

Exploring the impact of the framing of a laboratory experiment on the nature of student argumentation

Steven J. Petritis,* Colleen Kelley and Vicente Talanquer

Research on student argumentation in chemistry laboratories has mainly focused on evaluating the quality of students' arguments and analyzing the structure of such arguments (i.e. claims, evidence, and rationale). Despite advances in these areas, little is known about the impact of activity framing on the nature of student argumentation in laboratory settings. In this research study, we analyzed the arguments generated by college organic chemistry students working on a substitution reaction experiment that was framed in two distinct ways: predict-verify and observe-infer. The arguments constructed by students in their post-laboratory reports under each laboratory frame were characterized by paying attention to both domain-specific and domain-general features. Our analysis revealed significant differences in the chemical concepts and ideas that students under the two conditions invoked, as well as in the level of integration, specificity, alignment, and type of reasoning observed within and across different argument components. Our findings highlight the importance of paying attention to how experiments are framed in terms of the goals, procedures, information, and tools available to students as these decisions can have a major impact on the nature of the claims students make, their use of evidence, and the approach to reasoning that they follow.

Introduction

Scientists routinely engage in argument from evidence as they try to make sense of observable phenomena (National Research Council, 2012). As one of the eight foundational science practices identified in the US framework for K-12 science education, argumentation is expected to be a core component of science laboratories, presenting students with authentic opportunities to engage in this epistemic practice. Undergraduate chemistry laboratories should create spaces for students to carry out their own investigations, analyze and interpret data, and build arguments based on appropriate evidence. While engaging in these scientific practices, students are encouraged to coordinate theoretical constructs and physical observations as they search for meaning in data. This integration of the abstract and the empirical offers a ripe opportunity for students to build more meaningful understandings.

As educators are asked to engage students in scientific inquiry and reasoning, education researchers are trying to better understand how students construct arguments in different laboratory contexts and what instructional strategies support the development of this scientific practice. Although there is much agreement that engaging in productive argumentation is beneficial for student learning, our understanding of how different laboratory factors affect the nature of students' arguments is more limited. This paper seeks to provide insights in this area by characterizing how the framing of laboratory

activity impacts the types of arguments students build. We seek to characterize how changes in the way the laboratory goals, procedures, and tools are presented to students affect the types of arguments they built when communicating their results.

Argumentation in the laboratory

Argumentation is a process of "logical discourse whose goal is to tease out the relationship between ideas and evidence" (Duschl *et al.*, 2007). As a core scientific practice, this form of logical discourse has taken many forms across K-12 and undergraduate science education, including both classroom and laboratory learning (Erduran, Simon, & Osborne, 2004; Jimenez-Aleixandre, 2008; Jimenez-Aleixandre & Erduran, 2008; Osborne, 2010). In chemistry, research in this area has provided insights into how different types of learners engage in this practice in the classroom frequently using Toulmin's framework for argumentation. These investigations have focused on secondary school students (Abi-El-Mona & Fouad Abd-El-Khalick, 2006), college science students (Abi-El-Mona & Fouad Abd-El-Khalick, 2011), undergraduate organic (Cruz-Ramírez de Arellano & Towns, 2014) and physical chemistry students (Moon *et al.*, 2016), and preservice chemistry teachers (Pabuccu & Erduran, 2017).

In the chemistry laboratory, argumentation has been fostered through a variety of instructional models, such as Argument-Driven Inquiry (ADI) and the Science Writing Heuristic (SWH), and instructional styles, including inquiry and confirmatory experiments (Grooms, 2020; Moon *et al.*, 2019; Sampson *et al.*, 2010; Keys *et al.*, 1999; Katchevich *et al.*, 2013). Research on these various instructional forms has focused on two key areas: 1) evaluating the

quality of students' arguments in a variety of laboratory settings, and 2) characterizing the structure of students' argument (i.e. claims, evidence, and rationale).

Evaluating student arguments

Researchers interested in the effects of the implementation of the ADI and SWH instructional models in chemistry laboratories have developed different strategies to assess the quality of student written arguments. These two instructional models align in how they define a claim (a tentative conclusion or "an explanation that they have uncovered by their work") and evidence (articulated as a measurement, observation, difference, or relationships in the laboratory) (Sampson *et al.*, 2010; Burke *et al.*, 2006). Although their approaches to engaging students in argumentation differ, both models ultimately guide learners to develop questions or tasks that lead them to justify why their evidence supports their claims.

Burke and Greenbowe (2006) first implemented the SWH framework in chemistry laboratories based on research that showed its positive effects on student engagement and learning as compared to traditional laboratory courses. At its core, the SWH approach has students identify a researchable question, design and carry out investigations to answer it, make claims based on evidence, and reflect on how their ideas changed over the course of the experience (Hand & Choi, 2010). Choi, Hand, & Greenbowe (2013) developed analytical and holistic argument evaluation frameworks to score student arguments and observed correlation between argument scores and general chemistry classroom achievement.

Similarly, the ADI instructional model guides students to produce tentative arguments in the form of claims, evidence, and reasoning, and participate in a peer review process before revising their final arguments (Sampson *et al.*, 2010; Walker *et al.*, 2011). Using the ADI model, Walker & Sampson (2013) observed an increase in both written and oral argumentation scores over the course of a semester as students engaged with the argument-focused curriculum.

Domin (1999) has identified four different laboratory instructional styles, differentiated by their outcome (predetermined or undetermined), approach (deductive or inductive), and procedure (given or student-generated). Katchevich *et al.* (2011) argue that inquiry-based laboratory experiments provide students an effective platform to conduct argumentation contrasted with an argument-poor confirmatory experiment. Other studies have shown that guided-inquiry laboratory curricula afford students a greater opportunity to utilize an inductive approach to reasoning (Burke & Greenbowe, 2006). Consistent with similar studies in biology and physiology laboratories, Carmel *et al.* (2019) suggest that inquiry-based laboratory curricula create better opportunities for student argumentation and increased engagement with course content (Cronje *et al.*, 2013; Colthorpe *et al.*, 2010; Reiser *et al.*, 2001).

Characterizing argument components

As they construct arguments, students actively analyze data, build inferences, and justify them. These choices manifest in the claims students make, the evidence they provide, and the rationales they build. Besides assessing the quality of each of these components, researchers have investigated how students coordinate these pieces to craft written arguments in a laboratory setting.

A major goal of predominant argumentation frameworks is for students to justify why their evidence supports their claims. Analytical and holistic scoring frameworks have been used to evaluate the different components of an argument and their relationships (Choi, Hand, & Greenbowe, 2013). In these studies, overall argument quality has been shown to depend on the nature of the claims-evidence relationships that are built. Sandoval & Millwood (2005) describe how students often emphasize their claims without citing their data as evidence of their findings. Although they identify the need to reference laboratory data, students often fail to recognize patterns in the data that act as justification of their claims. Thus, students present evidence as self-evident and omit explicit reasoning to justify their argument (Brem & Rips, 2000).

McNeill & Krajcik (2007) suggested that students do have the ability to quickly link evidence to a claim but observed that students struggled to use empirical evidence as support in their arguments. A contrasting study using the ADI model observed that students relied much less heavily on theoretical knowledge when rationalizing their arguments (Walker, Van Duzor, & Lower, 2019). Choi, Hand, & Greenbowe (2013) also noted this struggle to distinguish between theoretical knowledge and empirical evidence when constructing arguments, echoing previous studies by Carey & Smith (1993) and Driver *et al.* (2000). Deciding which evidence to provide in support of a claim is a crucial step in constructing a quality argument. Yet, students often vary greatly with regards to their choice of evidence even when properly scaffolded using an argumentation model.

Existing research on argumentation in the chemistry laboratory has mainly focused on the structure of student arguments, including how various argument components (i.e. claims, evidence, and rationale) are combined to demonstrate scientific reasoning. This reasoning may be deductive, as students use a general principle to inform their hypothesis about an empirical phenomenon. Once they make their observations, students can either confirm or refute their original hypothesis. In contrast, student reasoning may be inductive when they ground their tentative hypotheses in the specific data gathered from their own empirical observations. Students can then make generalizable claims based on these empirical findings. Although most existing work on argumentation has focused on the analysis of domain-general aspects of expressed student reasoning, some researchers across various academic disciplines have also paid attention to the influence of domain-specific knowledge components on the nature of students' argumentation. These studies include work in biology (Babai & Levit-Dori, 2009), mathematics (Sommerhoff, Ufer, & Kollar, 2015), and physics (Syed, 2015), but equivalent studies in the chemistry domain are missing. The results of these different studies highlight the complex interplay of both domain-specific and domain-general components on the nature and quality of students' arguments (Engelmann *et al.*, 2018; McNeill & Krajcik, 2009).

As described in the next section, existing research also suggests that the nature of student reasoning is affected by how laboratory activity is framed and structured. In this investigation, we sought to further explore the effect of "activity framing" on student argumentation in an organic chemistry laboratory and characterize its impact on the arguments students build and the reasoning they manifest.

Activity framing

The concept of "framing" is rooted in anthropological, linguistic, and sociological research with Goffman (1974) defining a frame as an individual's experience of "what is it that's going on here?" (Bateson, 1972; MacLachlan & Reid, 1994; Tannen, 1993). Hammer, Scherr, & Reddish (2005) define a frame as a "set of expectations an individual has about the situation in which she finds herself that affect what she notices and how she thinks to act." More recent studies support this notion and demonstrate that different learning environments activate different cognitive resources as students seek to make sense of the phenomena they study (Berland & Reiser, 2010; Conlin, Gupta, & Hammer, 2010; Elby & Hammer, 2010; Scherr & Hammer, 2007).

Research across a variety of academic disciplines suggests that how students frame a classroom task strongly impacts how and why they engage in activity. For instance, Jimenez-Alexandre, Rodríguez, & Duschl (2000) contrasted examples of when students framed classroom work as "doing the lesson" versus "doing science." Students who approached their work using the former frame tended to just fulfill the task requirements, while students who used the latter frame were more likely to evaluate knowledge claims, discuss their findings, offer justifications for the different hypotheses, and support their justifications using varied approaches.

Similarly, Berland & Hammer (2012) investigated the effect of framing on when and how students participated in argumentation. The authors contend that students comply with the need to argue when the situation calls for it. However, students may also resort to "doing the lesson" just to please their teacher. Their findings exemplify how framing can foster either productive argumentation or "pseudoargumentation" depending on how students coordinate past experiences with their current environment.

Chemistry laboratories offer students the unique opportunity to experience and make sense of both chemical concepts and empirical observations. This meaning-making process often involves an intricate web of context, goals, actions, tools, and interactions that, together, establish the frame for a laboratory setting (Criswell, 2012). Results from recent studies in chemistry education have begun to highlight the major role that framing may have in shaping student argumentation. For example, Walker & Sampson (2013) suggested that the combination of materials and the framing of an investigation impact students' argumentative discourse. Thus, they proposed that experimental goals should be aligned with the tools and techniques recommended for students to use. An instructor, for instance, could modify the framing of an experiment to elicit either methods-focused or claims-focused arguments. In another instance, the goal

of student argumentative discourse was framed in two ways (dispute vs. deliberation), which led to higher argumentation scores for students deliberating to reach a consensus (Garcia-Mila *et al.*, 2013). The findings from this study specifically suggests that not all instances of argumentation promote scientific reasoning equally. Alternatively, laboratory questions could be reframed to promote student engagement with scientific practices (Rodríguez & Towns, 2018). Activity framing could also affect data interpretation. Stowe & Cooper (2019) recently suggested that not all students perceive the task of spectroscopic characterization and structural elucidation as one that requires argumentation. In alignment with Berland & Hammer (2012), they indicate that "pseudoargumentation" might occur as a result of students not perceiving the need to persuade their peers about their structural interpretations.

Guided by these ideas and results, this study was designed to explore how the framing of lab activity impacted student argumentation in organic chemistry laboratories. In particular, we characterized differences and similarities in the arguments built by students engaged in similar lab activity framed in two distinct ways. We hypothesized that the nature of student argumentation would be guided and constrained by how the goals of the activity were defined and what information and resources were available to students to successfully complete the task.

Research study

Research goals and questions

The main goal of this research study was to explore the impact of the framing of an experiment on the nature of student written argumentation in an undergraduate sophomore-level organic chemistry laboratory course. This research project followed a single experiment framed in two different ways: a predict-verify frame and an observe-infer frame. More specifically, we aimed to characterize similarities and differences in the claims students made about their laboratory findings, how they supported these claims with evidence, and the types of chemical rationales built to connect the evidence to their claims in both settings. The following research questions guided our investigation:

- (1) How did the frame of the laboratory activity impact the domain-specific chemical concepts and ideas expressed in the claims, evidence, and rationales of student post-lab arguments?
- (2) In what ways did activity framing influence the domain-general structure of arguments employed by students as they made sense of their laboratory results?

Table 1 Comparison of the characteristics of the two laboratory frames.

Predict-Verify Frame	Both Frames	Observe-Infer Frame
<ul style="list-style-type: none">Known molecular structures of eight alkyl halide starting materials.Predict reaction pathway and verify predictions.	<ul style="list-style-type: none">Same eight alkyl halide starting materials.Same two sets of reaction conditions ($S_N1 = \text{AgNO}_3$ in ethanol and $S_N2 = \text{NaI}$ in acetone).Same 16 reactions observed by students.Students provided identical table for recording qualitative observations.Provided identical pre-lab background materials in their laboratory workbook.	<ul style="list-style-type: none">Unknown molecular structures of eight alkyl halide starting materials.Observe the outcome of chemical reactions and infer reaction pathway and characteristics of reactants.

Laboratory setting

This research study was conducted in the first semester of a two-semester organic chemistry laboratory course for science and engineering majors at a public research-intensive university in the US. One experiment (Substitution Reactions) was chosen to explore the impact of the framing on the nature of student argumentation. All students enrolled in the course, regardless of laboratory frame, attended their laboratory section for approximately three hours each week. The first half hour of each laboratory session involved a pre-laboratory lecture and discussion that provided students with adequate background to relevant organic chemistry concepts, laboratory techniques, and experimental procedures. Students spent the next hour to two hours in a technical laboratory setting conducting the experiment as described in Kelley's (2019) laboratory workbook. The remaining time was used to analyze data and assemble post-lab reports following a claim-evidence-rationale framework for argumentation. Students were thus prompted to produce central claims "where you describe what happened", supporting evidence that "is your data to support your claim", and a rationale "that connect[s] your evidence to your claim" (Kelley, 2019). Fig. 1 shows an example of a post-lab report filled out using the scaffold provided to students in the course.

Substitution reaction laboratory experiment

The substitution reaction laboratory was the eighth experiment conducted during the course of interest. By this point, each of the concurrent lecture sections of the first semester organic chemistry course had already discussed nucleophilic substitution reactions. Topics included analysis of reaction mechanism and relative favorability of unimolecular (S_N1) and bimolecular (S_N2) pathways.

The laboratory course of interest reiterated these conceptual points and further emphasized the prediction of reaction pathway favorability (S_N1 and/or S_N2) based on both solvent environment (protic vs. aprotic) and molecular structure (1° , 2° , or 3° carbon compounds).

The objectives of this experiment, according to the laboratory workbook, were for students to investigate:

- 1) "mechanisms of nucleophilic substitution reactions,"
- 2) "the effect of substrate structure," and
- 3) "the rate of reaction based on the observation of precipitation of salt as a side product" (Kelley, 2019).

The substitution reaction experiment entailed students reacting each of eight alkyl halide compounds under two distinct sets of

reaction conditions. Students were told that the first set of conditions ($AgNO_3$ in ethanol – a polar, protic solvent environment) was more favorable for the S_N1 reaction mechanism whereas the second set of reaction conditions (NaI in acetone – a polar, aprotic solvent environment) was more favorable for the S_N2 reaction mechanism. If an attempted reaction was successful, students typically observed a color change from colorless to yellow and/or the presence of varying degrees of cloudiness or precipitation in their reaction vials. Students were provided a data table to use for their qualitative observations. For all 16 reactions prescribed by this laboratory experiment, students were instructed to record qualitative observations to use in constructing post-lab arguments. However, the experiment was framed in two distinctive ways (see Table 1) for different sets of students.

Predict-verify laboratory frame. Under this frame, students were provided with information about the identity of the eight alkyl halide starting materials. With this basic structural information, students were asked to *predict* the expected reactivity of each of these eight compounds based on their prior knowledge and the background information provided to them in their laboratory workbook. Students completed a total of 16 reactions (eight under S_N1 -favored conditions and eight under S_N2 -favored conditions) and collected qualitative observations in order to *verify* their predictions.

Observe-infer laboratory frame. In this case, students worked with the same eight compounds using the two sets of reaction conditions described above. However, the identities of these molecules remained concealed. Although students did know that each of these compounds were alkyl halides, they could not make specific predictions as to their expected reactivity. Students were asked to conduct the necessary chemical reactions to *observe* and explore the reactivity of the unknown substances under different conditions, record their observations, and use this qualitative data to *make inferences* about the nature of the reactants.

Data collection and participants

All data collection occurred between the Spring 2019 and Fall 2019 semesters. Most students were non-chemistry, science majors (e. g. biology, nutrition, physiology, psychology, etc.) concurrently enrolled in the first semester of the two-semester sequence of organic chemistry lecture. These students worked in laboratory sections of up to 24 students led by a graduate student instructor (GSI) with up to two undergraduate laboratory preceptors. A total of ten laboratory sections were chosen at random and consented for

Figure 1 Example of a student-constructed post-lab argument.

Major Outcomes			
Result and Discussion	Central Claims	Description of Supporting Evidence	Arguments That Connect Your Evidence to Your Claim
	unknown #2 is a tertiary alkyl halide	The reaction with ethanol was quick & immediate (S_N1) whereas with acetone it was not	In S_N1 favored reactions, tertiary alkyl halides favor S_N1 compared to S_N2 (why we saw a strong S_N1 rxn)
	unknown #3 may be a negatively charged nucleophile	The S_N2 reaction was much quicker and reactive in comparison to the S_N1 rxn.	Negatively charged nucleophiles favor S_N2 reactions as we saw here with the faster rxn in $S_N2 > S_N1$.
	unknown #1 and #4 may be secondary alkyl halides	For both #1 and #4, the reactions were equivalent to each other regardless of S_N1/S_N2 pathways; showing why these rxns	secondary (2°) alkyl halides react under S_N1 favored conditions through S_N2
	Implications and Reflections		

Handwritten notes on the right side of the table:

- and no S_N2 rxn
- would be the same regardless of S_N1 or S_N2

participation in the research study. Seven of these laboratory sections were instructed under the predict-verify frame (led by seven different GSIs) while the remaining three laboratory sections were instructed under the observe-infer frame (led by two different GSIs). Recruitment of research participants was done verbally in each randomly selected laboratory section at the beginning of the semester before any data were collected. All student participants in this study were informed of their rights as research participants and gave their written consent for participation. All data collection and analysis were conducted in accordance with Institutional Review Board (IRB) policy.

A total of 191 students participated in this research study. Students progressed through their substitution reaction laboratory experiments as instructed by their GSI. Each laboratory session ended with students putting together a post-lab report, in which they constructed arguments about their laboratory findings following the claim-evidence-rationale framework. Depending on the instructional decisions made by each GSI, students completed their post-lab reports individually, in pairs, or in teams of three to five. Reports written by more than one student were counted only once in the data collection process. Post-lab reports contained anywhere between one and eight arguments (most commonly three arguments), each comprised of a single claim, evidence, and a rationale as identified by the student(s). These arguments were constructed during the post-laboratory period and served as the primary source of data collected and analyzed for the purposes of this research project.

Table 2 provides a summary of the student participants, post-lab reports submitted, and total arguments analyzed for this study, broken down by laboratory frame of data collection. The post-lab reports belonging to consented students were collected, de-identified, scanned, and promptly returned to their respective GSI. Post-lab reports were transcribed verbatim to ensure student writing was legible and useful for coding purposes.

Table 2 Student participation and data collection by laboratory frame.

Laboratory Frame	Student N	Post-Lab Report N	Argument N
Predict-Verify	126	56	162
Observe-Infer	65	65	193
Total	191	121	355

Qualitative coding analysis

Data analysis began with the first author of this paper (a graduate student) reading transcripts of post-lab reports and compiling a tentative list of the domain-specific chemical knowledge expressed by students in their arguments. This researcher also explored the general structure of these arguments by focusing on the content of each claim, evidence, and rationale element and how these components were connected in the argument. This domain-general approach led to the development of another tentative list of argument-level codes that helped characterize how student arguments were constructed. Domain-specific and domain-general codes emerged from the data analysis and were not directly based

on categories used in prior work; they specifically characterized differences in the types of arguments generated by students working under different laboratory frames in our study. The initial lists of codes was discussed with the second author (a laboratory instructional manager) until the two reached agreement on a qualitative codebook to be used at both the domain-specific and domain-general levels of analysis.

To begin formal analysis, a randomly selected set of student reports (equally split between each laboratory frame) were selected and each argument component (i.e. claims, evidence, and rationale) was independently coded by the two researchers following the agreed upon codebook. The researchers met to discuss their coding of each argument component and reached complete agreement on their analysis before moving on to another set of reports. This iterative process was repeated until complete consensus was reached on close to 50% ($N = 62$) of the post-lab reports collected in this research study, split evenly between both laboratory frames. The codebook was refined during this process. The remaining ~50% of reports were coded by the graduate student researcher using the agreed upon qualitative codebook. Throughout this process, any additional changes to the qualitative codebook were discussed and agreed upon before being implemented in analyzing student reports. After coding the reports for their specific argument components, each report was then analyzed at the argument-level. The two researchers followed the same iterative process: analyzing the arguments individually, meeting to reach complete consensus on the assigned codes, and discussing any proposed revisions to the argument-level codes.

The codebook developed and used during the analysis included codes divided into two main categories: domain-specific and domain-general codes. Domain-specific codes helped us characterize the types of chemical concepts and ideas that students invoked in their claims, evidence, and rationales. Domain-general codes were used to characterize general characteristics of the claims, evidence, and rationales included in the reports or of the full arguments. These two sets of codes facilitated the identification of similarities and differences between arguments generated by students under different laboratory framing conditions

Domain-specific codes. The domain-specific level of analysis led to the identification of four major coding categories for the chemical concepts or ideas expressed in students' arguments: reaction pathway, molecular structure, chemical property, and reaction conditions. These codes were used to analyze each claim, evidence, and rationale element students included in their post-lab reports (see Table 3 for examples of the types of arguments built by students working under different laboratory frames).

The "reaction pathway" code identified references to the mechanistic pathway followed by the nucleophilic substitution reactions: unimolecular (S_N1), bimolecular (S_N2), both, or neither.

The "molecular structure" code highlighted references to the shape, size, or structure-related interactions (e.g. steric hindrance) of the molecules involved in the substitution reactions.

The "chemical property" code identified references to the properties of the molecules involved in their experiments, including reactivity ("lowest transition state energy"), stability ("primary unstable carbocation"), and electronegativity ("stable with negative charge").

Table 3 Examples of two student arguments, one from the predict-verify frame (report 92) and the other from the observe-infer frame (report 86).

	Claim	Evidence	Rationale
Predict-Verify Report 92	S_N1 favors a more stable carbocation.	2-chlorobutane undergoes an S_N2 reaction, quickly forming precipitate. <i>Tert</i> -butyl chloride undergoes an S_N1 reaction, quickly forming precipitate.	<i>Tert</i> -butyl chloride is a tertiary carbon, so the carbocation is very stable, so it undergoes an S_N1 reaction. 2-chlorobutane has little steric hindrance, so it undergoes an S_N2 reaction.
Observe-Infer Report 86	Unknowns 3, 4, and 7 are different from other substances by reactivity.	In S_N1 , reactions 3, 4, 7, and 8 had a form of reaction (+, ++, or +++) while in S_N2 there was no reaction for unknowns 3, 4, and 7.	The reactions that had precipitate were tertiary substrates as tertiary substrates only undergo S_N1 and not S_N2 reaction, explaining why unknowns 3, 4, 7 were not reactive in S_N2 reactions.

The "reactions conditions" code highlighted references to various laboratory procedures ("adding heat") and nature of the reactants ("bromine leaving group") or the chemical environment ("polar, protic solvent") that affected the chemical process.

Domain-general codes. The domain-general level of analysis included seven coding categories: specificity, explicitness, completeness, differentiation, integration, alignment, and approach to reasoning. These categories helped to characterize the individual components of an argument (i.e. claim, evidence, and rationale) and their connections.

The "specificity" coding category characterized the nature of students' claims and included two codes: case-specific inferences and class-level inferences. The case-specific code identified inferences referring to specific types of substances, or their molecules, used in the laboratory ("*Tert-butyl chloride will go through an S_N1 reaction but not S_N2* "). The class-level code identified inferences referring to general classes of substances or reactions, such as the claim made in report 92 of Table 3 (" S_N1 favors a more stable carbocation").

The "explicitness" coding category identified how clearly students expressed their evidence and rationale components and included two codes: explicit and implicit. With respect to evidence, the explicit code identified instances in which students clearly presented what they observed in their experiment ("*quickly forming +, ++, or +++ level of precipitate*"). The implicit code identified instances where students lacked clarity in describing their experimental observations, such as "*when comparing 2-chlorobutane and 2-bromobutane, bromine reacted much better*" which lists experimental evidence without explicitly describing what was observed during experimentation (i.e. precipitation and/or color change). With respect to rationale, the explicit code identified instances in which students clearly detailed their rationale, such as in the rationale for report 92 in Table 3. This explicit rationale referred to the stable carbocation formed from the tertiary carbon of *tert*-butyl chloride and related this stability to mechanistic favorability (S_N1). The implicit code identified instances where a rationale could not be clearly interpreted without further assumptions about student thinking. For example, in the rationale "*When placed in a water bath, a reaction took place whereas it didn't on its own,*" one has to assume that the student is referring to the effect of temperature on the chemical process in rationalizing claims about favored reaction pathway.

Codes in the "completeness" category characterized the extent to which students presented all the necessary evidence or rationale and included two codes: complete and incomplete. With respect to evidence, the complete code identified instances when students provided detailed observations ("*No reaction for unknown 6 in $AgNO_3$ in ethanol. However, heavy precipitate (opaque) occurred in NaI in acetone*"). The incomplete code identified instances where descriptions were insufficient ("*It required heat in NaI in acetone*" without reference to the observable outcome of the reaction using $AgNO_3$ in ethanol). Complete rationales included sufficient justification without missing any critical details needed to support their reasoning ("*Expected *tert*-butyl chloride to react in S_N1 , but it reacted in both. Expected to react in S_N1 because of tertiary carbon favorable conditions in S_N1* "). Incomplete rationales lacked key details in support of a claim. For example, the rationale: "*Expected to react in S_N2 because it is a primary carbon which typically observes the S_N2 reaction*" is missing the identity of the substance it refers to as well as the laboratory observations of how that alkyl halide reacted under S_N2 conditions.

The "differentiation" coding category identified arguments that explicitly compared the properties or behaviors of different chemical species or reactions. This category was applied in the analysis of claims, evidence, and rationale components and included two codes: multiple and single. The "multiple" code was applied to cases similar to the claim made in report 86 in Table 3 in which the student pointed to differences in the reactivity of several substances. The "single" code was used in cases where descriptions focused on the properties or behaviors of a single substance or reaction ("*Unknown 2 undergoes an S_N1 reaction mechanism*").

The "integration" coding category helped characterize the level of integration of the concepts and ideas invoked in student rationales and included two codes: integrated and fragmented. The integrated code was applied to rationales that demonstrated students reasonably connecting knowledge pieces and/or experimental data. For example, the rationale "*The reaction is S_N1 because it reacted in ethanol, not in acetone. Acetone favors S_N2 because it's polar, aprotic while ethanol favors S_N1 because it's polar, protic*" highlighted the link between solvent environment (polar, protic vs. polar, aprotic) and mechanistic pathway. The fragmented code identified student rationales that presented pieces of knowledge and/or observation in isolation. For example, the rationale "*The carbocation formed in S_N1 could be stabilized by resonance. It's a primary carbon which favors S_N2* " highlighted two pieces of chemical knowledge ("resonance"

stability and “primary carbon”) related to the S_N1 and S_N2 reaction pathways, respectively, without discussion of how these key concepts connect to each other or the substance being described.

The “alignment” coding category characterized arguments by the degree of coherence between the claims and rationale components and included two codes: aligned and misaligned. The aligned code highlighted consistency between the claim and rationale components as shown in report 92 of Table 3, in which both the claim and rationale emphasized carbocation stability and reaction pathway. The misaligned code highlighted inconsistency between the claims (“Cyclohexylbromide did S_N1 ”) and rationale components (“Bromine is attached to a secondary carbon which can do either S_N1 or S_N2 ”). The claim in this example emphasized the reaction pathway observed as a result of the laboratory experiment whereas the rationale discussed the structure (“secondary carbon”) of the substrate.

The “approach to reasoning” coding category characterized the type of reasoning employed by students in rationalizing their claims and included either deductive, inductive, or hybrid codes. The deductive code identified students supporting their specific claims (“2-chlorobutane reacted better by S_N2 mechanism than by S_N1 ”) with general chemical rules or principles (“chlorine is a relatively ‘bad’ leaving group, it needs more activation energy to undergo S_N1 ”). The inductive code identified students justifying general claims about reaction pathway or molecules (“ S_N1 reacts better with better leaving group”) with their specific observations and findings (“From observation, 2-bromobutane had ++ [precipitation] where 2-chlorobutane needed heat for reaction”). The hybrid code highlighted instances when student claims (“We expected no nucleophilic reactions would occur [for bromobenzene]”) were supported by both general chemical concepts and ideas (“an sp^2 hybridized carbon won’t undergo a nucleophilic substitution reaction”) and specific components of their experimental observations (“no reactions were observed for bromobenzene just as predicted”).

Statistical analysis of qualitative coding

The results from the qualitative coding analysis described above were exported to Microsoft Excel (2016). All qualitative codes were converted into numbers, which were arbitrarily assigned for each categorical variable described above. For example, the coding category of specificity has two codes, case-specific and class-level, which were assigned a 1 and 2, respectively. Each domain-specific and domain-general code were converted in a similar manner. A total

of 355 individual arguments, each consisting of one claim, one evidence, and one rationale component, were coded following this method.

Statistical data analysis was conducted using the R statistical software for Windows, version 3.6.1. The occurrence and co-occurrence of each qualitative code was tallied and broken down by laboratory frame (predict-verify vs. observe-infer). A series of χ^2 tests were run to explore the association between the framing of a laboratory experiment and the occurrence of each categorical code ($N = 355$ arguments). For χ^2 tests that demonstrated a statistically significant association between categorical variables (at the $\alpha = .05$ level), Cramer’s V was calculated to highlight the strength of the association between variables and interpreted as outlined by Cohen (1988): a small effect size (weak association) is between 0.1 and 0.3, a medium effect size (moderate association) is between 0.3 and 0.5, and a large effect size (strong association) is great than 0.5.

Results

In the subsections that follow, we present the results of our analysis of the impact of activity framing on: 1) the domain-specific chemical concepts and ideas invoked in the different components of students’ arguments, and 2) the domain-general structure of the arguments included in their post-laboratory reports.

How did the frame of the laboratory activity impact the domain-specific chemical concepts and ideas expressed in the claims, evidence, and rationales of student post-lab arguments?

Our analysis of the domain-specific chemical knowledge expressed in student arguments focused on four key categories: reaction pathway, molecular structure, chemical property, and reaction conditions. Table 4 summarizes the findings from this domain-specific qualitative coding analysis. The absolute and relative frequencies of occurrence are shown for each of the domain-specific codes across each argument component for both laboratory frames.

The results summarized in Table 4 indicate that a change in how the laboratory experiment was framed did significantly affect the types of chemical concepts and ideas that students invoked in their arguments, but the effect size was different across the various argument components. Claims made by students working in the predict-verify frame were more likely to relate to the molecular structure of the reactants than the claims made by students under the observe-infer experimental frame who, in contrast, more

Table 4 Domain-specific qualitative coding results conducted for the claims, evidence, and rationale components of student post-lab arguments.

	Claims		Evidence		Rationale	
	Predict-Verify	Observe-Infer	Predict-Verify	Observe-Infer	Predict-Verify	Observe-Infer
Domain-Specific Codes	Absolute Freq. (Relative Freq.)	Absolute Freq. (Relative Freq.)	Absolute Freq. (Relative Freq.)	Absolute Freq. (Relative Freq.)	Absolute Freq. (Relative Freq.)	Absolute Freq. (Relative Freq.)
Reaction Pathway	146 (90.1%)* ^S	155 (80.3%)* ^S	53 (32.7%)* ^M	113 (58.5%)* ^M	125 (77.2%)	163 (84.5%)
Molecular Structure	53 (32.7%)* ^M	17 (8.8%)* ^M	28 (17.3%)* ^S	5 (2.6%)* ^S	124 (76.5%)* ^L	43 (22.3%)* ^L
Chemical Property	2 (1.2%)	9 (4.7%)	3 (1.9%)	2 (1.0%)	33 (20.4%)	31 (16.1%)
Reaction Conditions	20 (12.3%)* ^S	46 (23.8%)* ^S	56 (34.6%)* ^M	136 (70.5%)* ^M	75 (46.3%)* ^M	151 (78.2%)* ^M

* Indicates $p < .05$ for the χ^2 test of the association between laboratory frame and domain-specific codes. ^S Indicates a Cramer’s V with a small effect size. ^M Indicates a Cramer’s V with a medium effect size. ^L Indicates a Cramer’s V with a large effect size.

Table 5 Examples of prototypical student arguments, one from the predict-verify frame (report 11) and the other from the observe-infer frame (report 37).

	Claim	Evidence	Rationale
Predict-Verify Report 11	All primary carbons reacted with S _N 2 and all tertiary carbons reacted with S _N 1.	Both primary carbons and our one tertiary carbon reacted with appropriate conditions.	Primary halides react only with S _N 2 mechanisms and tertiary halides react only with S _N 1 mechanisms.
Observe-Infer Report 37	Unknown 1 reacted best under S _N 2 conditions.	Under S _N 2, the reaction occurred immediately and became very cloudy; however, it took a long time in S _N 1 conditions.	Since it occurred in both conditions, we know it 'can't' work in both; however, it favors S _N 2 because it occurred much quicker in acetone.

frequently made claims referring to reaction conditions. This difference is illustrated by the representative examples included in Table 5. In these arguments, the claim made in report 11 (predict-verify) connects the degree of substitution of carbons to the pathway for nucleophilic substitution, while the claim in report 37 (observe-infer) connects reaction conditions (e.g. faster reaction in acetone) to reaction pathway. References to reaction pathway were the most common in the claims made in both types of reports, although with a slightly higher frequency in arguments built by students working under the predict-verify frame.

As can be inferred from the frequency data in Table 4, students in the observe-infer frame provided more pieces of evidence in their arguments than students working in the predict-verify frame, and their evidence was more likely related to reaction conditions and pathway than to molecular structure (which was more predominant in reports of students in the predict-verify condition). This difference is also illustrated by the examples presented in Table 5. The contrasting focus on reaction conditions in the observe-infer frame versus molecular structure in the predict-verify frame was also prevalent in the rationales built by students to justify their claims based on the evidence that they provided (as also exemplified by the examples shown in Table 5).

Laboratory framing did not seem to have an impact on the extent to which students referred to chemical properties across different components in their arguments. Chemical properties of the reactants were the least frequently cited compared to other major categories of concepts and ideas identified in the analyzed arguments. References to different types of mechanistic pathways were typically the most common across all argument components in both types of laboratory framing conditions. The most significant difference in this area was observed in the nature of the evidence presented by students working under different lab frames. References to reaction

pathway and reaction conditions were often entangled in the arguments built by students working under the observe-infer frame and that led to a higher frequency of codes assigned in both areas.

In addition to analyzing the frequency with which domain-specific qualitative codes occurred in each argument component, we tabulated the co-occurrence of these code across the entirety of student arguments. The absolute frequencies of co-occurrence of the domain-specific codes for the predict-verify and observe-infer frames are summarized in the different columns in Table 6. The absolute frequency of co-occurrence represents the number of instances that one code co-occurred with another code. The analysis summarized in this table highlights the higher relative frequency of co-occurrences between molecular structure and reaction pathway codes in arguments generated under a predict-verify experiment frame, whereas arguments built under the observe-infer frame exhibited a higher frequency of co-occurrence between the reaction pathway and reaction conditions codes.

The major types of co-occurrences identified in our analysis can be detected in the examples of arguments included in Table 5. In report 11, the students connected the reaction pathway to a structural feature of the reactants (degree of substitution of carbon atom) and the connection is present in both the claim and the rationale. On the other hand, in report 37 a connection was made between reaction pathway and reactions conditions ("occurred immediately", "very cloudy", and "quicker in acetone") in both the evidence and rationale elements. In report 11, the students referred to the molecular structure of the reactants to justify their claim about reactivity without mentioning empirical observations or specific reaction conditions. In contrast, the student in report 37 referred to the reactivity observed in "both conditions" and discussed reaction favorability based on reaction speed ("quicker") and solvent environment ("acetone") without reference to structural features.

Table 6 Co-occurrence of domain-specific qualitative codes for each laboratory frame in student post-lab arguments.

Domain-Specific Codes	Predict-Verify (N = 162)				Observe-Infer (N = 193)			
	Reaction Pathway	Molecular Structure	Chemical Property	Reaction Conditions	Reaction Pathway	Molecular Structure	Chemical Property	Reaction Conditions
Reaction Pathway	X	116 ^{*M}	25	86 ^{*M}	X	46 ^{*M}	31	233 ^{*M}
Molecular Structure		X	32	59 ^{*S}		X	15	37 ^{*S}
Chemical Property			X	24			X	34
Reaction Conditions				X				X

* Indicates $p < .05$ for the χ^2 test of the association between laboratory frame and domain-general codes. ^S Indicates a Cramer's V with a small effect size.

^M Indicates a Cramer's V with a medium effect size. ^L Indicates a Cramer's V with a large effect size.

Table 7 Domain-general qualitative coding results for the claims, evidence, and rationale components of student post-lab arguments.

Coding Category	Codes	Predict-Verify (N = 162)		Observe-Infer (N = 193)	
		Absolute Frequency	Relative Frequency	Absolute Frequency	Relative Frequency
Specificity	Case-Specific	102* ^L	63.0%* ^L	192* ^L	99.5%* ^L
	Class-Level	60* ^L	37.0%* ^L	1* ^L	0.5%* ^L
Explicitness	Explicit (Evidence)	69	42.6%	94	48.7%
	Implicit (Evidence)	93	57.4%	99	51.3%
	Explicit (Rationale)	64* ^L	39.5%* ^L	178* ^L	92.2%* ^L
	Implicit (Rationale)	98* ^L	60.5%* ^L	15* ^L	7.8%* ^L
Completeness	Complete (Evidence)	79* ^M	48.8%* ^M	145* ^M	75.1%* ^M
	Incomplete (Evidence)	83* ^M	51.2%* ^M	48* ^M	24.9%* ^M
	Complete (Rationale)	50* ^L	30.9%* ^L	141* ^L	73.1%* ^L
	Incomplete (Rationale)	112* ^L	69.1%* ^L	52* ^L	26.9%* ^L
Differentiation	Multiple (Claims)	45* ^M	27.8%* ^M	144* ^M	74.6%* ^M
	Single (Claims)	117* ^M	72.2%* ^M	49* ^M	25.4%* ^M
	Multiple (Evidence)	42* ^S	25.9%* ^S	99* ^S	51.3%* ^S
	Single (Evidence)	120* ^S	74.1%* ^S	94* ^S	48.7%* ^S
	Multiple (Rationale)	52* ^L	32.1%* ^L	154* ^L	79.8%* ^L
	Single (Rationale)	110* ^L	67.9%* ^L	39* ^L	20.2%* ^L
Integration	Integrated	111* ^M	68.5%* ^M	188* ^M	97.4%* ^M
	Fragmented	51* ^M	31.5%* ^M	5* ^M	2.6%* ^M
Alignment	Aligned	129* ^S	79.6%* ^S	179* ^S	92.7%* ^S
	Misaligned	33* ^S	20.4%* ^S	14* ^S	7.3%* ^S
Approach to Reasoning	Deductive	117* ^L	72.2%* ^L	16* ^L	8.3%* ^L
	Inductive	21* ^M	13.0%* ^M	101* ^M	52.3%* ^M
	Hybrid	24* ^M	14.8%* ^M	76* ^M	39.4%* ^M

*Indicates $p < .05$ for the χ^2 test of the association between laboratory frame and domain-general codes. ^S Indicates a Cramer's V with a small effect size.

^M Indicates a Cramer's V with a medium effect size. ^L Indicates a Cramer's V with a large effect size.

Although both rationales attempted to justify a claim related to reaction pathway (S_N1 vs. S_N2), these students invoked different chemical concepts, ideas, and data/observation in support of their argument. These differences were shown to be statistically significant for both the associations reaction pathway-molecular structure (predict-verify) and reaction pathway-reaction conditions (observe-infer) with each showing a moderate association with laboratory frame.

In general, our analysis elicited a common focus on reaction pathway for students in both the predict-verify and observe-infer frames. However, students under the predict-verify frame were more likely to connect the molecular structure of their reactants to the mechanistic pathway, while students working under the observe-infer frame were more likely to build connections between reaction conditions and pathway. These differences illustrate the major impact that laboratory framing can have on the nature of the connections that students built and the concepts they integrate.

In what ways did activity framing influence the domain-general structure of arguments employed by students as they made sense of their laboratory results?

Our analysis of the domain-general structure of arguments focused on seven key categories: specificity, explicitness, completeness, differentiation, integration, alignment, and approach to reasoning.

Table 7 summarizes the findings from this domain-general qualitative coding analysis. The absolute and relative frequencies of occurrence are shown for each of the domain-general codes (in their respective argument components) for both laboratory frames.

The results summarized in Table 7 indicate that the change in how the laboratory experiment was framed affected the structure of the arguments students constructed, but the effect varied for different categories of analysis. In terms of specificity of claims, students working in the predict-verify frame were more likely to make class-level claims in which they referred to general classes of substances or reactions, while students working under the observe-infer frame almost entirely constructed case-specific claims related to the specific substances used in their experiment. This difference is illustrated by the examples in Table 8. The claim made in report 102 (predict-verify) suggests that molecules containing tertiary carbons only react following the S_N1 mechanistic pathway, while report 22 (observe-infer) refers to a specific substance (unknown 2) used by the student in their laboratory experiment.

The explicitness of the evidence identified by students in their arguments was fairly comparable between the predict-verify (42.6% explicit) and observe-infer (48.7% explicit) laboratory frames. Although there was not a statistically significant difference in how explicitly evidence was presented by students in either frame, students working under the observe-infer frame were more likely to

Table 8 Examples of prototypical student arguments, one from the observe-infer frame (report 22) and the other from the predict-verify frame (report 102).

	Claim	Evidence	Rationale
Predict-Verify Report 102	Tertiary carbons only react as S_N1 .	<i>Tert</i> -butyl chloride reacted only as S_N1 in both solvents because it is a tertiary carbon.	This is a tertiary carbon and can only react as an S_N1 since it reacted in $AgNO_3$ but not in Nal , proving it is only an S_N1 .
Observe-Infer Report 22	Unknown 2 undergoes an S_N1 reaction mechanism.	No reaction occurred for unknown 2 in Nal in acetone. However, in $AgNO_3$ in ethanol, a heavy amount of opaque white precipitate formed.	The reaction is S_N1 because it reacted in ethanol, not in acetone. Acetone favors S_N2 because it's polar, aprotic while ethanol favors S_N1 because it's polar, protic. The S_N2 reaction failed to occur even after heating. The S_N1 reaction precipitate indicates a reaction occurred.

provide a complete description of supporting evidence as compared to the predict-verify frame. The evidence components shown in Table 8 help to illustrate these differences. The evidence for report 102 (predict-verify) mentions that the reaction proceeded only by the S_N1 mechanism without reference to what was physically observed to determine that *tert*-butyl reacted via an S_N1 mechanism. In contrast, report 22 (observe-infer) includes an explicit account of what was observed under S_N1 conditions ("a heavy amount of opaque white precipitate") as well as under S_N2 conditions ("no reaction occurred for unknown 2 in Nal in acetone").

Students working under the observe-infer laboratory frame were more likely to build chemical rationales in which supporting ideas were presented more completely and reasoning was more explicit in justifying their claims. On the other hand, most rationales built under the predict-verify frame were characterized as both implicit and incomplete. In the examples shown in Table 8, the student who built report 102 provided an incomplete rationale, stating that the "tertiary" nature of the *tert*-butyl chloride led the substrate to "only react as an S_N1 " without referring to the chemical knowledge that supported that justification (i.e. stability of the tertiary carbocation produced once the chloride leaving group leaves and subsequent stabilization by the polar, protic solvent). In contrast, the student in report 22 more completely justified the claim by discussing the reaction conditions that favored each reaction pathway and connecting them to the experimental results for the reactant under analysis.

Although there was no significant difference between laboratory frames on the "explicitness" of the evidence included in the reports, students working under the observe-infer laboratory frame were more likely to provide complete evidence and generate explicit and complete rationales than students who conducted the experiment under the predict-verify frame. When providing explicit evidence, students under the observe-infer frame included complete evidence 79.8% of the time, explicit rationale 96.8% of the time, and complete rationale 95.7% of the time compared to 50.7% (complete evidence), 58.0% (explicit rationale), and 49.3% (complete rationale) for students working under the predict-verify laboratory frame.

In the category of "differentiation", students under the observe-infer laboratory frame were more likely to craft claims in which they compared and contrasted the properties or behaviors of different chemical species or reactions in summarizing their laboratory findings. These students also more frequently differentiated substances or reactions in their evidence, although the effect size for this difference was small. In crafting their rationales, students working under the observe-infer frame were, again, more likely to

differentiate behaviors or properties of substances or reactions observed in their laboratory experience. Consider this example from report 29 (observe-infer).

Claim: "For both ethanol and acetone, unknowns 1 and 2 are similar."

Evidence: "For ethanol, unknowns 1 and 2 both turned pearly white quickly upon being shook and unknown 2 had a slight precipitate. For acetone, unknowns 1 and 2 both had no initial reactions, but once heat was applied both turned yellow and unknown 1 had a slight precipitate."

Rationale: "Since both unknowns 1 and 2 reacted similarly even with two different reagents, this means both have similar functional groups and structure. For the unknowns in ethanol, they were both S_N1 reactions and for the unknowns in acetone, they were both S_N2 reactions."

In this argument, the student made a claim about the similar reactivity of two of their unknowns under both sets of solvent conditions. In their evidence, the student described the similar empirical behaviors observed for both unknowns, including the color change observed in each set of conditions. This student concluded their argument by inferring the presence of similar functional groups for unknowns 1 and 2 from the observed similar reactivities in both experimental conditions. Overall, students working under the observe-infer frame crafted a total of 65 arguments (33.7%) in which a comparison or contrast was made between the chemical behavior of reactants whereas students in the predict-verify frame only made a total of 14 arguments (8.6%) in which a comparison or contrast in chemical reactivities was included.

Laboratory framing also impacted the level of integration of chemical concepts and ideas and data/observations included in the rationales built by students. Students working under the observe-infer frame were more likely to integrate key ideas and data/observation in making sense of their laboratory experience in contrast with students in the predict-verify frame, whose rationales were more often fragmented. Consider this example from report 39 (observe-infer).

Claim: "The activation energy of S_N1 mechanism for unknown 4 was lower than the activation energy of the S_N2 mechanism of unknown 4."

Evidence: "Heavy precipitation/cloudiness formed under both conditions, but S_N2 mechanism required heat input from water bath."

Rationale: "Since precipitation formed in both S_N1 and S_N2 , the reaction is product favored in both mechanisms. However, because S_N2 requires an input of energy (heat), it must have a

higher activation energy transition state in its rate-determining step and a greater activation energy than the S_N1 rate-determining step."

In this argument, the student made a claim about the activation energy of the S_N1 and S_N2 mechanistic pathways for unknown 4. The associated chemical rationale used empirical observations ("precipitation") as evidence for preferred reaction pathways. This student integrated these observations with theoretical ideas about relative activation energies for each reaction pathway when stating that the required "input of energy (heat)" implied that the S_N2 pathway for unknown 4 had a higher activation energy than the S_N1 pathway. This example illustrates how students integrated chemical concepts, ideas, and data/observation by linking their descriptions in support of the claim. In contrast, consider the following fragmented rationale included in report 5 (predict-verify).

Claim: "Tert-butyl chloride undergoes S_N1 type reaction but not S_N2."

Evidence: "Reaction observed (heavy precipitate) with AgNO₃ in EtOH [ethanol]. No reaction observed with NaI in acetone. Tertiary carbon."

Rationale: "EtOH is a solvent that favors S_N1 reactions. The molecule is bulkier, so harder for nucleophile to attack and forms intermediate."

In this argument, these students made a claim about the reaction pathway followed when using tert-butyl chloride as a reactant ("undergoes S_N1 type reaction"). In justifying their claim, the students referred to two separate chemical concepts, solvent environment and molecular structure (i.e. size). Although both chemical ideas (either explicitly or implicitly) relate to the S_N1 pathway, they were presented as independent statements, without discussion of how the nature of the solvent affected the likelihood of formation of the tertiary carbocation intermediate.

The analysis of the "alignment" of different components of the arguments included in the collected reports indicated that students working under the observe-infer frame were more likely (92.7%) to provide chemical rationales that were aligned with other components of their arguments than students working under the predict-verify frame (79.6%), although this difference had a small effect size. Both examples included in Table 8 represent aligned arguments in which the evidence and rationale were built using data and ideas that support the stated claims. Students working under the predict-verify frame more often generated misaligned arguments where their claims were disconnected from other argument components. Consider, for example, the following argument from report 4 (predict-verify).

Claim: "Tert-butyl chloride did not react as expected due to a possible contamination or other source of error."

Evidence: "Tert-butyl chloride should only exclusively favor an S_N1 mechanism, but ours reacted in both S_N1 and S_N2 mechanisms."

Rationale: "The tertiary carbon favors an S_N1 reaction because it is the most stable which means it has the lowest transition state energy."

In this example, students claimed that their reaction for tert-butyl chloride did not go as predicted due to a possible contamination. The evidence they identified highlights their prediction as to the expected reaction pathway and describes which mechanisms they observed ("both S_N1 and S_N2 mechanisms"). However, in their

rationale these students only discussed the expected reaction pathway for tert-butyl chloride without any reference to the effects of any potential contamination. Thus, this argument was misaligned as the claim referred to an empirical observation ("contamination") while the rationale only focused on the theoretical basis for their unverified prediction.

In addition to analyzing the alignment across different argument components, we also paid attention to the presence of misplacement of argument components (e.g. presenting a rationale in the evidence box). Consider the following example of the "evidence" from report 16 under the predict-verify frame.

Claim: "Benzyl chloride and 1-chlorobutane are both primary carbons, but 1-chlorobutane reacts faster."

Evidence: "Based on our results, for an S_N2 1-chlorobutane reacted faster than benzyl chloride because of less steric hindrance."

Rationale: "Even though both compounds are primary carbons, 1-chlorobutane reacted faster because it is less sterically hindered (no big neighbor molecules)."

In this example, these students included part of their rationale ("because of less steric hindrance") in their evidence component. Reports from students in the predict-verify frame more frequently included misplaced claims (3.7%) and rationale (6.8%) components into the evidence box compared to reports from students under the observe-infer frame that did not show this type of misplacement.

Analyzed reports also included cases in which evidence was presented in the rationale box, as exemplified in this example from report 96 from the predict-verify frame.

Claim: "If carbons are secondary, solvent and nucleophile determine pathway."

Evidence: "Protic = S_N1 favored. Aprotic = S_N2 favored."

Rationale: "Protic = 2-chlorobutane, secondary carbon, large precipitate in ethanol, no reaction in acetone."

In this case, these students made reference to specific empirical data ("precipitate in ethanol, no reaction in acetone") that was not previously included in the evidence component. Rationales built by students in the predict-verify frame also more frequently included misplaced claims (3.1%) and evidence (12.3%) component compared to rationales generated under the observe-infer frame (0% for claims and 5.7% for rationale).

In general, when students working under the observe-infer frame constructed an aligned argument their rationales were more likely to be integrated (98.9%) than the rationales included in aligned arguments from students working under the predict-verify frame (78.3%). Further, when students working under the predict-verify frame constructed misaligned arguments, they were more likely to also built fragmented rationales (69.7%) compared to students working under the observe-infer frame (21.4%).

Laboratory framing also had a significant impact on the approach to reasoning employed by students when building their arguments. The deductive approach to reasoning was invoked much more frequently by students working under the predict-verify frame compared to the observe-infer frame in which fewer than 10% of the students invoked general theoretical or empirical chemical principles to support their arguments. However, both inductive (which relied on specific laboratory observations) and hybrid approaches (combination of both deductive and inductive) were more frequent

under the observe-infer frame than the predict-verify frame. Differences in the manifestation of each of the three approaches to reasoning were significant between laboratory frames with either a large (deductive approach) or medium effect size (inductive and hybrid approaches). To better illustrate the differences in reasoning approach, consider the following example from report 112 of the predict-verify frame.

Claim: *"S_N1 reactions occurred in benzyl chloride and 2-chlorobutane with the ethanol solvent."*

Evidence: *"When the solvent and substrate mixed, the reaction occurred because the solution became cloudy."*

Rationale: *"Ethanol substitutes the various alkyl halides. The two-step process indicated the halogen bond being broken and the solvent binding to the substrate to form the carbocation. To negate the balance charge, more solvent bonded and the charge is 0."*

In this argument, the student made a specific claim that the benzyl chloride and 2-chlorobutane substrates proceeded via the S_N1 reaction pathway in the presence of an ethanol solvent. The student references evidence specific to what they observed upon reacting "the solvent and substrate". To rationalize their claim, the student discusses the mechanism of the reaction in which the "ethanol [solvent] substitutes" onto the carbocation intermediate formed through the two-step S_N1 reaction pathway. This "solvent binding" helps to stabilize the carbocation by "negat[ing] the balance charge" and allowing for a nucleophile to attack the carbocation and form the final product. In this rationale, the student invoked general principles about the nature of the S_N1 reaction mechanism, including concepts related to stabilization of the carbocation intermediate by interactions with the solvent, and did not elaborate on their empirical observations made in the laboratory. Students working under the predict-verify frame commonly invoked general structure-property relationships (e.g. primary substrates favoring S_N2 pathways, tertiary carbocations being most stable) in their rationales. Most students under the predict-verify frame (72.2%) took this deductive approach to reasoning, while only 8.3% of arguments in reports from the observe-infer laboratory frame followed this approach (large effect size).

Students working under the observe-infer frame were more likely to follow an inductive approach to reasoning in the construction of their rationales. Consider the following example from report 49 of the observe-infer frame.

Claim: *"Unknowns 2, 3, and 4 are more S_N1 favorable."*

Evidence: *"Unknown 2: S_N1 (cloudy white with big chunks of solid precipitate, no heating required) and S_N2 (no reaction happened even after heating). Unknown 3: S_N1 (greenish and cloudy solid with solid precipitate, no heating) and S_N2 (no reaction even after heating). Unknown 4: S_N1 (greenish cloudy color with big chunks of solid precipitate) and S_N2 (no immediate reaction, but when heated white precipitate is present)."*

Rationale: *"Unknowns 2, 3, and 4 did not require any energy source (heat) during the S_N1 reactions. All these unknowns had solid precipitate. On the other hand, during S_N2 reactions, all three unknowns required to be heated. Unknowns 2 and 3 did not react at all while unknown 4 had a small amount of solid precipitate."*

In this case, the student made a claim about the mechanistic favorability of unknowns 2, 3, and 4 while the student's evidence details their observations made for each unknown under both S_N1 and S_N2 reaction conditions. The rationale was built solely using empirical observations about the specific reaction outcomes ("solid precipitate") and conditions ("heat") that seemed to justify the claim made. Students under the observe-infer laboratory frame much more frequently manifested an inductive approach to reasoning (52.3%) as compared to students under the predict-verify frame (13.0%) (medium effect size). Students in the observe-infer laboratory frame were also more likely to express a hybrid approach to reasoning in which they incorporated elements of both deductive and inductive reasoning. Consider the following example from report 78 of the observe-infer frame.

Claim: *"Unknowns 2 and 3 are S_N1 reactions with AgNO₃ in ethanol."*

Evidence: *"When unknown 2 was added to solvent, a white cloudy precipitate formed immediately. Unknown 3 formed a yellow-white precipitate immediately."*

Rationale: *"Ethanol is a polar, protic solvent. It has an O-H bond, making it able to hydrogen bond. Polar, protic solvents favor S_N1 reactions. We made unknowns 2 and 3 favor S_N1 reactions by using a polar, protic solvent. The two did not react well in acetone (polar, aprotic), meaning they didn't undergo S_N2 reactions. Since they reacted immediately in polar, protic solutions, we can say that the alkyl halides have more steric hindrance and a stable leaving group."*

In this argument, the student made a claim about the specific reaction pathway of unknowns 2 and 3 (S_N1) when placed in "AgNO₃ in ethanol" conditions. Their evidence details the specific observations made for both unknown 2 and 3. In rationalizing this claim, the student first discussed the nature of the ethanol solvent ("polar, protic" and "able to hydrogen bond"), saying that this type of solvent favors the S_N1 reaction pathway. The first three sentences of this rationale highlight the student's deductive approach as they invoke general principles about solvent effects to rationalize why their reactions proceeded via this S_N1 pathway. Next, the student described their specific empirical observations ("did not react well in acetone (polar, aprotic)" and "reacted immediately in polar, protic") to justify why "unknowns 2 and 3 favor S_N1 reactions." This hybrid approach to reasoning was more commonly observed in reports from students working under the observe-infer frame (39.4%) than in those from the predict-verify frame (14.8%) (medium effect size).

Overall, students working under the observe-infer frame were more likely to construct an integrated rationale regardless of which approach to reasoning they invoked in their argument: deductive (100.0%), inductive (99.0%), or hybrid (94.7%). Comparatively, students under the predict-verify frame were less likely to construct an integrated rationale when their approach to reasoning was deductive (68.4%), inductive (71.4%), or hybrid (66.7%).

Discussion

As they participate in a laboratory experience, students are expected to engage in argument from evidence to make sense of their laboratory findings. As they seek to understand "what is it that's going on" (Goffman, 1974) in the laboratory, students must grapple

Table 9 General characterization of laboratory frames by domain-specific and domain-general codes.

Argument Code	Predict-Verify	Observe-Infer
<i>Domain-Specific</i>	Molecular Structure	Reaction Conditions
<i>Specificity</i>	Class-Level	Case-Specific
<i>Explicitness</i>	Implicit	Explicit
<i>Completeness</i>	Incomplete	Complete
<i>Differentiation</i>	Single	Multiple
<i>Integration</i>	Fragmented	Integrated
<i>Alignment</i>	Misaligned	Aligned
<i>Approach to Reasoning</i>	Deductive	Inductive and Hybrid

with data and theoretical concepts and ideas. The main findings of our study indicate that the way in which an experiment is framed significantly influences the chemical concepts and ideas that students invoke, as well as the general nature of the arguments that they built to communicate their results. Table 9 summarizes contrasting characteristics observed in the arguments constructed by study participants working under a predict-verify and an observe-infer frame in a substitution reactions lab in sophomore organic chemistry at our institution.

Students working under the predict-verify frame had access to information about the molecular structures of their starting materials and this knowledge affected the claims they made and the rationales they built. References to molecular structure were more common in their reports than in those of students working under the observe-infer frame who, lacking this information, were more prone to focus on the analysis of reactions conditions to justify their claims. Access to information about the molecular structure of reactants may have also been responsible for the presence of a greater number of class-level claims in which students working under the predict-verify frame referred to the behavior of general classes of substances. In contrast, students in the observe-infer frame more frequently made claims focused on the behaviors (reaction pathway) of the specific substances with which they worked.

Our analysis revealed no significant differences in the extent to which participants working under different framing conditions made their data explicit when presenting evidence in support of their claims. Nevertheless, written reports from students in the predict-verify frame more often lacked a complete description of supporting evidence compared to reports from students in the observe-infer frame. Existing research indicates that students often produce unsubstantiated claims in their arguments (Kuhn, 1991), missing key details needed to support their inferences (Sandoval, 2003; Sandoval & Millwood, 2005; Brem & Rips, 2000). We speculate that having information about the molecular structure of the reactants combined with theoretical knowledge about the reactivity of classes of electrophiles may have been responsible for observed differences in this area. Students in the predict-verify frame had theoretical expectations about the chemical behavior of their reactants which they may have thought needed no justification using experimental data. In the absence of this structural information, students working

under the observe-infer frame were implicitly forced to rely on experimental data to justify their claims.

Our results suggest that the predict-verify frame led students to conflate empirical and theoretical "evidence," and, thus, their arguments were more frequently misaligned and fragmented, and their rationales were more characteristically implicit and incomplete. These students tended to emphasize theoretical constructs related to molecular structure in their rationales instead of discussing the empirical data they collected. They seemed to consider their empirical data as self-evident and, thus, more frequently failed to make it explicit, analyze it thoroughly, and integrate it with theoretical concepts and ideas. Their focus on theoretical constructs to rationalize their claims is further supported by the overwhelming deductive approach to reasoning that characterized their arguments. Previous research has shown that students struggle to distinguish between theoretical knowledge and empirical evidence (Carey & Smith, 1993; Driver *et al.*, 2000). Our findings indicate that a predict-verify frame may pose additional challenges for students to coordinate theory and empirical observations (Bell & Linn, 2000; Havdala & Ashkenazi, 2007), and, thus, they may require additional scaffolding in building their arguments.

The rationales built by students working under the observe-infer laboratory frame were more frequently explicit, complete, and focused on the analysis and discussion of empirical observations about how reaction conditions affected the behavior of substances. These students more explicitly connected their evidence to their claim and built more aligned arguments. The greater focus on empirical data seemed to have led these students to more frequently draw comparisons and contrast the behaviors of unknown reactants across all argument components. Consequently, students in the observe-infer frame built arguments that more frequently followed an inductive approach to reasoning in which data served as the basis for discussions about chemical reactivity. Although these students did not have information about the molecular structure of their reactants, their rationales more often included references to both empirical observations and theoretical constructs. In general, they manifested a more sophisticated approach to reasoning in which different pieces of knowledge were presented and integrated (Sampson & Clark, 2006).

Implications

The hallmark characteristic of the predict-verify laboratory frame was that students were given the identity of the eight alkyl halide reactants used in their experiment. Structural information about whether a reactant was a primary, secondary, or tertiary alkyl halide (in addition to other structural features) afforded students the opportunity to *predict* expected reactivity on the basis of background theoretical knowledge and, ultimately, *verify* these predictions based on their own qualitative observations made in the laboratory. With access to both empirical data and theoretical knowledge, students in this laboratory frame more frequently built deductive rationales in which they invoked general chemical principles to support their specific laboratory findings (i.e. a tertiary alkyl halide having a strong preference for the S_N1 reaction mechanism). Our results suggest that when using this type of laboratory frame, student reasoning and argumentative skills may benefit from the inclusion of substrates

that exhibit unexpected reactivity and, thus, lead to differences between prediction and experimental observations. With proper guidance, the presence of conflicting predictions and results may help students to better coordinate experimental evidence with theoretical knowledge.

Students working under the observe-infer frame were more likely to engage in inductive reasoning when analyzing the chemical behavior of unknown reactants. These students often relied on empirical observations to make sense of their laboratory experience (i.e. quick and heavy precipitation in the AgNO_3 in ethanol reaction conditions suggesting a strong preference for the $\text{S}_{\text{N}}1$ reaction mechanism). They also more frequently generated more sophisticated arguments in which they considered how their substances reacted as informed by their qualitative observations and built inferences based on theoretical principles about expected reactivity (Grimberg & Hand, 2009). Our results suggest that when working under this laboratory frame, student reasoning would benefit from prompts that lead students to explicitly make arguments about the hypothesized molecular structure of their unknown substances based on their data. Their arguments could also be enriched by providing a set of potential molecular structures for the reactants that should be matched with their unknowns by integrating evidence with theoretical knowledge about chemical reactivity of different substrates.

In general, our findings highlight the significant effects that the framing of the laboratory goals, procedures, information, and tools available to students can have on the arguments they built and the reasoning they express when making sense of their results. Although in this research study only one aspect of the laboratory experiment was reframed for students (whether or not they knew the structure of their starting materials), this minor change drastically affected the chemical concepts and ideas that students invoked, the type of reasoning they manifested, and the extent to which different argument components were complete, integrated, and aligned. These results suggest that laboratory designers should carefully reflect on the alignment between the learning objectives and the framing and nature of the activities in which students are to engage. Anticipating how the information and guidance provided to students may affect how they interpret the goals of an experiment and the arguments they built can help lab designers figure out how to create laboratory tasks that best elicit the type of reasoning that they value and foster the core understandings that they want to target.

Our findings, although not characterizing the effect of GSIs on student argumentation, lead us to speculate as to the importance of training laboratory instructors to become aware of how their presentation and guidance of laboratory activity can impact student work. It is often the case that graduate students are the primary instructors (GSIs) for laboratory courses in large universities. To align instruction with the goals of the course designer, it would be advantageous to train GSIs on how to facilitate argumentation in the laboratory with attention to the effects of activity framing on student reasoning. Different GSIs may have different interpretations of the goals of an experiment and provide students with various types of information to guide or facilitate their work. For example, a recent study by Grooms (2020) suggested that students' conceptions of evidence are impacted by the nature of their course instruction. Additionally, training GSIs to not only look for the description of

expected or "correct" results but also encouraging the construction of thoughtful rationales for unexpected observations or behaviors might promote more meaningful scientific reasoning amongst students. Thus, collective discussion and reflection on how different types of information can affect how students engage in argument may help align learning opportunities and outcomes across lab sections taught by different GSIs.

In this investigation, we focused our analysis on the effects of reframing some aspects of laboratory activity, such as the information provided to students about reactants that affected the ability to make predictions based on theoretical knowledge. It is likely that changes on, for example, the types of techniques, instruments, or data that is gathered may also affect the nature of the arguments that students build. Further research in this area is needed to fully understand the effects of framing on student argumentation in chemistry laboratories.

Limitations

Our major findings emerged from the analysis of data for a single type of experiment (substitution reactions) and additional studies are needed to determine whether similar conclusions can be derived when students work on different types of experiments with distinct goals, procedures, and types of data. Researchers collected post-lab reports from laboratories guided by different GSIs who may have presented laboratory goals, procedures, and information in different ways. Our data sample is too small to evaluate the impact that laboratory instructors may have had on the arguments built by participating students. Similarly, some of the collected reports were built by individual students while others represent the output of the discussion of small groups of students. The size of our sample does not allow us to evaluate the potential impact of this variable on student argumentation.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

We would like to thank all the graduate student instructors who cooperated with this research study and offered us access to their laboratory sections and students for data collection. We would also like to thank all the organic chemistry laboratory students who participated in this research study.

References

- Abi-El-Mona, I. and Abd-El-Khalick, F., (2006), Argumentation Discourse in a High School Chemistry Course, *School Science and Mathematics*, **106**(8), 349-361.
- Abi-El-Mona, I. and Abd-El-Khalick, F., (2011), Perceptions of the Nature and 'Goodness' of Argument among College Students, Science Teachers, and Scientists, *Inter. J. Sci. Educ.*, **33**(4), 573-605.

- Babai, R. and Levit-Dori, T., (2009), Several CASE lessons can improve students' control of variables reasoning scheme ability, *J. Res. Sci. Educ. Tech.*, **18**, 439-446.
- Bateson, G., (1972), A theory of play and fantasy, In G. Bateson (Ed.), *Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology*, 177-193, New York: Ballantine.
- Bell, P. and Linn, M. C., (2000), Scientific arguments as learning artifacts: designing for learning from the web with KIE, *Inter. J. Sci. Educ.*, **22**, 797-817.
- Berland, L. K. and Hammer, D., (2012), Framing for Scientific Argumentation, *J. Res. Sci. Teach.*, **49**(1), 68-94.
- Berland, L. K. and Reiser, B. J., (2011), How classroom communities make sense of the practice of scientific argumentation, *Sci. Educ.* **95**(2), 191-216.
- Brem, S. K. and Rips, L. J., (2000), Explanation and evidence in informal argument, *Cognitive Science*, **24**(4), 573-604.
- Burke, K. A. and Greenbowe, T. J., (2006), Implementing the Science Writing Heuristic in the Chemistry Laboratory, *J. Chem. Educ.*, **83**(7), 1032-1038.
- Carey, S. and Smith, C., (1993), On understanding the nature of scientific knowledge, *Educ. Psych.*, **28**, 235-251.
- Carmel, J. H., Herrington, D. G., Posey, L. A., Ward, J. S., Pollock, A. M., and Cooper, M. M., (2019), Helping Students to "Do Science": Characterizing Scientific Practices in General Chemistry Laboratory Curricula, *J. Chem. Educ.*, **96**, 423-434.
- Choi, A., Hand, B., and Greenbowe, T., (2013), Students' Written Arguments in General Chemistry Laboratory Investigations, *Res. Sci. Educ.*, **43**, 1763-1783.
- Cohen J., (1988), *Statistical Power Analysis for Behavioral Sciences*, 2nd edition, Hillsdale, NJ.
- Colthorpe, K., Abraha, H. M., Zimbardi, K., Ainscough, L., Spiers, J. G., Chen, H.-J. C., and Lavidis, N. A., (2017), Assessing students' ability to critically evaluate evidence in an inquiry-based undergraduate laboratory course, *Adv. Physiol. Educ.*, **41**, 154-162.
- Cooper, M. M. and Stowe, R. L., (2018), Chemistry Education Research—From Personal Empiricism to Evidence, Theory, and Informed Practice, *J. Chem. Educ.*, **118**, 6053-6087.
- Criswell, B., (2012), Framing Inquiry in High School Chemistry: Helping Students See the Bigger Picture, *J. Chem. Educ.*, **89**, 199-205.
- Cronje, R., Murray, K., Rohlinger, S., Wellnitz, T., (2013), Using the Science Writing Heuristic to Improve Undergraduate Writing in Biology, *Inter. J. Sci. Educ.*, **35**(16), 2718-2731.
- Cruz-Ramírez de Arellano, D. and Towns, M. H., (2014), Students' Understanding of Alkyl Halide Reactions in Undergraduate Organic Chemistry, *Chem. Educ. Res. Pract.*, **15**, 501-515.
- Domin, D. S., (1999), A Review of Laboratory Instruction Styles, *J. Chem. Educ.*, **76**(4), 543-547.
- Driver, R., Newton, P., and Osborne, J., (2000), Establishing the norms of scientific argumentation in classrooms, *Sci Educ.*, **84**(3), 287-313.
- Duschl, R., Schweingruber, H., and Shouse, A. (Eds.), (2007), *Taking science to school: Learning and teaching science in grades K-8*, Washington, DC: National Academies Press.
- Elby, A. and Hammer, D., (2010), Epistemological resources and framing: A cognitive framework for helping teachers interpret and respond to their students' epistemologies, In L. Bendixen and F. Feucht (Eds.), *Personal Epistemology in the Classroom: Theory, Research, and Implications for Practice*, 409-434, Cambridge: Cambridge University Press.
- Engelmann, K., Chinn, C. A., Osborne, J., and Fischer, F., (2018), *Scientific Reasoning and Argumentation: The Roles of Domain-Specific and Domain-General Knowledge*, New York, NY: Taylor & Francis.
- Erduran, S., Simon, S., and Osborne, J., (2004), TAPPING into Argumentation: Developments in the Application of Toulmin's Argument Pattern for Studying Science Discourse, *Sci. Educ.*, **88**, 915-933.
- Garcia-Mila, M., Gilabert, S., Erduran, S., and Felton, M., (2013), The Effect of Argumentative Task Goal on the Quality of Argumentative Discourse, *Sci. Educ.*, **97**(4), 497-523.
- Goffman, E., (1974), *Frame analysis: An essay on the organization of experience*, Cambridge, MA: Harvard University Press.
- Grimberg, B. I. and Hand B., (2009), Cognitive pathways: analysis of students' written texts for science understanding, *Int. J. Sci. Educ.*, **31**(4), 503-521.
- Grooms, J., (2020), A Comparison of Argument Quality and Students' Conceptions of Data and Evidence for Undergraduates Experiencing Two Types of Laboratory Instruction, *J. Chem. Educ.*, **97**(8), 2057-2064.
- Hand, B. and Choi, A., (2010), Examining the Impact of Student Use of Multiple Model Representations in Constructing Arguments in Organic Chemistry Laboratory Classes, *Res. Sci. Educ.*, **40**, 29-44.
- Havdala, R. and Ashkenazi, G., (2007), Coordination of Theory and Evidence: Effect of Epistemological Theories on Students' Laboratory Practice, *J. Res. Sci. Teach.*, **44**(8), 1134-1159.
- Jimenez-Aleixandre, M. P., (2008), Designing argumentation learning environments. In Erduran, S. and Jimenez-Aleixandre, M. P. (Eds.), *Argumentation in science education: Perspectives from classroom-based research*, 91-115, Dordrecht: Springer Academic Publishers.
- Jimenez-Aleixandre, M. and Erduran, S., (2008), Argumentation in Science Education: An Overview. In *Argumentation in science education: perspectives from classroom-based research*, Jimenez-Aleixandre, M. and Erduran, S. (Eds.), 3-27, Springer: Dordrecht.
- Jimenez-Aleixandre, M., Rodriguez, A., and Duschl, R. A., (2000), "Doing the lesson" or "doing science": Argument in high school genetics, *Sci. Educ.*, **84**(3), 287-312.

- Katchevich, D., Hofstein, A., and Mamlok-Naaman, R., (2013), Argumentation in the Chemistry Laboratory: Inquiry and Confirmatory Experiments, *Res. Sci. Educ.*, **43**, 317-345.
- Kelley C., (2019), Thinking Through the Laboratory: An Organic Chemistry I Workbook, 1st edition, Dubuque, IA: Kendall Hunt.
- Keys, C. W., Hand, B., Prain, V., and Collins, S., (1999), Using the Science Writing Heuristic as a Tool for Learning from Laboratory Investigations in Secondary Science, *J. Res. Sci. Teach.*, **36**(10), 1065-1084.
- Kuhn, D., (1991), The skills of argument, Cambridge, England: Cambridge University Press.
- MacLachlan, G. L. and Reid, I., (1994), Framing and interpretation, Victoria: Carlton.
- McNeill, K. L. and Krajcik, J., (2007), Middle school students' use of appropriate and inappropriate evidence in writing scientific explanations. In M. Lovett and P. Shah (Eds.), Thinking with data: The proceedings of 33rd Carnegie Symposium on Cognition. Mahwah, NJ: Erlbaum.
- McNeill, K. L. and Krajcik, J., (2009), Synergy Between Teacher Practices and Curricular Scaffolds to Support Students in Using Domain-Specific and Domain-General Knowledge in Writing Arguments to Explain Phenomena, *J. Learn. Sci.*, **18**, 416-460.
- Microsoft Excel 2016, (2016), Redmond, WA: Microsoft.
- Moon, A., Moeller, R., Gere, A. R., and Shultz, G. V., (2019), Application and testing of a framework for characterizing the quality of scientific reasoning in chemistry students' writing on ocean acidification, *Chem. Educ. Res. Pract.*, **20**, 484-494.
- National Research Council, (2012), A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas, Washington, DC: National Academies Press.
- Osborne, J. F., (2010), Arguing to learn in science: the role of collaborative, critical discourse, *Science*, **328**, 463-466.
- Pabuccu, A. and Erduran, S., (2017), Beyond Rote Learning in Organic Chemistry: The Infusion and Impact of Argumentation in Tertiary Education, *Inter. J. Sci. Educ.*, **39**, 1154-1172.
- Reiser, B. J., Tabak, I., Sandoval, W. A., Smith, B. K., Steinmuller, F., and Leone, A. J., (2001), BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms, In S. M. Carver and D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress*, 263-305, Mahwah, NJ: Erlbaum.
- Rodriguez, J.-M. G. and Towns, M. H., (2018), Modifying Laboratory Experiments To Promote Engagement in Critical Thinking by Reframing Prelab and Postlab Questions, *J. Chem. Educ.*, **95**, 2141-2147.
- Sampson, V. and Clark, D. B., (2006), Assessment of Argument in Science Education: A Critical Review of the Literature, *ICLS 2006 - International Conference of the Learning Sciences, Proceedings*, **2**, 655-661.
- Sampson, V., Grooms, J., and Walker, J. P., (2010), Argument-Driven Inquiry as a Way to Help Students Learn How to Participate in Scientific Argumentation and Craft Written Arguments: An Exploratory Study, *Sci. Educ.*, **95**, 217-257.
- Sampson, V. and Walker, J., (2012), Learning to write in the undergraduate chemistry laboratory: The impact of Argument-Driven Inquiry, *Inter. J. Sci. Educ.*, **34**(10), 1443-1485.
- Sandoval, W. A., (2003), Conceptual and epistemic aspects of students' scientific explanations, *J. Learn. Sci.*, **12**(1), 5-51.
- Sandoval, W. A. and Millwood, K. A., (2005), The Quality of Students' Use of Evidence in Written Scientific Explanations, *Cognition and Instruction*, **23**(1), 23-55.
- Sommerhoff, D., Ufer, S., and Kollar, I., (2015), Research on mathematical argumentation: A descriptive review of PME proceedings, In K. Beswick, T. Muir, & J. Wells (Eds.), *Proceedings of the 39th conference of the international group for the psychology of mathematics education*, **4**, 193-200, Hobart, Australia: TME.
- Stowe, R. L. and Cooper, M. M., (2019), Arguing from Spectroscopic Evidence, *J. Chem. Educ.*, **96**(10), 2072-2085.
- Syed, M. Q., (2015), Going beyond equations with disciplinary thinking in first-year physics, *J. Coll. Teach. Learn. (Online)*, **12**(2), 127.
- Tannen, D., (1993), Framing in discourse, New York: Oxford University Press.
- Walker, J. and Sampson, V., (2013), Argument-Driven Inquiry: Using To Improve Undergraduates' Science Writing Skills through Meaningful Science Writing, Peer-Review, and Revision, *J. Chem. educ.*, **90**, 1269-1274.
- Walker, J. and Sampson, V., (2013), Learning to Argue and Arguing to Learn: Argument-Driven Inquiry as a Way to Help Undergraduate Chemistry Students Learn How to Construct Arguments and Engage in Argumentation During a Laboratory Course, *J. Res. Sci. Teach.*, **50**(5), 561-596.
- Walker, J. P., Sampson, V., Grooms, J., Anderson, B., and Zimmerman, C. O., (2012), Argument-Driven Inquiry in Undergraduate Chemistry Labs: The Impact on Students' Conceptual Understanding, Argument Skills, and Attitudes Toward Science, *J. Coll. Sci. Teach.*, **41**(4), 74-81.
- Walker, J. P., Sampson, V., and Zimmerman, C. O., (2011), Argument-Driven Inquiry: An Introduction to a New Instructional Model for Use in Undergraduates Chemistry Labs, *J. Chem. Educ.*, **88**, 1048-1056.
- Walker, J., Van Duzor, A. G., and Lower, M. A., (2019), Facilitating Argumentation in the Laboratory: The Challenges of Claim Change and Justification by Theory, *J. Chem. Educ.*, **96**, 435-444.