that this social activity captures the intended learning outcomes is necessary for developing an effective curriculum. This evaluation is achieved through discourse analysis (Mercer, Dawes, Wegerif, & Sams, 2004; Mercer, 2007).

In this study, qualitative methods were employed to understand the nature of the discursive products resulting from student interactions with different types of curricular prompts.

Curriculum Design and Evaluation Model

To frame this investigation and the results presented in this work, a curriculum conjecture map (Sandoval, 2014) was generated and provided in Figure 1. Originally intended as an argument structure for educational design research, conjecture maps can make salient relationships that are worthy of understanding for a particular design. For our study, the map provided a way to theorize the relationship between the curriculum (POGIL physical chemistry: Thermodynamics) and the discursive products. In this way, the conjecture map serves to generate a tentative model based on the theorized relationship between student interaction with mediating tools and learning outcomes. Ultimately, an understanding of this relationship helps to evaluate the curriculum, provide guidance to developers, and support teachers in their implementation (Bismack, Arias, Davis, & Palincsar, 2015). This conjecture map is by no means comprehensive. Rather, it served to guide analysis by directing us to investigate specific relationships according to the two desired outcomes we focused on in this study: causal reasoning and appropriate use of evidence.

Figure 1. Conjecture map for POGIL Thermodynamics materials (Spencer, Moog, & Farrell, 2004)

Our high-level conjecture is that inquiry-oriented physical chemistry materials can support student learning of concepts and practices in physical chemistry where the curricular materials mediate student participation in classroom discourse. This conjecture is embodied in the use of Process-Oriented Guided Inquiry Learning, a widespread inquiry approach for undergraduate chemistry. In addition to use of this specific set of materials developed by Spencer et al. (2004) for upper-level physical chemistry courses, instructor- and classroom-driven participation structure and discourse practices serve to embody our conjecture. We posit that processes of constructing arguments and interacting with the inquiry materials serve to mediate student development of the intended outcomes. In this study, we considered two desired outcomes, though we recognize there are many associated with this curricular approach. The first outcome was the complexity of causal reasoning employed in constructing arguments, in accordance with national (NRC, 2012) and disciplinary (Russ, Scherr, Hammer, & Mikeska, 2008; Talanquer, 2010) calls for consideration of students' reasoning about causal mechanisms. The second desired outcome was developing students' use of evidence. Particularly relevant in characterizing the appropriate use of evidence for physical chemistry is the use of both mathematical and conceptual reasoning.

Our high-level conjecture is partially embodied in the POGIL Thermodynamics materials (Spencer, Moog, and Farrell, 2004). Because our desired outcomes include a cross-cutting concept (causal reasoning) and scientific practice (use of mathematics and computational thinking), we aimed to characterize the task structures using the Task Analysis Guide for Science (TAGS), which provided an understanding of the practices and concepts targeted by the POGIL activities (Tekkumru-Kisa, Stein, & Schunn, 2015). Task structure was the primary consideration for embodiment, with participation structure and discourse practices being secondary. More targeted investigation of the participation structure and discourse practices can be found in other work (Stanford et al., 2016).

Research Questions

This study was guided by the following research question:

How do curricular prompts in physical chemistry inquiry materials support and/or constrain (1) students' use of evidence and (2) students' causal reasoning?

Methods

Participants and Setting

Two physical chemistry classes, which used the POGIL approach and the Spencer, Moog and Farrell POGIL thermodynamics materials (2004), were observed. Table 1 includes relevant demographics, including important similarities and differences between the two classes. Arguments from both of these classes were pooled in order to consider the relationship between arguments and the curricular materials. The POGIL activities covered by both classes were included in the curricular analysis. Table 2 shows the activities and the corresponding content covered by each.

Data Collection

Data collection modeled a methodology originating in mathematics education for documenting collective activity (Cobb & Whitenack, 1996; Rasmussen & Stephan, 2008; Cole et al., 2012). This method is especially appropriate for considering classroom discourse over a period of time. To this end, whole class periods were videotaped. Classroom A was videotaped for approximately half of the total course time, though only thermodynamics activities will be evaluated for this work. In the case of classroom B, video data of two months of the course were collected. To capture small group interactions, one small group was videotaped during each class period. All videos were transcribed verbatim. Students were assigned pseudonyms to protect their identity. This study received necessary IRB approval for data collection and analysis.

Analytic Framework

Argument logs were generated using the Toulmin Argument Pattern (Toulmin, 1958), featured in Figure 2. Transcripts were coded using the components from the Toulmin Argument

Pattern (TAP). This involved identifying claims, which usually corresponded to answers to critical thinking questions in the POGIL inquiry activities. If a claim was supported with data or evidence, they were extracted as an argument. The whole episode containing the claim and data was framed according to TAP. Two graduate students individually coded transcripts for arguments and then met with the entire research team to confirm interpretation and generate a consensus argument log. A majority of the arguments included paraphrased statements aimed at capturing the meaning the students and instructor aimed to convey. Paraphrases were also used when multiple statements conveyed only one component of an argument. If the meaning could not be clearly derived from the student's words or the argument component corresponded to one statement, their exact statements were used and italicized in the argument logs. This method provided a means of condensing the large amount of text resulting from classroom discourse to smaller, clearer episodes that could then be analyzed.

Analysis of Arguments

Arguments were analyzed using the modes of reasoning from the Chemical Thinking Learning Progression (Sevian & Talanquer, 2014; Szteinberg et al., 2014; Banks et al., 2015; Cullipher et al., 2015; Moon et al., 2016). The modes of reasoning, as interpreted by the primary researcher, are presented in Table 3. These were used to characterize the complexity of reasoning students employed in their arguments.

Arguments were labeled as descriptive, relational, linear, or multicomponent based on the features provided in Table 3. The features were clear and effective at differentiating arguments from each other. Descriptive arguments tended to provide little new information, instead repeating back features provided in the problem. Relational arguments used reasoning like "because of X, Y happens" or used a relationship without explanation to justify an output or claim. These arguments were distinguished from linear and multicomponent arguments by the presence of a mechanism. In the case of linear arguments, this mechanism was linear with stepwise causal reasoning. Multicomponent arguments, on the other hand, showed evidence of students considering and weighing multiple variables as contributing to a possible outcome. This coding step was completed through multiple iterations. Upon receiving feedback from the research team and chemistry education researchers, codes were refined and reassigned accordingly.

Reliability was also ensured iteratively. During the first round of coding, members of the research team independently coded the argument log from one class period. After this, the research team met and discussed the meaning and interpretation of the labels and arguments until consensus was reached; that is, all arguments involving disagreement were resolved. The primary researcher applied any revisions made to interpretation of the codes as a result of this discussion to the rest of the arguments. Later in analysis, the primary researcher worked with two researchers outside of the research team to discuss the meaning of the labels and the ways that they were assigned. This discussion similarly continued until all disagreements and questions were resolved. The aim of these measures was to ensure that the primary researcher was interpreting arguments and applying labels consistently.

Analysis of POGIL Curriculum

Prompts from the POGIL thermodynamics inquiry activities were coded using the Task Analysis Guide for Science (TAGS) (Tekkumru-Kisa, et al., 2015). If prompts were coded as “scientific practices” or “integration of content and practices”, they were further categorized according to which scientific practice they targeted (NRC, 2012). The TAGS framework evaluates tasks along two dimensions, cognitive demand and integration of content and practices.

Table 4 shows the TAGS framework, replicated from Tekkumru-Kisa et al. (2015). This framework was appropriate for directly characterizing and evaluating the inquiry activities used in these classrooms.

To assign a TAGS label to a prompt, multiple features including context, placement in the activity, and information provided were all taken into consideration. Placement in the activity and information provided were especially important for delineating cognitive demand levels. For example, if the students had already derived all relevant equations before the prompt, the prompt would tend to be scripted. In contrast, if the prompt is the first in an activity and requires the generation of new information, it was more likely to receive a guided label. Generally, cognitive demand was assigned according to the following criteria, as interpreted by the primary researcher, available in Table 5 (Tekkumru-Kisa et al., 2015).

To distinguish between practices, only what was explicitly elicited in the prompt was considered. For example, a prompt must tell the students to generate a question in order to be target the practice of ‘asking questions.’ For this reason, most prompts were easily distinguishable according to the TAGS framework. In considering prompts that were difficult to categorize for the primary researcher, however, feedback was sought from other chemistry education researchers. To gather this feedback, a group of chemistry education researchers was instructed about the TAGS framework and provided examples of each practice and each cognitive demand level and content-practices integration. Upon instruction and discussion, the group independently coded multiple prompts that the primary researcher found particularly difficult to categorize. After individual coding, a discussion about the TAGS labels assigned was conducted until consensus was reached about interpretation and application of the TAGS framework. The primary researcher assigned labels to those prompts considered in accordance with the feedback received. Further, the primary researcher changed TAGS assignments for other prompts as was appropriate after negotiating the interpretation and application of the TAGS framework.

Results

Task structure of POGIL Thermodynamics Inquiry Activities

The inquiry activities analyzed in this study overwhelmingly engage students in scripted tasks, which means a majority of prompts require students to follow a clear set of steps to complete the task. Most of the prompts target the integration of both content and practices, as seen in Figure 3. There are two levels of guided cognitive demand because guided integration tasks are thought to be higher in cognitive demand than guided practice and guided content. There were no tasks that targeted only a scientific practice at any cognitive demand level. Given that these inquiry activities largely target the integration of content and practice; which practices were being targeted was of interest. Figure 4 shows the distribution of practices targeted by integrated prompts. The POGIL Thermodynamics curriculum overwhelmingly uses scripted integration tasks and targets the practices of 'using mathematics and computational thinking' and 'constructing explanations.' This is not unexpected for a physical chemistry course as physical chemistry content draws heavily on the use of mathematical models to understand chemical processes.

Supports and constraints of appropriate use of evidence

To investigate this relationship, variation in the curricular prompts and practices and content targeted by these prompts were considered to explain trends in the resulting arguments. This investigation revealed that features of the prompt (i.e. which practice(s) were targeted) impacted the type of information students used as evidence in their arguments. Prompts that supported the construction of arguments using mathematical reasoning, arguments using conceptual or phenomenal reasoning, and arguments using data were identified. It was found that the practice targeted by the prompt that elicited the argument largely determined the types of evidence use that were observed. That is, prompts that targeted mathematical reasoning elicited arguments that used mathematical reasoning or manipulation as evidence. This relationship is evident in qualitative consideration of prompts and resulting arguments.

Supporting Mathematical Reasoning

When responding to prompts targeting the practice of 'using mathematics and computational thinking', students frequently describe their mathematical manipulations as warrants for their claims. These prompts incorporated very little conceptual or phenomenal reasoning. The excerpt presented in Table 6 illustrates how physical chemistry students completed mathematical tasks and constructed arguments about them.

Both of these arguments followed a linear mathematical pattern in which the claim is the final mathematical output, the data was the starting equation(s), and the warrant provided how the data led to the final outcome. Students used the information provided in the prompt as a starting point for solving the problem, which is a well-documented approach to problem-solving (Sweller, 1988). Arguments resulting from prompts that targeted the practice of using mathematics were generally descriptive of the computations. This is promising as it reveals that following mathematical instructions is an achievable task for students. However, these tasks could also be serving to constrain students' conceptual reasoning about the mathematical operations evidenced by the resulting arguments that reveal very little about their conceptual reasoning.

Supporting conceptual reasoning

Targeting the practice of ‘constructing explanations’ in addition to ‘using mathematics’ can support both mathematical and conceptual reasoning. Consider the arguments presented in Table 7. This prompt required the students to predict the outcome of a phenomenon. In order to do this, students must use the mathematical relationship between temperature, heat capacity, and entropy. This prompt reflects how physical phenomena, observables, symbolic reasoning, and abstract mathematical relationships are integrated in the study of thermodynamics.

Quentin and Melody used markedly different data to support their claims. Melody used an equation from the curricular materials, while Quentin used a qualitative description of the phenomenon being considered. A sufficient argument would have incorporated both an understanding of the phenomenon and an understanding of the underlying mathematical relationships. Melody concluded that the change in entropy of reaction was impossible to determine and used mathematical reasoning to justify her claim. Her warrant and backing were indicative of some misunderstandings of the phenomenon and the impact on the mathematics. Her warrant that “it depends on whether the natural log of T_2/T_1 is greater or less than 1” failed to incorporate the condition of increasing temperature, as that would mean that there was no way for the ratio of final temperature to initial temperature to be less than one. Melody’s warrant was not incorrect, but rather failed to incorporate reasoning about the phenomenon being considered. She expressed a meaningful concern when she considered the magnitude of change, which can be largely affected by logarithmic math, but this did not justify her claim that it was ultimately impossible to determine. Quentin used simpler reasoning when he warranted that “you’re increasing your final, then you’re going to receive a bigger number”, referring to increasing final temperature leading to a larger change in entropy. The arguments in Table 7 point to the difficulty of relating mathematical reasoning and conceptual reasoning about a phenomenon, which has been well documented in the literature (Kuo, Hull, Gupta, & Elby, 2013). However, explicitly prompting for explanation supported students in the incorporation of conceptual reasoning in their arguments. Simply revealing students' understanding and ideas improves the quality of the discourse and provides opportunities for the students to negotiate meaning.

Supporting argument from evidence

Prompts that targeted 'Engaging in argument from evidence' elicited argument sequences with multiple claim-data-warrant units as well as rebuttals and qualifiers. This is unsurprising as multiple studies have shown how tasks with certain argument features support the construction of arguments (Berland & McNeill, 2010). However, considering an argument task at this level reveals how upper-level undergraduate students interact with data provided to them. An example of this type of task and the resulting arguments is presented in Table 8.

This prompt required students to make a decision as to whether they agreed or disagreed with the statement and support their claim with evidence provided in the curriculum materials, which included heat capacity values and constant parameters for many chemical species. This prompt, then, explicitly provided students with data to consider and directed them to consider specific features of the data (molar heat capacity and chemical identity). Though it may seem intuitive that given data, students will use data to construct arguments, research has suggested that using actual data to make claims (or evaluate them) can be difficult for students (Kuhn, 1991; Brem & Rips, 2000). Furthermore, students can sometimes rely on explanations or conceptual reasoning and be over-confident in their claims. The arguments in this excerpt illustrate the difficulty that students encountered when interpreting the data to come to a consensus.

In contrast with previously discussed arguments, these claims and arguments appear quite tentative. This was evidenced by the qualifying words in the claims (e.g. “This statement appears to be *somewhat* true” or “The more complex does *not necessarily* mean larger”). All of the arguments used information provided in the data table to make their claims, but the way that the data was used and interpreted varied. There seem to be two difficulties with using this evidence to make a claim. The first difficulty was using the data to make a generalizable claim. This was captured by the students’ addition of qualifying words to the claim that they were evaluating. In the whole group discussion, Jerome modified the claim, arguing that it is not always true. Part of this difficulty clearly sourced from the data for monatomic hydrogen serving as an outlier, which was referenced by Rosalind, Dominique, and Reed. In none of the arguments that considered monatomic hydrogen did the students correctly incorporate it to qualify or support their claim. Though Jerome posits that the phase difference could explain why the heat capacity of monatomic hydrogen may not follow the trend, this line of reasoning does not seem to be adopted by other students and it is not used to evaluate the claim. At the very end of the whole class argument sequence, Instructor B models appropriate use of the exceptions (monatomic hydrogen) to evaluate the generalizability of the claim by qualifying that the “Statement is true if you are comparing similar phases.”

The second difficulty is reflected in the relative absence of warrants that explain how the data gives rise to the claim. Though the students clearly cite the data to evaluate the claim, they demonstrate a preoccupation with consideration of the exception. This is not a negative feature of argumentation as considering outliers is certainly a challenge in analyzing and interpreting data. However, explicitly connecting the data to the claim is a core task in constructing arguments. In this specific example, explicit consideration of how the data supported the claim may have helped them in considering data that did not. When provided with data, students cited the data, but did not use it to its fullest extent. That is, they did not use the data comprehensively to support the claim and they demonstrated difficulties in considering data that did not seem to fit. These difficulties reveal the need for more scaffolding to support engaging in argument from evidence.

The prompts and resulting arguments shown above illustrated how different types of prompts supported arguments that relied on a variety of evidence types. The information provided to the students as well as the practice targeted by the prompt drove these differences. Tasks targeting the completion of a mathematical operation effectively support the use of mathematical reasoning to reach an output, but possibly constrain students' conceptual reasoning if students are not asked to reflect on the meaning or utility of the result. Tasking students to construct explanations resulted in arguments that were much more revealing of their conceptual understanding, including evidence of interpretation and reflection. Providing students with experimental data resulted in similar evidence of students attempting to interpret the data, though it revealed a need for more explicit scaffolding to effectively support using data to make claims.

Supports and constraints of complex causal reasoning

The features of the curricular prompts that had the largest impact on the complexity of causal reasoning were the cognitive demand and the integration of content and practices. Specifically, the majority of multicomponent reasoning, the most complex, resulted from prompts with a cognitive demand of guided. Table 9 illustrates this effect, indicating that the majority of descriptive, relational, and linear causal arguments result from scripted integration (SI) prompts, while the majority of multicomponent arguments result from guided prompts (both GC and GI). In this Table 8, scripted tasks were split into scripted content and scripted integration, even though according to the original framework they are equivalent in cognitive demand. Qualitative comparison of prompts with differing cognitive demand and resulting arguments contributes to a better understanding of how these prompts influence students' reasoning. Students employed primarily relational reasoning; that is, they treated one variable or relationship as the sole cause of an outcome. This reasoning is sufficient in some situations, but it falls short for considering many thermodynamic problems.

Constraining complex, multicomponent reasoning

Tasks that targeted the practices of 'constructing explanations' and 'using and developing models' were highlighted in the previous section for supporting the use of conceptual reasoning to construct arguments. When these tasks were memorized or scripted and provided students with explicit instructions that required them to recall or use information previously encountered, students largely employed lower-level reasoning. This was the case even for prompts that required some interpretation of previously encountered material. That is, prompts with low cognitive demand potentially constrained students' use of complex reasoning. Consider the following prompts.

> Focus Question: A hot brick is placed into cold water in an isolated container. The final temperatures of the brick and water are identical. What is the total energy change in this process:
> a) Positive
> b) Negative
> c) Zero
> d) Cannot determine without further information
>
> [T2, Focus Question (Spencer et al., 2004)]

This question provides a phenomenon with associated temperature changes and prompts students to determine the total energy change. This prompt was coded as memorized content as it targets the concept of energy conservation, which students in an upper-level chemistry class have previously encountered. Though the prompt ideally requires some interpretation (e.g. the temperature change of the brick and water do not matter because it is in an isolated system), students could employ equations or principles from their prior knowledge to answer the question easily. The invocation of prior knowledge without simultaneous prompting for interpretation can limit their access to more complex reasoning.

(a) $Mg(s) + CO(g) + O_2(g) \rightarrow MgCO_3(s)$
(b) $MgO(s) + CO_2(g) \rightarrow MgCO_3(s)$
(c) $Mg(s) + C(s) + \frac{3}{2} O_2(g) \rightarrow MgCO_3(s)$
(d) $BaCO_3(s) \rightarrow BaO(s) + CO_2(g)$
(e) $CO(g) + \frac{1}{2} O_2(g) \rightarrow CO_2(g)$
(f) $C(s) + O_2(g) \rightarrow CO_2(g)$

In which of the above reactions is the product the result of the reaction of the elements that compose it, each of the elements being in their stable states at 1 bar? [T3, CTQ 12 (Spencer et al., 2004)]

In contrast to the previous prompt, this one provides an explicit script the students must follow to evaluate the chemical reactions presented to them. In response to this, students applied the script, employing descriptive reasoning. A prompt like this may seem like a straightforward way to develop students' understanding of the definition of enthalpy of formation. In effect, however, students applied the script without having a meaningful conversation about the conditions of an enthalpy of formation reaction. In this way, this prompt served to constrain students' access to more complex reasoning. Similar to the previous prompt, providing the students with a script without simultaneous prompting them to negotiate the script constrained their reasoning.

A particularly revealing prompt and resulting arguments are provided in Table 10. In this example, the prompt elicits relational reasoning from the students, but linear causal reasoning from the instructor. This type of question is important to explore as it could point to the role that expertise plays in interpreting what a question is asking for and what information is necessary to consider in order to answer it. This prompt explicitly elicited a calculation, which required use of an equation previously provided in the activity, earning it a scripted integration code. This problem could also have been solved using conceptual information provided just prior to the prompt (the internal energy of an ideal gas is dependent only on temperature). Excluding Instructor A's argument in classroom A, the arguments drew from the relationship between internal energy and temperature for an ideal gas. This type of reasoning was targeted by the prompt, evidenced by the prompts leading up to this one in which students constructed the concept of temperature dependence of internal energy. This was confirmed by Melody's backing in which she said "They really want us to know that energy is only a function of temperature."

Instructor A provided a more complex linear causal argument in which he incorporated the definition of an ideal gas as having no forces between particles. This argument was sophisticated and explicit, representing the type of reasoning that would ultimately be desired of students; however, the prompt did not serve to elicit this type of reasoning. Instructor A's argument illustrated how expertise prompted a more sophisticated argument in response to a prompt that did not elicit sophisticated reasoning from more novice students. Even at this level (upper-level undergraduate), it cannot be assumed that students will think more deeply than what is explicitly asked of them. That is, unless prompts are scaffolded by the instructor to support the type of reasoning demonstrated by the expert, students will not extend their reasoning beyond what is explicitly demanded by the prompt.

Supporting complex causal reasoning

Prompts that supported more complex causal reasoning made explicit multiple variables that needed to be considered by students and/or required the students to design something. Guided prompts required students to generate new information that was not provided to them. They often involved synthesizing prior knowledge, resulting in more complex reasoning. The guided content (GC) example found in Table 11 illustrates the effect of explicitly prompting students to consider multiple variables to make a prediction.

This prompt required students to make a prediction for which they needed to synthesize their prior knowledge. This task was situated at the beginning of an activity that introduces Gibbs energy and Helmholtz energy as concepts for determining the direction of chemical processes. Further, students needed to consider and weigh at least two variables (enthalpy change and entropy change) to answer the prompt. As a result, their arguments revealed how they weighed and considered these two variables. Liam claimed that the reaction would not proceed due to a decrease in entropy and drew on his knowledge that spontaneity is determined by entropy, which ultimately determines direction. Brian drew on knowledge of a chemical reaction that he was familiar with that could be applied to the hypothetical reaction in the question in order to justify that the reaction proceeded forward. Jerome claimed that "maybe the reaction will occur." His reasoning was that it was possible for the bond strength gains to outweigh entropy losses. In all three of these arguments, students were considering and weighing the change in entropy and enthalpy. In order to do this, they incorporated prior knowledge in the form of additional variables, such as spontaneity or an example chemical reaction. In Jerome's case, weighing both variables resulted in a less conclusive claim that the reaction was possible. The last argument served to synthesize multiple small groups' answer to this question, after which Instructor B introduced the concept of Gibbs energy as direction determining. Explicitly tasking students with multiple variables to consider to predict something unknown or synthesize something new can support the use and development of multicomponent reasoning.

Another feature of prompts that promoted more complex causal reasoning was requiring the students to design something. The prompt in Table 12 required students to plan an investigation that would allow them to determine heat capacity for a constant volume process. This prompt provided no specific steps the students needed to follow to reach an output and specifically targeted the practice of ‘planning and carrying out an investigation.’ The argument sequence generated demonstrates complex causal reasoning.

In the initial argument, Reed claimed that a bomb calorimeter could be used to determine C_v. He considered multiple variables and described how he intended to manipulate those variables to measure a value for molar heat capacity for a constant volume process. Callum sought elaboration on how Reed can measure internal energy. Quentin introduced the consideration of “base changes” and “difficult reaction processes”, for which he provided combustion as an example. Finally, Callum [incorrectly] incorporated the concept of constant density so as to have constant volume. Quentin and Callum contributed new variables for consideration by the whole group. The guided level of cognitive demand and targeted practice of ‘planning and carrying out an investigation’ in this question facilitated the use of complex causal reasoning and consideration of multiple variables. Higher cognitive demand along with scaffolding to prompt consideration of multiple variables can support the type of complex causal reasoning that is desired of upper-level undergraduate chemistry students.

Conjecture Map

Incorporating the findings above into our conjecture map illustrates how the results help us answer our two-part research question. Figure 5 shows that components of the task structure impacted the arguments generated in both classes considered in this study. Specifically, the cognitive demand was shown to impact the complexity of causal reasoning students employed in their arguments, while which practices were targeted influenced how students used information provided in the questions to construct their arguments. More importantly, the targeted practices influenced how revealing an argument was of students' conceptual understanding. When multiple practices were targeted (e.g. 'using mathematics and computational reasoning' and 'constructing explanations'), resulting arguments included consideration of the meaning of the mathematics and connections between the mathematics and the phenomenon being investigated. Explicitly targeting the practice of 'engaging in argument from evidence' by providing students with data elucidated difficulty students had with justifying claims with the data.

Discussion

Results from this study show that the curricular prompts can support and constrain students' use of evidence and the complexity of their reasoning. Findings regarding the complexity of causal reasoning are consistent with previous research on the conditions of learning environments for supporting argumentation and reasoning in primary and secondary classrooms (Duschl & Osborne, 2002; Jimenez-Aleixandre, 2007; Berland & McNeil, 2010). In particular, tasks with higher cognitive demand that require collaboration support argumentation and complex reasoning. A meaningful understanding of physical chemistry requires the use of complex causal models. However, students have a difficult time developing those models (Smith, Carey, & Wiser, 1985; Perkins & Grotzer, 2005; Sevian & Talanquer, 2014). One desired goal, then, of an inquiry curriculum for physical chemistry is to support the construction of arguments that use complex causal models. From a sociocultural perspective, this means that the learning environment, including the curricular materials and classroom discourse, will mediate the degree to which this goal is achieved. As a curriculum supports the collaborative building of these complex arguments, individual students are able to then internalize complex causal models into their own reasoning (Mercer & Howe, 2012). Findings from this study showed that this inquiry curriculum primarily supported lower-level causal reasoning, but also revealed features that supported more complex reasoning. Explicitly providing multiple variables that the students must consider promoted consideration of those variables. Pairing the explicit instructions for what to consider with a task requiring them to generate something new can scaffold multicomponent reasoning. In this way, prompts are cognitively demanding by requiring students to engage authentically (i.e. making predictions) and support students in building an understanding of how variables are related and interact dynamically to give rise to phenomena. The second feature that promoted complex reasoning was the task to design something. With little information provided, the students were forced to draw on their prior knowledge and synthesize information encountered previously in the class.

These features identified in this study that support complex reasoning contrast with Jimenez-Aleixandre's (2007) and Berland and McNeil's (2010) call for little to no scaffolding in order to support complex argumentation. Part of this contradiction derives from the difficulty of this particular content for students, even upper-level tertiary students. When students are given little to no scaffolding, they either generate an argument that is not very meaningful or no argument at all. This is why the difficulties students encountered in 'engaging in argument from evidence' revealed the need for more explicit scaffolding to support these practices. These areas that demand more scaffolding were elucidated through a qualitative analysis of prompts that targeted different practices and the resulting arguments. It was found that the practice(s) targeted by the question afford certain types of reasoning and potentially constrain others. Of particular interest in physical chemistry is how students relate mathematics to conceptual understanding of phenomena. When scripted tasks are assigned that target only the 'use of mathematics and computational reasoning', students' arguments used the equations provided and describe steps taken to arrive at a final output. The resulting arguments reveal little reflection on the mathematics or connection between the mathematics and physical phenomena. However, when scripted tasks targeted both using mathematics and constructing explanations, the resulting arguments did reveal the desired reflection and connections between the mathematics and phenomena. This means that simply the way the prompt is written, namely, explicit calls for explanation, can elicit the desired conceptual reasoning associated with effective argumentation (Jimenez-Aleixandre, 2007). In accordance with this reasoning, we expected that providing students with experimental data and explicitly calling for engaging in argument should promote the practice of engaging in argument from evidence. As has been demonstrated previously, this is not always the case (Kuhn, 1991; Brem & Rips, 2000). Students encountered difficulties in using the data. Specifically, students had difficulty incorporating all the data to make a generalizable claim and in demonstrating how the data supports the claim.

The POGIL thermodynamics inquiry activities investigated in this study largely include scripted tasks, which might limit students' opportunities to use more complex causal reasoning. The majority of scripted questions can partially be explained by principles guiding how POGIL activities are designed. The POGIL approach draws on directed questions, which can be answered directly with provided information, to introduce a concept or explore a model, followed by convergent questions, which require multiple pieces of information and some level of synthesis to answer, to further construct and apply concepts. Divergent questions are a third type of question included in the POGIL approach, which are open-ended with multiple possible solutions (Hanson, 2006). Within the TAGS framework, these directed questions likely explain the abundance of scripted questions. Coding prompts for these three types of questions (directed, convergent, and divergent) revealed a lack of divergent questions and the presence of questions that could not be categorized into these three types (mathematical procedural questions). A more detailed analysis of this aspect of the materials is addressed in a separate manuscript (Authors, in preparation).

Implications for Research and Practice

Results from students' interaction with the physical chemistry inquiry activities provide implications for revisions to the activities and for the design of inquiry activities for other instructional approaches. In particular, to promote the type of discourse that supports students' development of the conceptual tools necessary to learn thermodynamics, more prompts pairing the 'use of mathematics and computational thinking' with other practices are necessary. There should be no inquiry cycles targeting the use of mathematics and computational thinking that do not also include explicit prompts for students to justify, reflect, or explain their reasoning. As upper-level students did not use reasoning that extended beyond what the question demanded, explicitly prompting for this is key to supporting desirable scientific discourse. Additionally, targeting a variety of practices is necessary. The arguments shown here that resulted from a task targeting "engaging in argument from evidence" revealed important obstacles encountered and strategies used by students when using data. These difficulties reveal the need for more explicit scaffolding. Particularly, calling for explanations of how the data give rise to the claim can support the construction of warrants, while more opportunities to consider outliers are required to support the construction of generalizable claims.

Similarly, in order to support discourse that employs complex causal models, more questions with guided cognitive demand are necessary. The majority of the prompts in these activities were scripted. We are not arguing that scripted prompts must be eliminated or do not have a role in this curriculum, but rather that they are not sufficient for supporting the development of complex causal reasoning. To this end, we suggest that the "application" portion of the POGIL cycle be expanded to include most, if not all, guided prompts. For any type of inquiry activities that are targeting the development of complex causal models, there must be opportunities in the form of tasks with high cognitive demand to support that development. The results from this work indicate that explicitly prompting students to consider multiple variables and to design something are promising features to incorporate into those tasks.

The qualitative analysis of resulting discourse was perhaps the most revealing of the need for more research into how students interact with and use the curricular materials to construct arguments. Further investigation of how students' prior knowledge informs their argumentation practices is necessary to understand how students generate arguments with varying complexity in response to a single prompt. While this work considered only the task structure form of our embodied conjecture, we recognize that it is not independent from the participation structure and discursive structure. In the case of this work, considering only task structure supported investigation of specific relationships of interest, but it also makes clear the need to better understand the relationship between embodiments (task structure, participation structure, and discursive structure) and between embodiments, mediating processes, and outcomes.

The Task Analysis Guide for Science was especially useful in this work for considering the prompts with which students were working. Comparing the prompts, as categorized by TAGS, to the students' discursive products provided insight into how students use and respond to prompts. We argue that this framework has potential for closing the gap between an expert's intended outcome and novices' classroom experience. That is, instructors and curriculum designers are encouraged to use this framework to evaluate tasks that they write from the perspective of the students who may be encountering these tasks. The use of a conjecture map served to guide investigation that can aid instructors and curriculum designers in this evaluation.

References

Asterhan, C. S. C., & Schwarz, B. B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of Educational Psychology*, *99*(3), 626–639.

Banks, G., Clinchot, M., Cullipher, S., Huie, R., Lambertz, J., Lewis, R., Ngai, C., Sevian, H., Szteinberg, G., Talanquer, V, & Weinrich, M. (2015). Uncovering Chemical Thinking in Students ' Decision Making: A Fuel- Choice Scenario. Journal of Chemical Education.

Becker, N., Rasmussen, C., Sweeney, G., Wawro, M., Towns, M., & Cole, R. (2013). Reasoning using particulate nature of matter: An example of a sociochemical norm in a university-level physical chemistry class. *Chemistry Education Research and Practice*, *14*, 81. http://doi.org/10.1039/c2rp20085f

Becker, N., Stanford, C., Towns, M., & Cole, R. (2015). Translating across macroscopic, submicroscopic, and symbolic levels: the role of instructor facilitation in an inquiry-oriented physical chemistry class. *Chem. Educ. Res. Pract.* http://doi.org/10.1039/C5RP00064E

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*, *94*(5), 765–793.

Bismack, A. S., Arias, A. M., Davis, E. A., & Palincsar, A. S. (2015). Examining Student Work for Evidence of Teacher Uptake of Educative Curriculum Materials. *Journal of Research in Science Teaching*, *52*(6), 816–846.

Brem, Sarah, K., & Rips, Lance, J. (2000). Explanation and Evidence in Informal Argument. *Cognitive Science*, *24*(4), 573–604.

Bricker, L. a., & Bell, P. (2008). Conceptualizations of argumentation from science studies and the learning sciences and their implications for the practices of science education. *Science Education*, *92*, 473–498. http://doi.org/10.1002/sce.20278

Brown, N. J. S., Nagashima, S. O., Fu, A., Timms, M., & Wilson, M. (2010). A Framework for Analyzing Scientific Reasoning in Assessments. Educational Assessment, 15(3-4), 142–174.

Cetin, P. (2014). Explicit argumentation instruction to facilitate conceptual understanding and argumentation skills. *Research in Science & Technological Education*, *32*(1), 1–20.

Chin, C., & Osborne, J. (2010). Students' questions and discursive interaction: Their impact on argumentation during collaborative group discussions in science. *Journal of Research in Science Teaching*, *47*(7), 883–908.

Cho, K.-L., & Jonassen, D. (2002). The effects of argumentation scaffolds on argumentation and problem solving. *Educational Technology Research and Development*, *50*(3), 5–22.

Cobb, P., & Whitenack, J. W. (1996). A method for conducting longitudinal analyses of classroom videorecordings and transcripts. Educational Studies in Mathematics, 30, 213–228.

Cobb, P., Stephan, M., McClain, K., & Gravemeijer, K. (2001). Participating in Classroom Mathematical Practices. *Journal of the Learning Sciences*, *10*, 113–163. http://doi.org/10.1207/S15327809JLS10-1-2_6

Cole, R., Becker, N., Towns, M., Sweeney, G., Wawro, M., & Rasmussen, C. (2012). Adapting a methodology from mathematics education research to chemistry education research: Documenting collective activity. *International Journal of Science and Mathematics Education*, *10*(1), 193–211.

Cullipher, S., Sevian, H., & Talanquer, V. (2015). Reasoning about benefits, costs, and risks of chemical substances: Mapping different levels of sophistication. Chem. Educ. Res. Pract., 16, 377–392.

Daubenmire, P. L., Bunce, D. M., Draus, C., Frazier, M., Gessell, A., & van Opstal, M.,T. (2015). During POGIL implementation the professor still makes a difference. Journal of College Science Teaching, 44(5), 72-81.

Duschl, R. a., & Osborne, J. (2002). Supporting and Promoting Argumentation Discourse in Science Education. *Studies in Science Education*, *38*(1), 39–72. http://doi.org/10.1080/03057260208560187

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. Science Education, 88(6), 915–933.

Erduran, S. (2007). Methodological foundations in the study of argumentation in science classrooms. In *Argumentation in science education* (pp. 47–69).

Fay, M. E., Grove, N. P., Towns, M. H., & Bretz, S. L. (2007). A rubric to characterize inquiry in the undergraduate chemistry laboratory. Chemistry Education Research and Practice, 8(2), 212–219.

Garcia-Mila, M., Gilabert, S., Erduran, S., & Felton, M. (2013). The Effect of Argumentative Task Goal on the Quality of Argumentative Discourse. *Science Education*, *97*(4), 497–523. http://doi.org/10.1002/sce.21057

Hanson, D. (2006). Instructor's guide to process-oriented guided-inquiry learning. Lisle, IL: Pacific Crest.

Jimenez-Aleixandre, M. P. (2007). Designing argumentation learning environments. *Argumentation in Science Education*, 91–115. http://doi.org/10.1007/978-1-4020-6670-2

John-Steiner, V., & Mahn, H. (1996). Sociocultural approaches to learning and development: A Vygotskian framework. *Educational Psychologist*, *31*(3-4), 191–206. http://doi.org/10.1080/00461520.1996.9653266

Khishfe, R. (2014). Explicit Nature of Science and Argumentation Instruction in the Context of Socioscientific Issues: An effect on student learning and transfer. *International Journal of Science Education*, *36*(6), 974–1016.

Kuhn, D. (1991). The skills of argument. Cambridge: Cambridge University Press.

Kulatunga, U., Moog, R. S., & Lewis, J. E. (2013). Argumentation and participation patterns in general chemistry peer-led sessions. Journal of Research in Science Teaching, 50(10), 1207–1231.

Kulatunga, U., Moog, R. S., & Lewis, J. E. (2014). Use of Toulmin's Argumentation Scheme for Student Discourse to Gain Insight About Guided Inquiry Activities in College Chemistry. Journal of College Science Teaching, 43(5), 78–86.

Kulatunga, U., & Lewis, J. E. (2013). Exploration of peer leader verbal behaviors as they intervene with small groups in college general chemistry. Chemistry Education Research and Practice, 14(4), 576–588.

Kuo, E., Hull, M. M., Gupta, A., & Elby, A. (2013). How students blend conceptual and formal mathematical reasoning in solving physics problems. *Science Education*, *97*(1), 32–57. http://doi.org/10.1002/sce.21043

McNeill, K. L., & Krajcik, J. S. (2012). Supporting grade 5–8 students in constructing explanations in science: The claim, evidence, and reasoning framework for talk and writing. Boston: Pearson Education

Mercer, N. (2007). Sociocultural discourse analysis: analysing classroom talk as a social mode of thinking. *Journal of Applied Linguistics*, *1*(2), 137–168. http://doi.org/10.1558/jal.v1i2.137

Mercer, N., & Howe, C. (2012). Explaining the dialogic processes of teaching and learning: The value and potential of sociocultural theory. *Learning, Culture and Social Interaction*, *1*(1), 12–21. http://doi.org/10.1016/j.lcsi.2012.03.001

Mercer N., Dawes L., Wegerif R. and Sams C., (2004), Reasoning as a Scientist: Ways of Helping Children to Use Language to Learn Science, Br. Educ. Res. J., 30, 359–377.

Moon, A., Stanford, C., Cole, R., & Towns, M. (2016). The nature of students' chemical reasoning employed in scientific argumentation in physical chemistry. *Chem. Educ. Res. Pract.*, *17*, 353. http://doi.org/10.1039/C5RP00207A

National Research Council. A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Idea; National Academies Press: Washington, DC, 2012.

National Research Council. Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering; National Academies Press: Washington, DC, 2012.

Loughlin, M. O. (1992). Rethinking Science Education : Beyond Piagetian Constructivism Toward a Sociocultural Model of Teaching and Learning *. *Journal of Research in Science Teaching*, *29*(8), 791–820. http://doi.org/10.1002/tea.3660290805

Perkins, D. N., & Grotzer, T. a. (2005). Dimensions of Causal Understanding: The Role of Complex Causal Models in Students' Understanding of Science. Studies in Science Education, 41(1), 117–165.

Rasmussen, C., and Stephan, M. (2008). A methodology for documenting collective activity. in A. E. Kelly, R. A. Lesh and J. Y. Baek (ed.), Handbook of innovative design research in science, technology, engineering, mathematics (STEM) education (pp. 195–215). New York: Taylor and Francis.

Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education*, *92*, 499–525.

Sadler, T. D. (2009). Situated learning in science education: socio-scientific issues as contexts for practice. Studies in Science Education, 45, 1–42.

Sampson, V., & Clark, D. (2011). A Comparison of the Collaborative Scientific Argumentation Practices of Two High and Two Low Performing Groups. *Research in Science Education*, *41*(1), 63–97. http://doi.org/10.1007/s11165-009-9146-9

Sandoval, W. (2014). Conjecture Mapping : An Approach to Systematic Educational Design Researc. *Journal of the Learning Sciences*, *23*(1), 18–36.

Sevian, H., & Talanquer, V. (2014). Rethinking chemistry: a learning progression on chemical thinking. Chemistry Education Research and Practice, 15(1), 10–23.

Smith, C., Carey, S., & Wiser, M. (1985). On differentiation: A case study of the development of the concepts of size, weight, and density. Cognition, 21, 177–237.

Spencer, J. N., Moog, R. S., & Farrell, J. J. (2004). Physical Chemistry: A Guided Inquiry Thermodynamics. Boston: Houghton Mifflin Company.

Stanford, C., Moon, A., Towns, M., & Cole, R. (2016). Analysis of Instructor Facilitation Strategies and Their Influences on Student Argumentation: A Case Study of a Process Oriented Guided Inquiry Learning Physical Chemistry Classroom. *Journal of Chemical Education*, *93*(9), 1501–1513. http://doi.org/10.1021/acs.jchemed.5b00993

Sweller, J. (1988). Cognitive Load During Problem Solving : Effects on Learning. *Cognitive Science*, *12*, 257–285.

Szteinberg, G., Balicki, S., Banks, G., Clinchot, M., Cullipher, S., Huie, R., Lambertz, J., Lewis, R., Ngai, C., Weinrich, M., Talanquer, V., & Sevian, H. (2014). Collaborative Professional Development in Chemistry Education Research: Bridging the Gap between Research and Practice. Journal of Chemical Education, 91(9), 1401–1408.

Talanquer, V. (2010). Exploring Dominant Types of Explanations Built by General Chemistry Students. *International Journal of Science Education*, *32*(18), 2393–2412.

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.

Toulmin, S. E. (1958). The uses of argument. Cambridge University Press.

van Roon, P. H., van Sprang, H. F., & Verdonk, A. H. (1994). International Journal of Science ‘Work' and 'Heat': on a road towards thermodynamics 'Work' and 'Heat': on a road towards thermodynamics. *International Journal of Science Education*, *16*(2), 131–144.

Wertsch, J. (1991). Voices of the mind: A sociocultural approach to mediated action. Cambridge, MA: Harvard University Press.

Conflict of Interest

The authors declare no conflict of interest, relationships, financial, or otherwise, that may have affected this study.

Table 1. Demographics about participants and settings

	Classroom A		**Classroom B**	
Instructor Experience	7 years of implementing POGIL		10 years of implementing POGIL	
Setting	Private College, ~1000 students		Public University, ~14,000 students	
	Physical Chemistry I & II		Thermodynamics	
	Spencer, Moog, and Farrell POGIL materials			
Number of Participants	10 students	3 Females	18 students	5 Females
		7 Males		13 Males
Participant Demographics	Second through Fourth years		Third & Fourth years	

Table 2. Content targeted by each POGIL activity covered in both classrooms

POGIL Activity	Content
T1	Work
T2	The First Law of Thermodynamics
T3	Enthalpy

T4	Heat Capacity
T5	Temperature Dependence of Enthalpy of Reaction
T6	Entropy
T7	Entropy Changes as a Function of Temperature
T9	Gibbs Energy and Helmholtz Energy

Table 3. Modes of reasoning from Chemical Thinking Learning Progression (CTLP) (Adapted from Sevian and Talanquer, 2014)

Categories	Features	Example
Descriptive	• Salient properties are recognized • Explicit properties are verbalized • Phenomenon is instantiation of reality • Reasoning based on experiences from daily life	"Work is done"
Relational	• Explicit and implicit properties are highlighted • Spatial and temporal relationships are identified • Phenomenon is effect of single variable (no mechanism)	"Because the volume changes, work is done"
Linear	• Mechanisms proposed that involve linear cause-effect relationships • Step-wise mechanism	"The reaction produces more moles resulting in an increased volume so the system does work on the surroundings"
Multi-component	• Mechanism considers and weighs effects of several variables	"Reaction produces more moles increasing pressure pushing the piston up doing work on the surroundings. Exothermic reaction releases energy, which can go into doing work on the surroundings."

Table 4. TAGS framework (replicated from Tekkumru-Kisa et al., 2015)

Cognitive Demand Levels

		Scientific Practices (e.g., argumentation and investigation)	Science Content (i.e., scientific body of knowledge)	Integration of Content and Practices
5	Doing Science Tasks			**Doing Science (DS)** Engaging in practices to make sense of content and recognize how scientific body of knowledge is developed

4	Tasks involving guidance for understanding			**Guided Integration (GI)** Guidance for working with practices tied to a particular content
3		**Guided Practices (GP)** Being guided for understanding practices	**Guided Content (GC)** Being guided for understanding particular content	
2	Tasks involving scripts	**Scripted Practices (SP)** Following a script to work on practices	**Scripted Content (SC)** Following a script about a content	**Scripted Integration (SI)** Following a script to work on practices tied to content
1	Memorization tasks	**Memorized Practices (MP)** Reproducing definitions/explanations of practices	**Memorized Content (MC)** Reproducing definitions, formulas, or principles about particular content	

Table 5. Interpretations of cognitive demand levels used to code POGIL prompts

Cognitive demand	Interpretation used for coding
Doing Science	Requires students to engage in scientific practices and scientific knowledge construction relatively unaided
Guided	Requires the generation of new knowledge, information, practice
Scripted	Requires students to follow a script to complete
Memorized	Requires students to regurgitate provided information

Table 6. Arguments illustrating students' use of mathematical reasoning to complete a mathematical task

Prompt Let $dU = C_v dT$ and rearrange equation (4) [$dU = TdS - PdV$]to provide an expression for *dS* for one mole of an ideal gas in terms of T, V, and Cv. [T7, CTQ 4 (Spencer et al., 2004)]	
Classroom A	**Claim:** $dS = C_v dT + PdV$ **Data:** $dU = C_v dT$ (implied) **Warrant:** We just substitute this one into *dU* and then rearrange this equation right here (Mark).
Classroom B	**Claim:** $dS = \frac{C_v}{T} dT + \frac{R}{V} dV$

	Data: $dU = C_v dT = TdS - PdV$ **Warrant:** so substitute in $C_v dT$ for U. add PV down here to that side. And divide by T. Flip it around, Well, we want volume and temperature, so I need to get rid of pressure. Replace pressure with its definition. nRT divided by V. So I've got nRT over V times 1 over TdV, T's cancel out. Actually, we did molar volume, we don't need the n.

Table 7. Arguments illustrating how students respond to a mathematical task that also targets constructing explanations

Prompt Consider a constant pressure process in which $\Delta_r C_p$ is greater than zero and does not depend on temperature. If temperature is raised, does the value of $\Delta_r S$ increase, decrease, stay the same, or is it impossible to determine? Explain your reasoning. [T7, CTQ 11 (Spencer et al., 2004)]	
Classroom A	**Claim:** It is impossible to determine what will happen to the value of $\Delta_r S$ when you raise the temperature. (Melody) **Data:** $\Delta_r S\ T_2 - \Delta_r S\ T_1 = \Delta_r C_p \ln T_2/T_1$ (book) **Warrant:** It depends on whether the natural log of T_2/T_1 is greater or less than 1. (Melody) **Backing:** Or even if it's not such a tiny amount, if it's much bigger, because like the natural log of 1.5 is less than 1. (Melody)
Classroom B	**Claim:** The change in entropy of the reaction gets bigger. (Quentin) **Data:** Constant pressure process in which $\Delta_r C_p$ is greater than zero, and doesn't depend on temperature. The temperature increases. (POGIL Materials) **Warrant:** You're increasing your final, then you're going to receive a bigger number (inaudible) (math of logs). (Quentin)

Table 8. Arguments generated in response to task targeting 'engaging in argument from evidence.'

Prompt Critique the following statement: The more complex the species, the larger is $\bar{C}$	
Classroom B: Small Group	**Claim:** *The statement* [The more complex the species the larger is C_p^o and the larger in C_p^o with increasing temperature.] *appears to be somewhat true*. (Thaddeus) **Data:** Table 4, page 88 (POGIL Materials) **Qualifier:** With the exception of monatomic hydrogen.

	(Rosalind) **Data for Qualifier:** *Maybe, the fact that it's a gas instead of a solid* (Jerome) **Rebuttal:** *But H_2 is a gas. The only difference is that it's monatomic instead of diatomic.* (Rosalind) **New Claim:** Cp increases with temp, so long as b and c are really small. (Thaddeus/Rosalind) **Data:** Table 4, page 88 (POGIL Materials) **Warrant:** *Well Cp increase with temperature is true right?* (Thaddeus)
Classroom B: Whole Class	**Claim:** *Not sure that the change* [in heat capacity with increasing temperature] *is always larger for the more complex species* (Jerome) **Data:** Table 4, page 88 (POGIL Materials) **Warrant:** *It did seem to be more complex species, there was a larger initial value. Assuming in the case of hydrogen gas, it's ionized at that point, so it's the same amount of ionized gas, so I guess you can consider that complex. But the more interaction going on such, despite the fact that it's ionizing.* (Jerome) **Rebuttal:** *It's not ionizing, it's just atomic hydrogen, so you don't have to, there's no charge on it.* (Instructor B) **Alternate Claim:** *We agreed with the statement* [The more complex the species the larger is Cp^o and the larger in Cp^o with increasing temperature.] (Dominique) **Alternate Data:** Table 4, page 88 (POGIL Materials) **Qualifier:** *The only exception is when you have small, uncomplex molecules, then it's not dependent on temperatures* (Dominique) **Alternate Claim:** *the more complex does not necessarily mean larger.* (Reed) **Alternate Data:** Table 4, page 88 (POGIL Materials) **Alternate Warrant:** *carbon graphite would be more complex than hydrogen, just bigger molecule and everything. But it's got a lower cp.* (Reed) **Qualifier:** Statement is true if you are comparing similar phases. (Instructor B)

Table 9. The percentage of arguments from each mode of reasoning resulting from each TAGS label (Values given in percentages)

TAGS	Descriptive	Relational	Linear	Multicomponent
Memorized content (MC)	2	3	1	0
Scripted content (SC)	41	14	3	7
Scripted integration (SI)	44	64	70	7
Guided content (GC)	5	4	5	43
Guided integration (GI)	7	15	20	43

Table 10. Student and instructor arguments in response to scripted prompt

Prompt Calculate ΔU for an isothermal process for an ideal gas in which the pressure increases from 1 bar to 10 bar. [T4, CTQ 13 (Spencer et al., 2004)]	
Classroom B: Small Group	**Claim:** Δ U = 0 (Qi) **Data:** It is isothermal. (Melody/book) **Warrant:** Because isothermal means Δ T is zero. (Melody) **Claim:** Δ U = 0 (Qi) **Data:** Isothermal, Δ T = 0 (Melody) **Warrant:** *For an ideal gas, the energy is only dependent on temperature.* (Melody) **Backing:** *They really want us to know that energy is only a function of temperature.* (Melody) **Claim:** The energy of an ideal gas is a function of the temperature only. (Instructor A) **Data:** Temperature is a direct measure of the average kinetic energy. (Instructor A) **Data:** Ideal gas does not have forces between particles. (Melody) **Warrant:** But if there are no forces between the particles, which is what an ideal gas is, we assume there are no forces between the particles, that means there is no potential energy, so every energy change is a kinetic energy change, so it all goes back to the fundamental idea of what an ideal gas is. (Instructor A) **Backing:** Some thermal energy can go into vibrational and rotational modes, but most goes into translational modes (Instructor A)
Classroom B: Whole Class	**Claim:** ΔU = 0 for an isothermal process for an ideal gas in which the pressure increases from 1 bar to 10 bar. (boards) **Data:** isothermal process (Male)

	Warrant: delta T = 0, therefore delta U = 0 (Instructor B)

Table 11. Student arguments in response to guided content prompt

<table>
<tr><td colspan="2">Prompt
Suppose a given chemical mixture has the potential to produce products so that the sum of the bond strengths is larger than those of the reactants but that the number of moles of reactant gases are decreased. Will the reaction occur?</td></tr>
<tr><td>Classroom B:
Small Group</td><td>Claim: Reaction will not go. (Liam)
Data: More moles of reactant, less moles of product (entropy decreases). (Liam)
Warrant: Spontaneity determined by entropy to determine direction (reaction won't go towards a lower entropy). (Liam)
Claim: Reaction does occur. (Brian)
Data: Sum of products bond strengths is larger than reactants, number of moles reactant decreases. (POGIL Materials)
Warrant: So the combustion of hydrogen forms water so you have 1.5 moles to every mole of product and we know that ΔH for that is negative release of energy and we know that reaction does occurs so we know that's ΔS is positive. (Brian)

Claim: Maybe the reaction will occur. (Jerome)
Data: Sum of products bond strengths is larger than reactants, number of moles reactant decreases. (POGIL Materials)
Warrant: We argued if your bond strength gains, or if your change, if it's energetically favorable enough, your product, your bond strength gains, then it will outweigh any uh, entropy losses so it could still happen. (Jerome)

Claim: The reaction will not occur. (2 groups)
Data: Sum of products bond strengths is larger than reactants, number of moles reactant decreases. (POGIL Materials)
Data: Entropy of the system is negative (2 groups). (entropy won't increase because there are fewer moles and greater bond strength)
Warrant: Reaction will not occur is entropy of the system is negative. (Instructor B/Caprice)
Rebuttal Claim: Only entropy of the universe determines spontaneity (Instructor B/Kayden)
Qualifier: Gibbs energy is necessary to consider the perspective of the system (Instructor B)

Rebuttal Data: $\Delta H < 0$, $\Delta S < 0$ (Instructor B)
Rebuttal Warrant: One favors reaction, the other disfavors. (Instructor B/Kayden)
Rebuttal Backing: Must consider magnitudes. There is not enough information to determine if the reaction will go or not</td></tr>
</table>

	(Instructor B)

Table 12. Student arguments in response to a guided integration prompt

Prompt Describe a process that could be used to determine $\bar{C}$v. [T4, CTQ 9 (Spencer et al., 2004)]	
Classroom B: Small Group	**Claim:** A bomb calorimeter could be used to determine C_v. (Reed) **Data:** By using a standard, then you know U. **Data:** Bomb is constant volume (Reed) **Data:** C_v equals [writes (du/dt)] (Reed/Instructor B) **Warrant:** so you can solve for the change for U, or the, you can solve for the U over dT, so you can solve for the change in energy with respect to temperature … {additional discussion} Because you're going to make the temperature change. And you know what your change in energy is, because you know how much energy you put in, and assuming you know how much it used, then you know much is used. (Reed) **Warrant:** We use the bomb because that gives me the process at constant volume. (Instructor B) **Request for clarification:** How would you know how much energy you used? (Callum) **Clarification:** Use standard mass to know how much internal energy there is. (Reed) **Rebuttal:** you're not really accounting for base changes or difficult reaction processes. (Quentin) **Rebuttal Data:** you're burning a substance so you're going to have combustion, you're going to be breaking down bonds and everything. (Quentin) **Rebuttal Warrant:** So you're not really accounting for the molar heat capacity, isn't that just increasing the temperature of a substance by a certain amount? (Quentin) **Counter Claim:** Apply a certain amount of energy to a substance and measure the temperature change [to determine heat capacity]. (Quentin) **Data/Warrant:** You need to know how much of the substance you have (Quentin) **Data/Warrant:** Assuming density doesn't change. He said the volume has to be constant. (Callum)

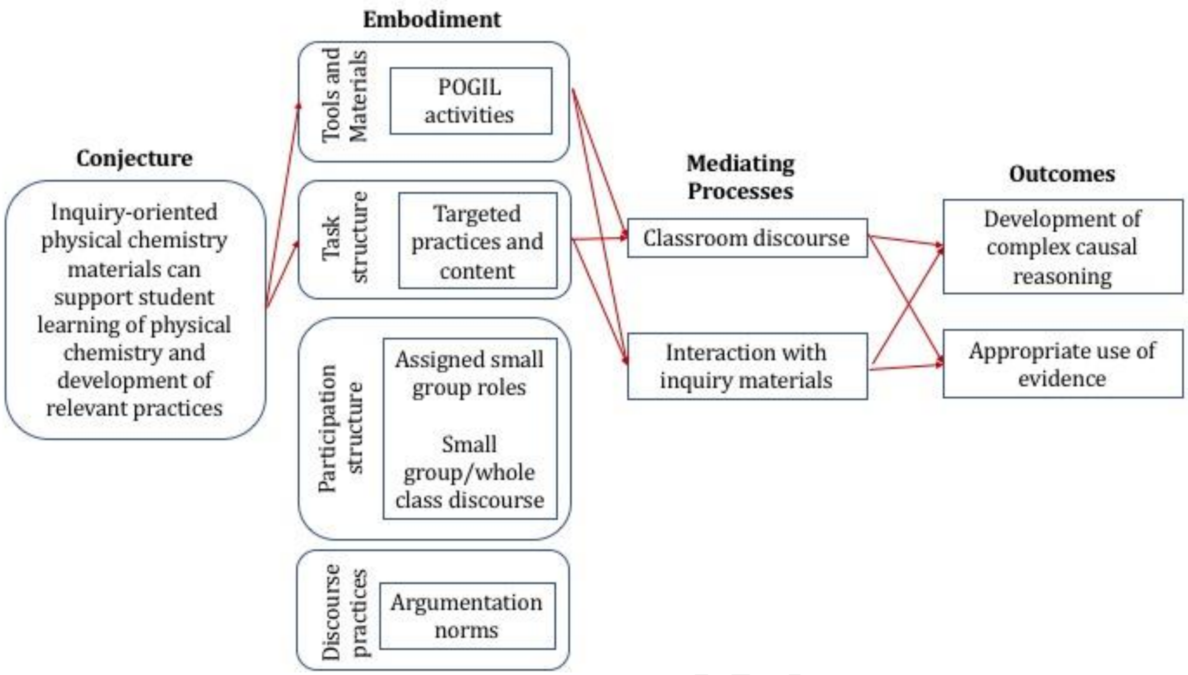
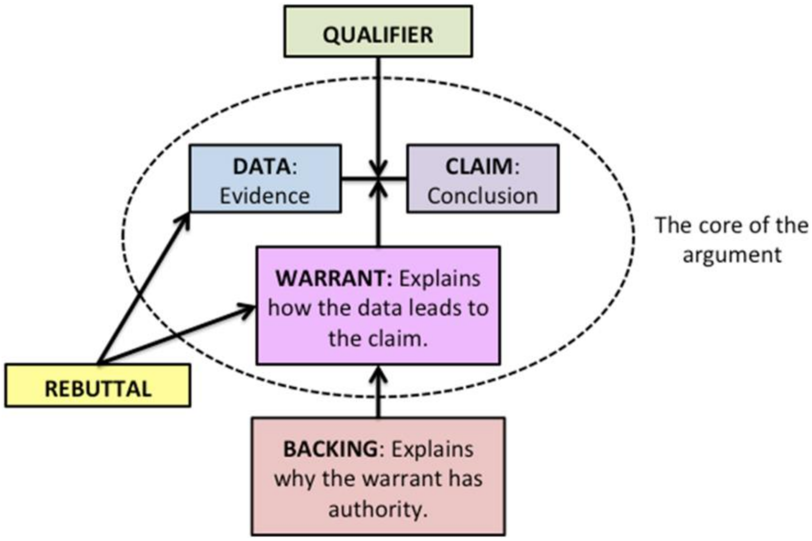

Embodiment
Tools and Materials
POGIL activities
Conjecture
Inquiry-oriented physical chemistry materials can support student learning of physical chemistry and development of relevant practices
Task structure
Targeted practices and content
Mediating Processes
Classroom discourse
Interaction with inquiry materials
Outcomes
Development of complex causal reasoning
Appropriate use of evidence
Participation structure
Assigned small group roles
Small group/whole class discourse
Discourse practices
Argumentation norms

Figure 1


QUALIFIER
DATA: Evidence
CLAIM: Conclusion
The core of the argument
WARRANT: Explains how the data leads to the claim.
REBUTTAL
BACKING: Explains why the warrant has authority.
Figure 2

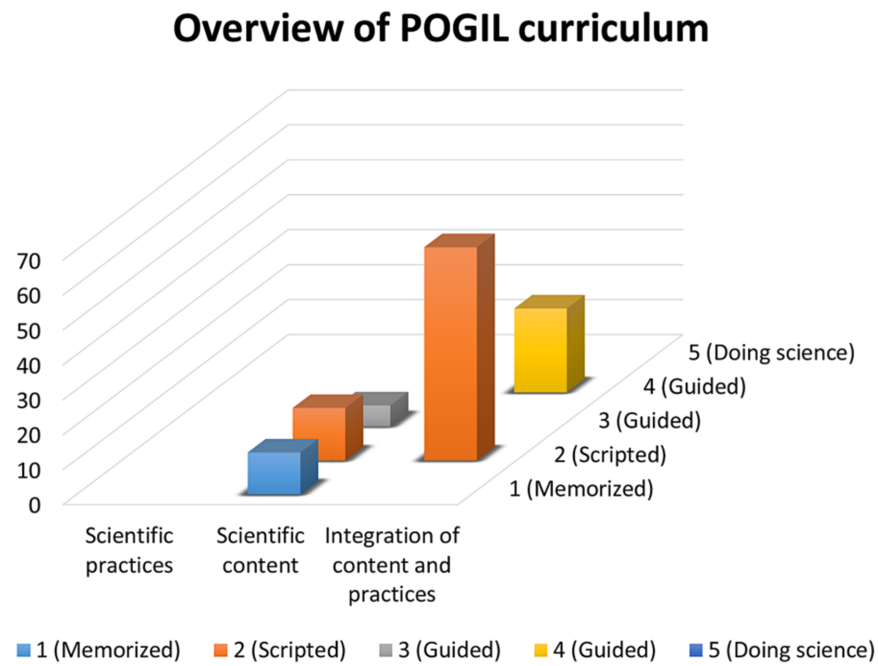


Figure 3

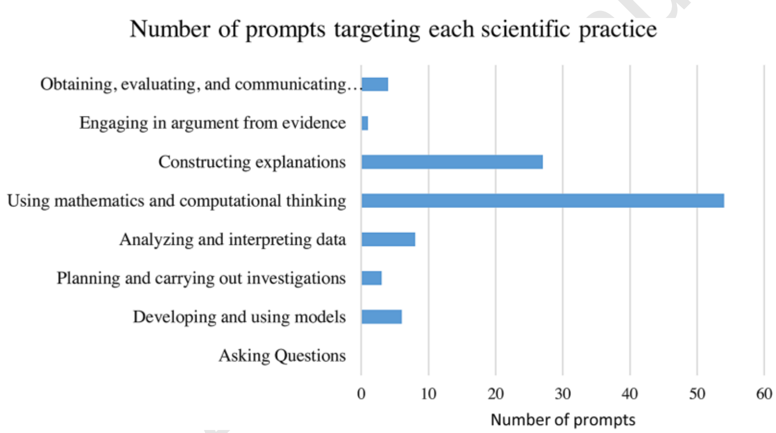


Figure 4

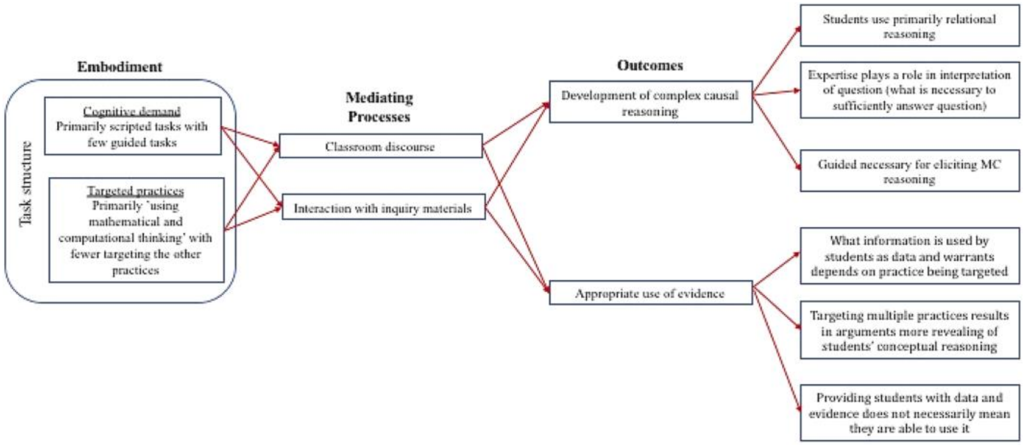

Embodiment
Task structure
Cognitive demand
Primarily scripted tasks with few guided tasks
Targeted practices
Primarily 'using mathematical and computational thinking' with fewer targeting the other practices
Mediating Processes
Classroom discourse
Interaction with inquiry materials
Outcomes
Development of complex causal reasoning
Appropriate use of evidence
Students use primarily relational reasoning
Expertise plays a role in interpretation of question (what is necessary to sufficiently answer question)
Guided necessary for eliciting MC reasoning
What information is used by students as data and warrants depends on practice being targeted
Targeting multiple practices results in arguments more revealing of students' conceptual reasoning
Providing students with data and evidence does not necessarily mean they are able to use it

Figure 5