# EXPLORING ARGUMENTATION IN THE UNDERGRADUATE ORGANIC CHEMISTRY LABORATORY

by

Steven Petritis

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2022

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by: Steven Petritis

titled: Exploring Argumentation in the Undergraduate Organic Chemistry Laboratory

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Vicente Talanquer Date: Jan 7, 2022

Molly Bolger Date: Jan 7, 2022

Hamish S. Christie
Hamish Christie Date: Jan 7, 2022

John Jewett
John Jewett Date: Jan 7, 2022

Colleen Kelley
Colleen Kelley Date: Jan 7, 2022

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Date: Jan 7, 2022

Vicente Talanquer
Department of Chemistry and Biochemistry

**Signature:** Steven Petritis
**Email:** petritis@email.arizona.edu

**Signature:** Kristi R Davenport
**Email:** kdavenport@email.arizona.edu

# Acknowledgments

Throughout my educational journey, I have been extremely fortunate to have been in the presence of some amazing people, including my family, friends, and academic colleagues at the various stages of my career. Finishing graduate school is an amazing accomplishment that, at times, I felt would never come to fruition. They deserve all the credit for making me the person I am today and none of this would have been possible without their continual guidance and support.

First and foremost, I would like to thank Vicente Talanquer for taking a chance on me when I felt lost and without purpose as a chemistry graduate student. You opened my eyes to education research and transformed the way I see the chemistry classroom and laboratory. Your guidance reinvigorated my passion for research. You always challenged me to think more critically about education from every angle, search for answers to the impossible questions, and strive to find my own path in my research career. Thank you for leading by example and guiding me towards a discipline that is the perfect mesh of my interests in teaching and learning.

Thank you very much, Colleen Kelley. When you came to the University of Arizona at the same time that I was transitioning into chemistry education research, I had a gut feeling that I needed to follow through with this decision. That turned out being one of the best decisions I have ever made and I would have never realized my true passion without your support, understanding, and the sheer joy you bring to teaching chemistry. Thank you very much for all the guidance you have given me over the years, as a teaching supervisor, research committee member, FLC co-facilitator, inter-rate reliability coder, co-author, and, most importantly, a wonderful mentor and dear friend. Without your endless help, none of this would have been possible.

I want to thank my other research committee members, Molly Bolger, Hamish Christie, and John Jewett. John and Hamish, I want to thank you both for your genuine support from the very start of my time in graduate school. You both encouraged me to push myself to become a better chemist, whether in the research lab, teaching lab, or as an observer of those learning chemistry. You challenged me to consider how to further support students, as both a graduate teaching assistant and as a researcher, and I will never forget the many insights and thought-provoking conversations I have shared with both of you. Molly, through my interactions with you and your research group, I was further pushed to think more carefully about the intricacies of conducting education research. You always found the perfect balance between supporting my research interests and productively challenging my ideas and findings, which ultimately helped to shape the type of researcher I am today. Thank you so much for all of your support.

I would also like to thank the various groups of academic and professional colleagues that have each provided with me unique opportunities to engage with teaching and learning. I would like to thank Erin Galyen, Mascha Gemein, and Gretchen Gibbs from the Office of Instruction and Assessment (OIA) at the University of Arizona. The three of you revolutionized the way I see teaching and learning and I am forever in debt to you for your support and kindness throughout our interactions.

Thank you to Gail Burd, Sarah Grace, and Kristin Winet for all of our wonderful conversations through my work with the Faculty Learning Communities (FLCs) and the Center for Integration of Research, Teaching, and Learning (CIRTL). You provided a wonderful outlet for my passion of teaching, expanded my connections to like-minded academic colleagues, and further transformed my vision for education.

Thank you very much to Katherine Byrd and Will Schneller for helping me foster a connection to the educational industry and for providing me the invaluable experience of seeing educational technology through a commercial lens. Through our interactions and various projects, you both helped me grow to think about complex problems regarding the implementation of educational materials and technology to a wider audience of students.

Thank you very much to my Growing Mathletes team, including Christina Baze, Sanlyn Buxner, Chris Rendes, Erin Turner, and Ricardo Valerdi. You all provided me the unique opportunity to conduct research with youth learning about America's greatest sport, baseball, and propelled me to utilize my experience in educational research in an exciting and different way.

Thank you to everyone working in the Teaching Support Office (TSO) in the Department of Chemistry & Biochemistry (CBC), my fellow graduate teaching assistants, and, of course, all of my students from over the years. Some of my greatest memories of graduate school have been my time teaching in the chemistry laboratories and I am forever grateful of the opportunity to interact with all of you.

I would also like to say a special thank you to a few people whose mentorship has been particularly impactful to me in my academic journey. Ritika Gautam and Eman Akam, thank you so much for welcoming me to graduate school with open arms and for your undying support through my early struggles in graduate school. Both of you are tremendous scientists and even better people. Needless to say, I would not have made it through my early years of graduate school if it was not for both of you. Erick Leggans, thank you for mentorship and friendship during my time at Grinnell College and beyond. You single-handedly fueled my passions for chemistry and, without your guidance, I would not have even known that graduate school or a career in chemistry was an option. Thank you so much for being such a kind-hearted, compassionate, and all around amazing person who was always there to lend a helping hand.

Thank you so much to all of my closest friends who have helped to make the journey through school much more enjoyable. I am so lucky to have gotten to know all of you, through the good times and bad. I would not be the person I am today if it was not for your friendship, support, and everything I have learned through getting to know each and every one of you. Thank you to my graduate school friends, including Nathan Bedard, Jacob Fussell, and Chris Foley. Thank you very much to my Grinnell College baseball friends, including Samuel “Atticus” Galanek, Matthew Hammond, Yuki Kawahara, Andrew Kelley, Jason Ormond, Will Pahos, Matthew Schaeffer, and Ross Voelker. Getting to know you all has meant the world, I will forever cherish our time together, and am excited for all of our memories to come.

I would like to extend a special thank you to my two fur children, Luigi and Ember. You both have brought me nothing but pure joy and I am so thankful for the happiness you both have brought me. Your unceasing affection has helped me through many challenging and tough times and your presence and loyalty has left an eternal paw print on my life.

Garrett Davis, your friendship have been the one constant throughout our time together in graduate school. Our many laughs, cries, and fun memories from the past several years have left a lasting impact on my life and I could not have asked for a better best friend to have gone through it all with. You are one of the most genuine, honest, and thoughtful guys that I know and I am so thankful that you have always been there to have my back. Although graduate school was certainly not easy and was downright awful at times, having you there as my best friend helped make it bearable and worth it in the end. Thank you for always being there for me when I needed it most and when I didn’t even know I needed it. I love you, man.

Eliza Short, you mean more to me than anything else in the world and I am so lucky and thankful to have you in my life. Not only did you make the end of my time in graduate school the best year in my life, but you gave me hope for the future and helped me see a life beyond school. You have taught me so much by being such a thoughtful, caring, honest, and amazing person. I will always hold the last year of graduate school in a special place in my heart and am looking forward to a lifetime of memories that await us. I love you so much.

Of course, I want to thank my amazing family for their constant love and support from my first day of Kindergarten through my final days of graduate school. Although school has certainly pulled me away from creating more wonderful memories with all of you, I am forever grateful of all the time we have had together and of the boundless support you have shown me over the years. Aunt Paula, thank you for always being there to ask me how school is doing and listen to me talk about the highs and lows of every step along the way. Your enthusiasm for learning from others served as the perfect example of how to go about my own academic career. I will always remember your love, support, and the role you played throughout my educational journey.

Dad, you were the very first person who got me interested in science and always pushed me to learn more and work hard. More importantly, you instilled in me the value of being a good person above all. Even though we have not always seen eye-to-eye, you have had a tremendous impact on my life and on my journey through school. I will always be grateful of the example you set for me and how much you have taught me through your selflessness, love, and support of others. Thank you and I love you.

Mom, you have been there for me since day one and have always been my number one supporter. You pushed me to excel in school from the start and never settled for anything less than my best. But more importantly, you sacrificed a lot to make sure that I had the opportunity to be successful and go to school to learn about the things that I loved. You put me above you every step of the way and helped build me into the man I am today. Even when school was tough, busy, and took me away from home, you remained my number one supporter and always encouraged me to do my best and follow my dreams. Above all, you just wanted me to be happy and, though all of your love and support, that is exactly what I am. I love you so much, Mom.

# Dedication

Mom,

This work and all of my schooling over the years is dedicated to you.

Thank you for always pushing me to be my best and follow my dreams.

I love you to infinity and beyond!

# Table of Contents

# List of Figures

# List of Tables

# Abstract

Recent research emphasis has been placed on studying how students engage in scientific argumentative reasoning in the chemistry laboratory setting. Although several studies have evaluated the quality and characterized the structure of student arguments following the claims, evidence, and rationale (CER) framework, little is known about the influence of various laboratory factors on student reasoning. In this study, we seek to better understand what factors affect student argumentation in laboratories and how these factors foster or hinder students' integration of core chemistry concepts and laboratory data. We have identified several factors associated with students' laboratory experiences and aim to gain additional insight by exploring how students' use of specific chemical data in various types of experiments impacts both the nature and quality of their post-lab arguments. Ultimately, this work highlights the need for explicit consideration of these factors in designing opportunities for undergraduate chemistry laboratory students to engage in productive argumentation from evidence.

In the first part of this project, 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.

Building on our previous work involving activity framing, the second part of this project involved the analysis of student arguments produced following eight experiments that comprise the first semester of a college organic chemistry laboratory to identify other factors that may significantly affect the nature and quality of student argumentation in undergraduate organic chemistry labs. Our analysis revealed no trends on the effect of experiment order or general type on the quality of student arguments; however, the amount and types of data sources as well as the level of scaffolding provided both had an impact on student argument quality. Although the undergraduate laboratory offers a ripe opportunity for students to engage in argument from evidence, laboratory activity involves a complex web of components each with the potential to affect productive and quality sensemaking. Our findings highlight the importance of explicit consideration of various laboratory factors and their impact on how students express their chemical reasoning through written argumentation.

# Chapter 1: Introduction

Scientists routinely engage in argument from evidence as they try to make sense of observable phenomena (National Research Council, 2012). To train the next generation of research and industry professionals, laboratory courses need to create opportunities for students to develop various epistemic practices. 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 formulate their own research questions, carry out their own investigations, analyze and interpret data, and engage in argumentation with the appropriate evidence as they search for meaning in their laboratory findings. While engaging in these scientific practices, students are encouraged to coordinate theoretical constructs and physical observations through their experimental work as they search for meaning in data collected from the observable world. This integration of the abstract and the empirical offers a ripe opportunity for students to build more meaningful understandings. By participating in these authentic scientific experiences, students can also develop a more meaningful understanding of the role of chemistry in everyday life.

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. Building arguments from evidence is one of the core epistemic practices of science (Crujeiras-Pérez and Jiménez-Aleixandre, 2017), and chemistry educators should strive to better understand how to create robust opportunities for undergraduate students to formulate scientific arguments in the instructional laboratory. Although there is much agreement that engaging in productive argumentation is beneficial for student learning, our understanding of how different laboratory factors and aspects of experimental task design, implementation, and assessment affect the nature and quality of students argumentation in the college chemistry laboratory is much more limited.

The central goal of this work is to characterize students' arguments in the organic chemistry laboratory course and qualify the impact of various laboratory factors on the nature and quality of student written argumentation. Identifying and characterizing the impact of such laboratory factors serves to benefit both the laboratory instructional team and students taking laboratory courses. Findings from this work can help inform laboratory course designers and managers on considerations that must be made when engaging their students in laboratory-based argumentation. With these ideas in hand, our hope is that more evidence-based implementations can be utilized to foster meaningful engagement in the epistemic practice of argument from evidence. Our work adds to the existing body of literature (albeit small) on student laboratory argumentation and highlights various aspects of students' laboratory experiences and their impact on how they manifest their meaning-making process as they communicate their findings. Ultimately, greater awareness of how students approach laboratory challenges in constructing their reasoning can inform the chemistry education community on instructional strategies that better engage students in the science practice of argumentation.

This work provides significant contributions to both the research and practice communities within the chemistry educational community. First, by establishing a robust analytical tool for argumentation analysis we provide the framework for other researchers to dig more deeply into their own argumentation analysis and add to a growing body of literature that involves laboratory argumentation. Our novel considerations of laboratory factors that impact student reasoning sets the stage for follow up studies (both by us and others) that involved the impact of different factors on laboratory argumentation. Further, this work could easily be extended to the chemistry classroom and other contexts to investigate how aspects of different learning environments impact the way in which students invoke chemical ideas in their epistemic practice. Practically speaking, our findings lay the groundwork for laboratory course designers to consider the impact of different components of the lab on how students construct reasoning based on their laboratory findings. Explicit consideration of these components can ultimately help laboratory managers more meaningfully engage students in argumentation and help them develop this epistemic practice over the course of a semester in the chemistry laboratory.

The first part of this dissertation provides a review of argumentation literature related to the chemistry laboratory, drawing from other disciplines and instructional environments to provide more context for our research studies. This chapter also includes a brief summary of several laboratory factors that are relevant to the studies we describe herein. The second part of this work sets the stage for our two major research projects, including the research and practice problems we intend to address with our work. Following this overview, we detail our first major work involving the activity framing of a laboratory experiment (Chapter 3) and present our evidence and findings as to the impact that frame change has on students' approaches to reasoning. The third part of this dissertation builds on the second part by identifying additional laboratory factors and characterizing their impact on both the nature and quality of student arguments from experiments spread across an entire semester of an organic chemistry laboratory course. Finally, we summarize our findings, discuss the research and practical implications of our work, and present additional data collection and interventions that have been conducted involving argumentation in the organic chemistry laboratory.

# Chapter 2: Review of Literature

## Research on chemistry laboratory argumentation

Argumentation is a process of "logical discourse whose goal is to tease out the relationship between ideas and evidence" (Duschl *et al.*, 2007). Further, arguments in the classroom and the laboratory are "tool[s] for understanding student reasoning, engagement in scientific practices, and development of conceptual and epistemic understanding" (Kelly & Takao, 2002). Of the eight science practices identified by the National Research Council (2012), "engaging in argument from evidence" has become increasingly commonplace in science education at all educational levels (Berland & Reiser, 2011). 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). Thus, in recent years chemistry education researchers have sought to gain a better and deeper understanding of how students engage in this epistemic practice.

In chemistry, educational research in this area has provided insights into how different types of learners engage in this practice in the classroom. Research in chemistry classrooms has often relied on Toulmin's framework for argumentation to characterize how students coordinate "evidence and theory to support or refute an explanatory conclusion, model or prediction" (Jiménez-Aleixandre & Erduran, 2008). Research on student argumentation in chemical contexts has been carried out in diverse educational settings, including elementary schools (Park *et al.*, 2020; Soysal and Yilmaz-Tuzun, 2021), secondary schools (Çetin, 2021; Juntunen and Aksela, 2014; Grooms *et al.*, 2018), and in undergraduate courses in a variety of chemistry subdisciplines (Cruz-Ramírez de Arellano and Towns, 2014; Kelly and Takao, 2002; Moon *et al.*, 2019).

Meanwhile, the undergraduate chemistry laboratory has seen deployment of a variety of evidence-based instructional models to facilitate productive engagement in science practices (Abi-El-Mona and Abd-El-Khalick, 2006; Abi-El-Mona and Abd-El-Khalick, 2011). 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). Educational 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).

## Evaluation of student argument quality

A significant body of work analyzing chemistry laboratory argument quality has been completed using the Science Writing Heuristic (SWH) and the Argument-Driven Inquiry (ADI) frameworks, which were developed to help foster and assess the efficacy of argumentation-based laboratory curricula (Keys *et al*., 1999; Sampson *et al*., 2010). 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.

The SWH framework was designed to increase student engagement and learning in the chemistry laboratory while allowing students to carry out their own investigations, making claims based on their own data, and reflect on what they learned from their laboratory experience (Burke and Greenbowe, 2006; Hand and Choi, 2010). 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). Subsequent research under the SWH umbrella has developed analytical and holistic frameworks to evaluate the quality of student argumentation and showed correlation between argument quality and academic performance in the general chemistry classroom course achievement (Choi *et al*., 2013).

The ADI instructional model, on the other hand, guides students to produce tentative arguments in the form of claims, evidence, and reasoning, and share their arguments with their classmates by participating in a peer review process before revising their final arguments (Sampson *et al.*, 2010; Walker *et al.*, 2011). Using the ADI model, Walker & Sampson (2012) observed an increase in both oral and written argumentation scores over the course of a semester as students engaged with the argument-centered curriculum. Subsequent work has supported these findings by noting similar increases in oral and written argumentation quality when following this type of argument-centered laboratory curriculum (Walker and Sampson, 2013a; Çetin and Eymur, 2017). In both frameworks, learners' competence in developing arguments is fostered by creating multiple opportunities for students to make claims, support their findings with evidence, and provide solid rationales to support their ideas (Walker, *et al*., 2012; Walker, Van Duzor, and Lower, 2019). Overall, work within these two approaches has demonstrated increased quality of written arguments over time in both general chemistry and organic chemistry laboratory courses (Hosbein *et al*., 2021; Hand and Choi, 2010; Walker and Sampson, 2013a).

Other instructional approaches have also demonstrated efficacy at promoting increased student argument quality. 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, Hofstein, and Mamlok-Naaman (2013) argue that inquiry-based laboratory experiments provide students an effective platform to conduct argumentation and, thus produced higher quality arguments compared to those produced as a result of 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). Results from studies in chemistry laboratories are in line with those from investigations in biology and physiology education that have shown increased engagement with argumentation and conceptual knowledge as a result of using inquiry-based laboratory curricula (Carmel *et al.*, 2019; Cronje *et al.*, 2013; Colthorpe *et al.*, 2010; Reiser *et al.*, 2001). These studies reveal clear alignment between higher argument quality and laboratory curricula that promote engagement in science practices and argumentation. However, more research is needed to identify which specific features of these laboratory implementations lead students to construct higher quality arguments.

## Characterization of student 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. Characterization of student arguments has emphasized a domain-general approach to understanding how students build their claims, select evidence, and generate rationales. Analytical and holistic scoring frameworks have been used to evaluate the different components of an argument and their relationships (Choi, Hand, & Greenbowe, 2013). For example, several studies have shown that students are more likely to construct higher quality arguments when they adequately connect their claims and evidence (Sandoval and Millwood, 2005; Choi *et al.*, 2013; Katchevich *et al.*, 2013; Walker and Sampson, 2013b; Grimberg and Hand, 2009). 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).

Unfortunately, Kuhn (1991) showed that students often provide unsubstantiated claims that lack evidentiary support, and Brem and Rips (2000) indicated that students sometimes replaced missing evidence with highly implicit reasoning. Students may also fail to explicitly link empirical evidence to their claims, present their arguments as self-evident without adequate support from their experimental findings (Brem and Rips, 2000; McNeill and Krajcik, 2007), and focus on individual pieces of laboratory data or evidence when building arguments (Bell and Linn, 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. Nevertheless, they seem capable of properly connecting claims and evidence as well as coordinating theory with empirical data in particular contexts (McNeill and Krajcik, 2009; Jiménez-Aleixandre, 2008).

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).

Although students' general abilities to build arguments have been investigated thoroughly, there are fewer reports on how learners use specific disciplinary knowledge to coordinate various argument components. For example, Stowe and Cooper (2019) found that students could analyze and integrate spectroscopic data from various sources in their arguments, but they struggled to connect their evidence to a reasonable chemical claim. Thus, more research is needed to understand when and how students build such connections, and what educational structures and strategies help them coordinate these elements properly (Carey and Smith, 1993; Driver *et al.*, 2000; Sampson and Clark, 2008).

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. Additionally, we identified several other factors associated with students' laboratory experience including: experiment order, experiment type, the amount and types of data sources available to students, and the explicit argumentation scaffolding provided to students. We sought to characterize the impact of these factors on both the nature and quality of student post-lab written argumentation.

## Factors in the chemistry laboratory

The chemistry laboratory is a complex learning environment in which students need to consider a wide array of empirical evidence and integrate their data with diverse chemical ideas and concepts to build meaning based on their laboratory findings. Adding to this intricate web, different experiments are designed with various explicit and implicit goals, outcomes, approaches, and procedures (Domin, 1999). Criswell (2012) described the chemistry laboratory as a complex combination of context, goals, actions, tools, and interactions, each of which has the potential to impact student argumentation. Existing research on argumentation in the laboratory alludes to various factors that can impact students' experiences and ability to coordinate empirical evidence and theoretical concepts (Kadayifci *et al.*, 2012). These factors include the activity framing of a laboratory experiment (Berland and Hammer, 2012), explicit prompts and instructional scaffolds (Cooper, 2015; Cooper and Stowe, 2018; McNeill, Lizotte, Krajcik, and Marx, 2006), instructional contexts and styles (Berland and McNeill, 2010; Smith, Wiser, Anderson, and Krajcik, 2006; Grooms, 2020), and the use of experimental data in arguments (Stowe and Cooper, 2019). Below, we provide a brief summary of some of the previous research involving these laboratory factors.

### *Activity framing*

The frame of a student's educational experience is defined 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" (Hammer *et al.*, 2005). Existing research suggests that students' perceived frames impact their engagement and participation in argumentation (Berland and Hammer, 2012).

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-Aleixandre, 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, in which students focus their attention on what they think their instructor will value (Berland and Hammer, 2012), depending on how students coordinate past experiences with their current environment.

### ***Argumentation scaffolds***

The concept of scaffolding was originally introduced in the context of adult-child interactions in the elementary classroom (Wood, Bruner, and Ross, 1976), in which it was noted that learners can begin to develop deeper understandings of topics and practices through the use of instructional scaffolds (Bransford, Brown, and Cocking, 2000). Similarly, students often enter the science classroom or laboratory with little to no experience with various science practices including argument from evidence. Previous studies have also demonstrated that students struggle to interpret the purpose of the augmentative task largely due to their inexperience with the conventions of this epistemic practice (Berland and Hammer, 2012; Garcia-Mila *et al*., 2013).

To better engage students in productive argumentation with their laboratory data, explicit scaffolding is likely to benefit students who become familiar with this practice in the context of analyzing their data and communicating their laboratory findings. Several researchers have noted that fading away scaffolded prompts helps to promote higher quality argumentation across the course of a semester (McNeill *et al*., 2006; Berland and McNeill, 2010; Smith *et al*., 2006). Other examples have also highlighted that designing curricula around writing scaffolds can increase student engagement in scientific inquiry and argumentation in addition to encouraging reflection and metacognitive practices (White and Frederiksen, 1998; White and Frederiksen, 2000). Thus, the degree of scaffolding present in laboratory activities has the potential to impact student argumentation.

### ***Instructional contexts and styles***

Several researchers have noted that the practice of argumentation holds an important place in the science classroom as students construct knowledge with their peers (Berland and McNeill, 2010; Duschl, Schweingruber, and Shouse 2007). Despite the need to argue, typical classroom contexts focus students' argumentative efforts on engaging in persuasive discourse rather than knowledge construction (Newton, Driver, and Osborne, 1999; Hogan and Corey, 2001; Lemke, 1990). These findings have emphasized the need to foster a productive environment for argumentation as well as setup instructional design that affords students a fair chance at developing argumentation skills over the course of a semester (Smith *et al.*, 2006). Stemming from these challenges, researchers have suggested the need to utilize a learning progression for argumentation in the science classroom (Berland and McNeill, 2010). However, instructional implementations are needed to afford students the opportunity to develop these practices. Smith *et al.* (2006) commented that "learning progressions are not developmentally inevitable but depend on instruction" and argumentative discourse requires intentional design by classroom and laboratory designers. Thus, utilizing argumentative scaffolds has the potential to promote productive engagement in argument from evidence.

### *Experimental data in arguments*

Another major component of the undergraduate chemistry laboratory that is imperative to the process of argumentation is data collection, analysis, and interpretation. Although several researchers have noted the relationship between high-quality arguments and a sufficient link between students' claims and evidence components (Sandoval and Millwood, 2005; Choi *et al.*, 2013; Katchevich *et al.*, 2013; Walker and Sampson, 2013b; Grimberg and Hand, 2009; Sandoval and Millwood, 2005), much less is known about how the nature of experimental data impacts student written argumentation. Thus, it is important to consider how the types of data students collect in the chemistry laboratory may impact the nature and quality of their arguments produced in different laboratory contexts.

Stowe and Cooper (2019) found that students were proficient with analyzing both IR and $^{1}H/^{13}C$ NMR spectroscopic data, but demonstrated an inability to develop claims that were consistent with the entirety of their data sets even when provided with additional scaffolding. Choi, Hand, and Greenbowe (2013) also described how students often conflate theoretical constructs and experimental evidence in their arguments. Meanwhile, Grooms (2020) noted that students shifted their conceptions of permissible data towards more investigation-driven, empirical evidence rather than theoretical concepts; however, this study also attributed this shift to changes in explicit instruction. Nonetheless, there is not a clear cut idea of how experimental data impacts the nature and quality of student argumentation, especially in the context of the undergraduate chemistry laboratory.

# Chapter 3: Exploring Argumentation in the Undergraduate Organic Chemistry Laboratory – Part I: Activity Framing

## Problem statement and research questions

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 these advances, relatively little is known about how various aspects of the chemistry laboratory affect student argumentation. Guided by these ideas and literature precedent, our work attempted to fill this void by identifying factors within an undergraduate organic chemistry laboratory course and characterize their impact on both the nature and quality of student written argumentation.

In the first stage of our research project, we developed a qualitative coding framework that helped us characterize the nature of student argumentation and characterized differences and similarities in the arguments built by students engaged in a single laboratory 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.

In the second part of our research project (see Chapter 4), we expanded the use of our analytical framework for student arguments to a wider set of organic chemistry experiments seeking to identify other factors that may also significantly affect the nature of student argumentation in undergraduate organic chemistry labs. We also developed a quantitative method for measuring the quality of written arguments to assess the impact of additional laboratory factors on student argument quality. These factors included experiment order, experiment type, the amount and type of data sources, and scaffolding for student argumentation.

## Exploring the impact of activity framing on student argumentation

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).

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 (Rodriguez & 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.

Despite advances in argumentation research, 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 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.

## 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:

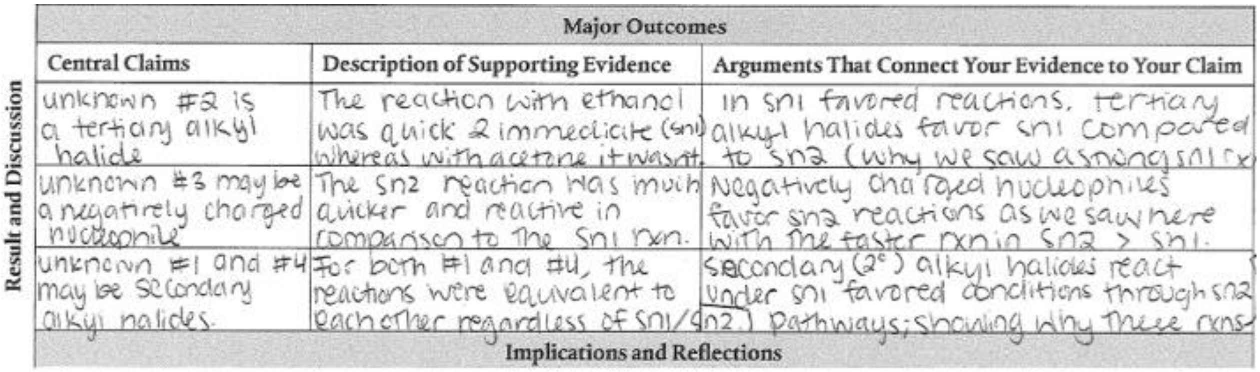
(1) How did the frame of the laboratory activity impact the domain-specific chemical concepts and ideas expressed in the claims, evidence, and rationales of student post-lab arguments?

(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?

### 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

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



| | Major Outcomes | | |
|---|---|---|---|
| | Central Claims | Description of Supporting Evidence | Arguments That Connect Your Evidence to Your Claim |
| Result and Discussion | unknown #2 is a tertiary alkyl halide | The reaction with ethanol was quick & immediate (sn1) whereas with acetone it wasn't. | in sn1 favored reactions, tertiary alkyl halides favor sn1 compared to sn2 (why we saw a strong sn1 rx |
| | unknown #3 may be a negatively charged nucleophile | The sn2 reaction was much quicker and reactive in comparison to the sn1 rxn. | Negatively charged nucleophiles favor sn2 reactions as we saw here with the faster rxn in sn2 > sn1. |
| | unknown #1 and #4 may be secondary alkyl halides. | For both #1 and #4, the reactions were equivalent to each other regardless of sn1/s | secondary (2°) alkyl halides react under sn1 favored conditions through sn2 n2.) pathways; showing why these rxns |
| Implications and Reflections | | | |

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).

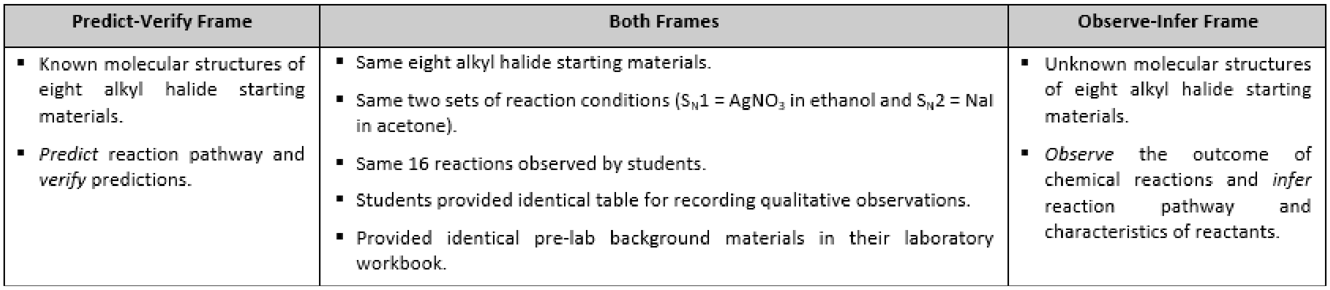
The objectives of this experiment, according to the laboratory workbook, were for students to investigate:

(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).

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



| Predict-Verify Frame | Both Frames | Observe-Infer Frame |
|---|---|---|
| ▪ Known molecular structures of eight alkyl halide starting materials.<br>▪ *Predict* reaction pathway and *verify* predictions. | ▪ Same eight alkyl halide starting materials.<br>▪ Same two sets of reaction conditions ($S_N1 = AgNO_3$ in ethanol and $S_N2$ = NaI in acetone).<br>▪ Same 16 reactions observed by students.<br>▪ Students provided identical table for recording qualitative observations.<br>▪ Provided identical pre-lab background materials in their laboratory workbook. | ▪ Unknown molecular structures of eight alkyl halide starting materials.<br>▪ *Observe* the outcome of chemical reactions and *infer* reaction pathway and characteristics of reactants. |

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.

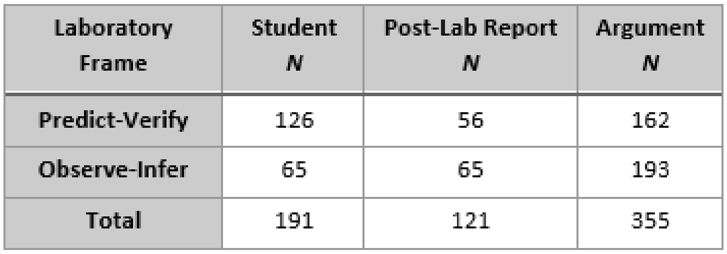
***Observe-infer laboratory frame.*** In this case, students worked with the same eight compounds using the two sets of reaction conditions described above. However, the identities of these molecules remained concealed. Although students did know that each of these compounds were organic halides, they 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, biochemistry, 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 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

**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 |

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.

### 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 were 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.

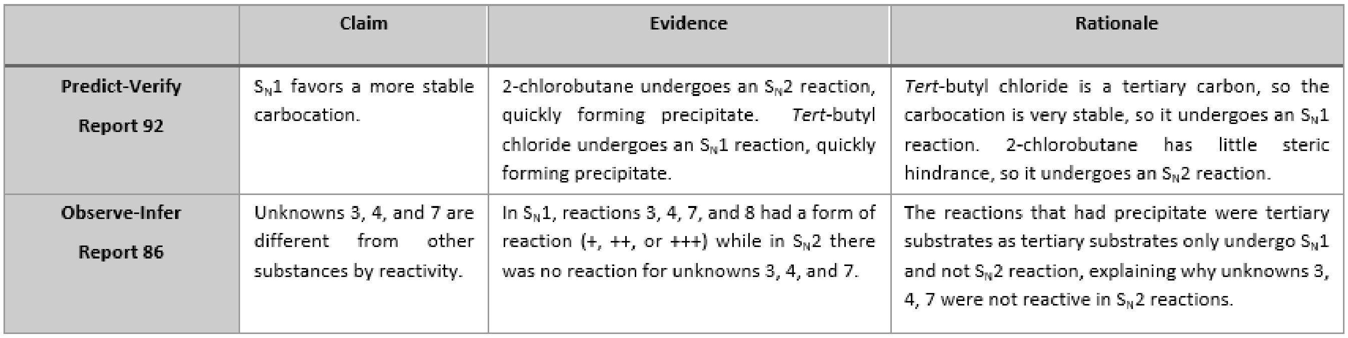
The "molecular structure" code highlighted references to the shape, size, or structure-related interactions (*e.g.,* steric hindrance) of the molecules involved in the substitution reactions.

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”).

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

**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**<br>**Report 92** | $S_N1$ favors a more stable carbocation. | 2-chlorobutane undergoes an $S_N2$ reaction, quickly forming precipitate. *Tert*-butyl chloride undergoes an $S_N1$ reaction, quickly forming precipitate. | *Tert*-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**<br>**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. |

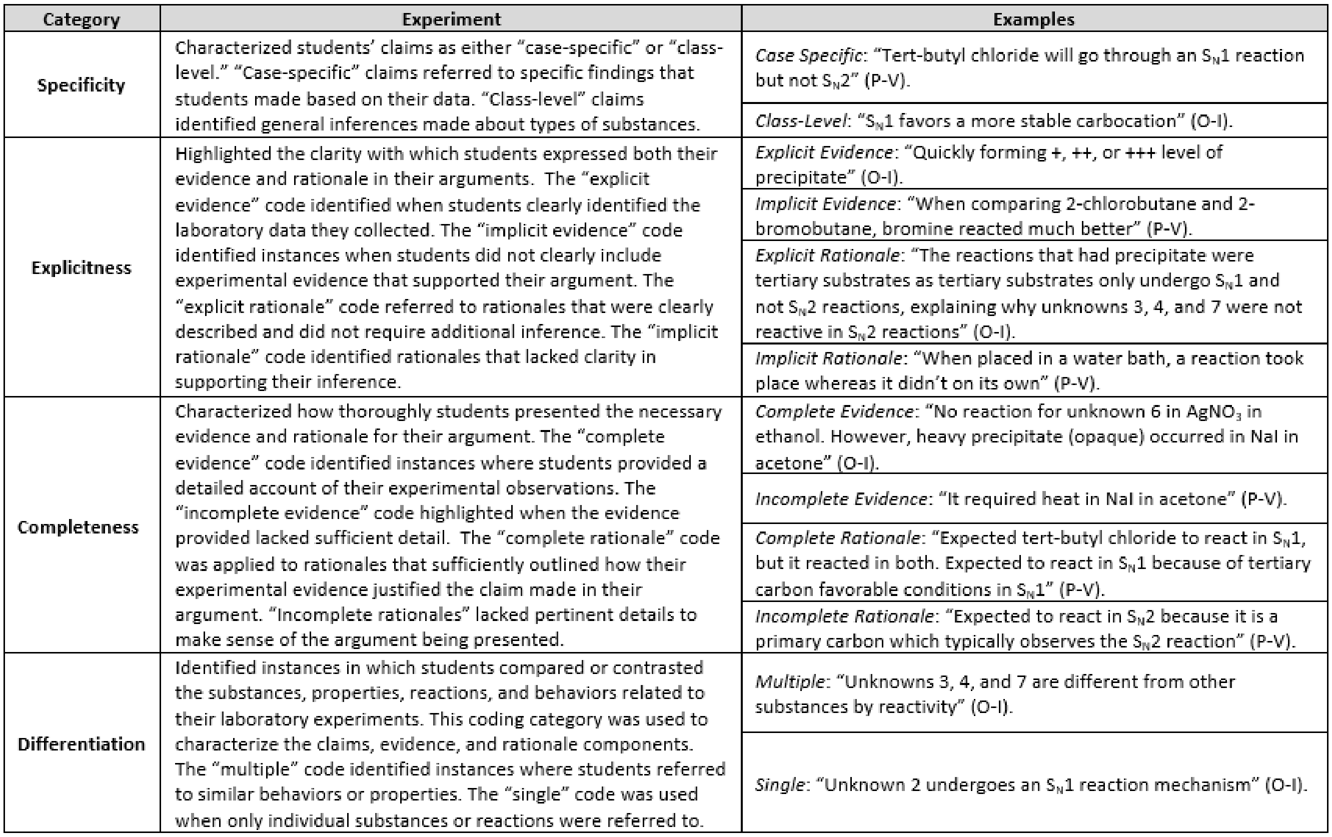
approach to reasoning. These categories helped to characterize the individual components of an argument (*i.e.*, claim, evidence, and rationale) and their connections.

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.

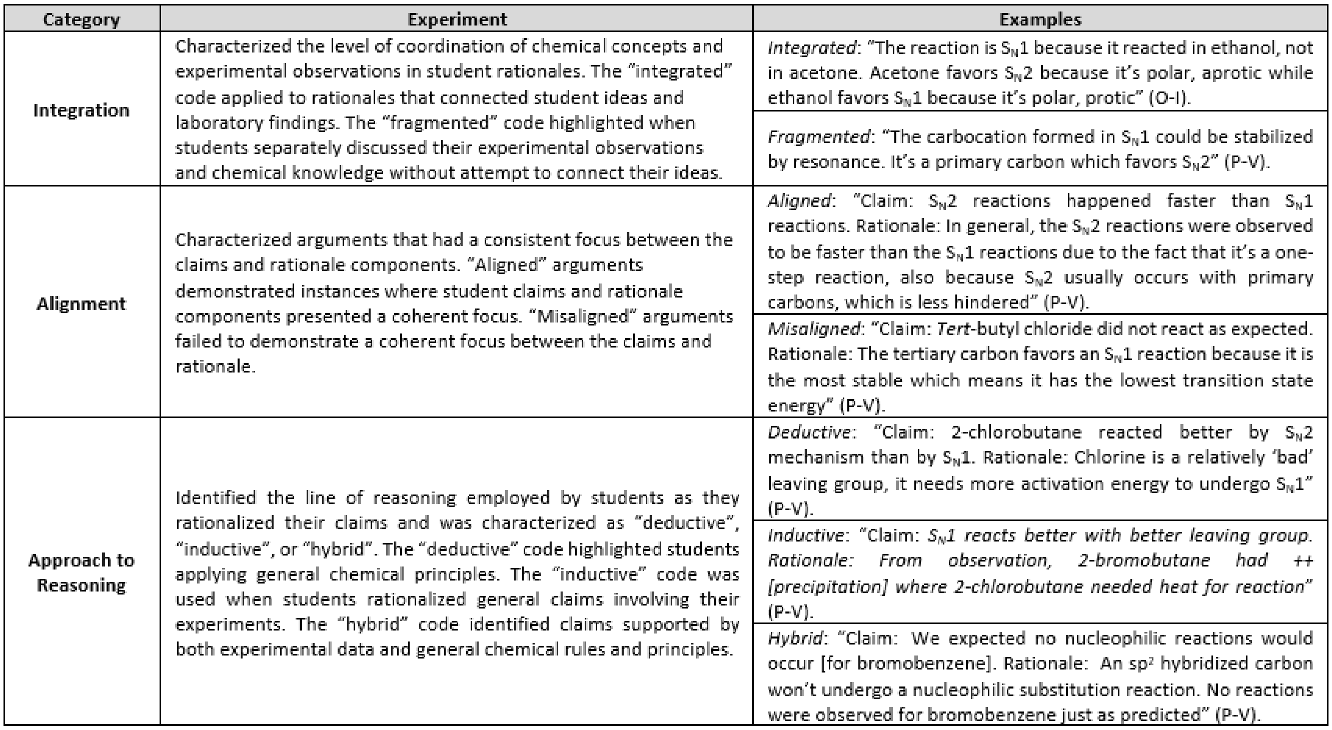
Codes in the "completeness" category characterized the extent to which students presented all the necessary evidence or rationale and included two codes: complete and incomplete. With respect to evidence, the complete code identified instances when students provided detailed

**Table 4** Qualitative scheme utilized for domain-general coding categories and examples of each code from either the predict-verify (P-V) or observe-infer (O-I) laboratory frame.



| Category | Experiment | Examples |
|---|---|---|
| **Specificity** | Characterized students' claims as either "case-specific" or "class-level." "Case-specific" claims referred to specific findings that students made based on their data. "Class-level" claims identified general inferences made about types of substances. | *Case Specific*: "Tert-butyl chloride will go through an $S_N1$ reaction but not $S_N2$" (P-V). |
| | | *Class-Level*: "$S_N1$ favors a more stable carbocation" (O-I). |
| **Explicitness** | Highlighted the clarity with which students expressed both their evidence and rationale in their arguments. The "explicit evidence" code identified when students clearly identified the laboratory data they collected. The "implicit evidence" code identified instances when students did not clearly include experimental evidence that supported their argument. The "explicit rationale" code referred to rationales that were clearly described and did not require additional inference. The "implicit rationale" code identified rationales that lacked clarity in supporting their inference. | *Explicit Evidence*: "Quickly forming +, ++, or +++ level of precipitate" (O-I). |
| | | *Implicit Evidence*: "When comparing 2-chlorobutane and 2-bromobutane, bromine reacted much better" (P-V). |
| | | *Explicit Rationale*: "The reactions that had precipitate were tertiary substrates as tertiary substrates only undergo $S_N1$ and not $S_N2$ reactions, explaining why unknowns 3, 4, and 7 were not reactive in $S_N2$ reactions" (O-I). |
| | | *Implicit Rationale*: "When placed in a water bath, a reaction took place whereas it didn't on its own" (P-V). |
| **Completeness** | Characterized how thoroughly students presented the necessary evidence and rationale for their argument. The "complete evidence" code identified instances where students provided a detailed account of their experimental observations. The "incomplete evidence" code highlighted when the evidence provided lacked sufficient detail. The "complete rationale" code was applied to rationales that sufficiently outlined how their experimental evidence justified the claim made in their argument. "Incomplete rationales" lacked pertinent details to make sense of the argument being presented. | *Complete Evidence*: "No reaction for unknown 6 in $AgNO_3$ in ethanol. However, heavy precipitate (opaque) occurred in NaI in acetone" (O-I). |
| | | *Incomplete Evidence*: "It required heat in NaI in acetone" (P-V). |
| | | *Complete Rationale*: "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 $S_N1$" (P-V). |
| | | *Incomplete Rationale*: "Expected to react in $S_N2$ because it is a primary carbon which typically observes the $S_N2$ reaction" (P-V). |
| **Differentiation** | Identified instances in which students compared or contrasted the substances, properties, reactions, and behaviors related to their laboratory experiments. This coding category was used to characterize the claims, evidence, and rationale components. The "multiple" code identified instances where students referred to similar behaviors or properties. The "single" code was used when only individual substances or reactions were referred to. | *Multiple*: "Unknowns 3, 4, and 7 are different from other substances by reactivity" (O-I). |
| | | *Single*: "Unknown 2 undergoes an $S_N1$ reaction mechanism" (O-I). |

**Table 5** Qualitative scheme utilized for domain-general coding categories and examples of each code from either the predict-verify (P-V) or observe-infer (O-I) laboratory frame.



| Category | Experiment | Examples |
|---|---|---|
| **Integration** | Characterized the level of coordination of chemical concepts and experimental observations in student rationales. The "integrated" code applied to rationales that connected student ideas and laboratory findings. The "fragmented" code highlighted when students separately discussed their experimental observations and chemical knowledge without attempt to connect their ideas. | *Integrated*: "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" (O-I). |
| | | *Fragmented*: "The carbocation formed in $S_N1$ could be stabilized by resonance. It's a primary carbon which favors $S_N2$" (P-V). |
| **Alignment** | Characterized arguments that had a consistent focus between the claims and rationale components. "Aligned" arguments demonstrated instances where student claims and rationale components presented a coherent focus. "Misaligned" arguments failed to demonstrate a coherent focus between the claims and rationale. | *Aligned*: "Claim: $S_N2$ reactions happened faster than $S_N1$ reactions. Rationale: In general, the $S_N2$ reactions were observed to be faster than the $S_N1$ reactions due to the fact that it's a one-step reaction, also because $S_N2$ usually occurs with primary carbons, which is less hindered" (P-V). |
| | | *Misaligned*: "Claim: *Tert*-butyl chloride did not react as expected. Rationale: The tertiary carbon favors an $S_N1$ reaction because it is the most stable which means it has the lowest transition state energy" (P-V). |
| **Approach to Reasoning** | Identified the line of reasoning employed by students as they rationalized their claims and was characterized as "deductive", "inductive", or "hybrid". The "deductive" code highlighted students applying general chemical principles. The "inductive" code was used when students rationalized general claims involving their experiments. The "hybrid" code identified claims supported by both experimental data and general chemical rules and principles. | *Deductive*: "Claim: 2-chlorobutane reacted better by $S_N2$ mechanism than by $S_N1$. Rationale: Chlorine is a relatively 'bad' leaving group, it needs more activation energy to undergo $S_N1$" (P-V). |
| | | *Inductive*: "*Claim: $S_N1$ reacts better with better leaving group. Rationale: From observation, 2-bromobutane had ++ [precipitation] where 2-chlorobutane needed heat for reaction*" (P-V). |
| | | *Hybrid*: "Claim: We expected no nucleophilic reactions would occur [for bromobenzene]. Rationale: An $sp^2$ hybridized carbon won't undergo a nucleophilic substitution reaction. No reactions were observed for bromobenzene just as predicted" (P-V). |

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 $\chi^2$ 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 $\chi^2$ 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 greater than 0.5.

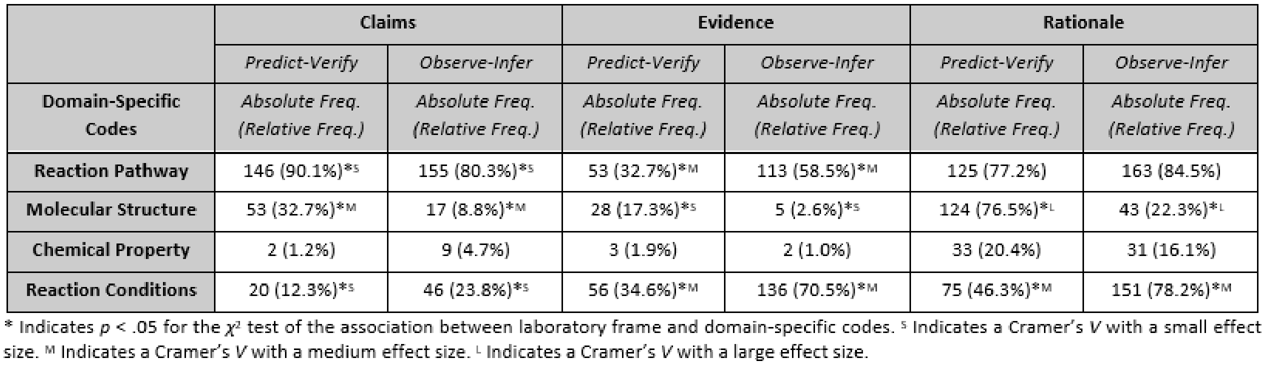
## Results

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

**Table 6** Domain-specific qualitative coding results conducted for the claims, evidence, and rationale components of student post-lab arguments.



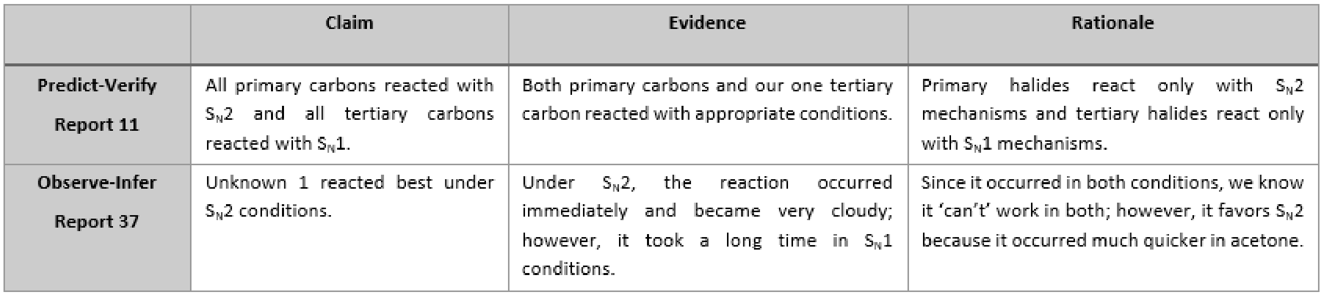
| | **Claims** | | **Evidence** | | **Rationale** | |
|---|---|---|---|---|---|---|
| | *Predict-Verify* | *Observe-Infer* | *Predict-Verify* | *Observe-Infer* | *Predict-Verify* | *Observe-Infer* |
| **Domain-Specific Codes** | *Absolute Freq. (Relative Freq.)* | *Absolute Freq. (Relative Freq.)* | *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 $\chi^2$ 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.

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 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.

**Table 7** 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**<br>**Report 11** | All primary carbons reacted with $S_N2$ and all tertiary carbons reacted with $S_N1$. | Both primary carbons and our one tertiary carbon reacted with appropriate conditions. | Primary halides react only with $S_N2$ mechanisms and tertiary halides react only with $S_N1$ mechanisms. |
| **Observe-Infer**<br>**Report 37** | Unknown 1 reacted best under $S_N2$ conditions. | Under $S_N2$, the reaction occurred immediately and became very cloudy; however, it took a long time in $S_N1$ conditions. | Since it occurred in both conditions, we know it 'can't' work in both; however, it favors $S_N2$ because it occurred much quicker in acetone. |

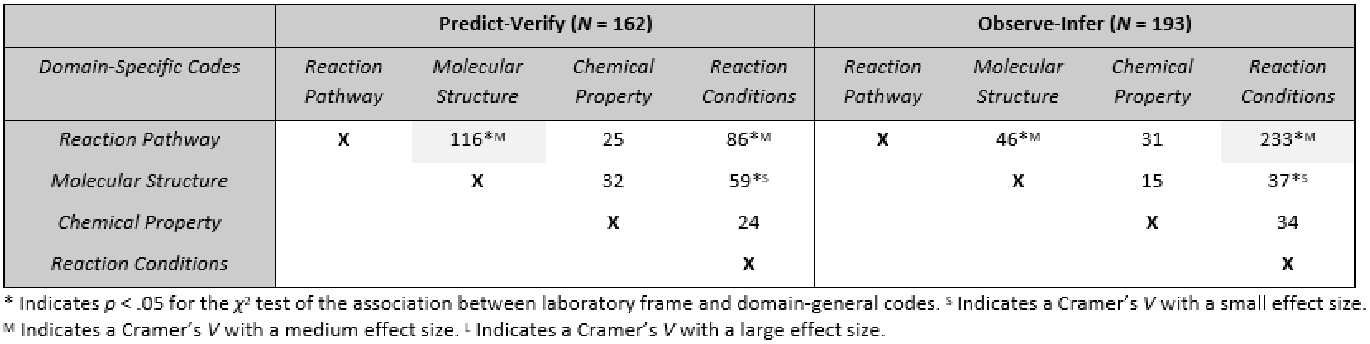
As can be inferred from the frequency data in Table 4, students in the observe-infer frame provided more pieces of evidence in their arguments than students working in the predict-verify frame, and their evidence was more likely related to reaction conditions and pathway than to molecular structure (which 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 reaction 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

**Table 8** Co-occurrence of domain-specific qualitative codes for each laboratory frame in student post-lab arguments.



| | **Predict-Verify ($N$ = 162)** | | | | **Observe-Infer ($N$ = 193)** | | | |
|---|---|---|---|---|---|---|---|---|
| *Domain-Specific Codes* | *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 $\chi^2$ 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.

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. 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.

**Table 9** Domain-general qualitative coding results for the claims, evidence, and rationale components of student post-lab arguments.

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

*Indicates $p < .05$ for the $\chi^2$ 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.

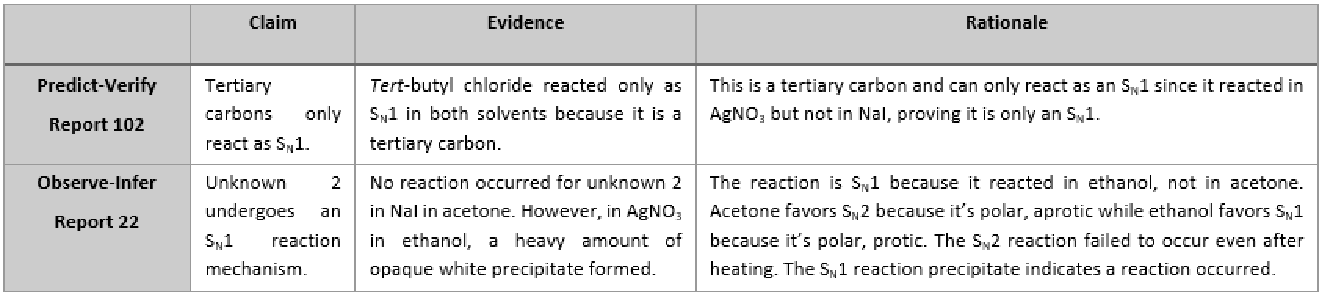
The results summarized in Table 7 indicate that the change in how the laboratory experiment was framed affected the structure of the arguments students constructed, but the effect varied for different categories of analysis. In terms of specificity of claims, students working in the predict-verify frame 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 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 NaI 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

**Table 10** 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 $S_N1$. | *Tert*-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 NaI, 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 NaI 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. |

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 $AgNO_3$ 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 build 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* $AgNO_3$ *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."*

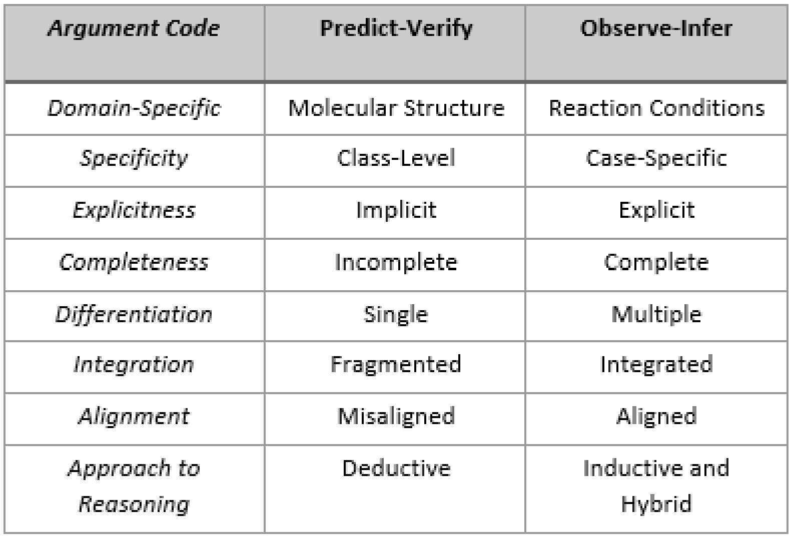
In this argument, the student made a claim about the specific reaction pathway of unknowns 2 and 3 ($S_N1$) when placed in "$AgNO_3$ in ethanol" conditions. Their evidence details the specific observations made for both unknown 2 and 3. In rationalizing this claim, the student first discussed the nature 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 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.

**Table 11** General characterization of laboratory frames by domain-specific and domain-general codes.



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

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 $S_N1$ 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.

The content of this chapter has been reproduced, with permission from the Royal Society of Chemistry, from Petritis, S. J., Kelley, C., and Talanquer, V., (2021), Exploring the impact of the framing of a laboratory experiment on the nature of student argumentation, *Chem. Educ. Res. Pract.*, **22**, 105-121.

# Chapter 4: Exploring Argumentation in the Undergraduate Organic Chemistry Laboratory – Part II: Laboratory Factors

The overarching goal of our research was to identify and characterize various laboratory factors that impact student argumentation in the undergraduate organic chemistry laboratory. 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.

Building on our previous work involving activity framing, our second research project involved the analysis of student arguments produced following the eight experiments that include argumentation during the first semester of our college organic chemistry laboratory course. Arguments were characterized by the same set of domain-general coding categories that were related to the nature of student arguments. Four of these coding categories (explicitness, completeness, integration, and alignment) were also used to characterize student argument quality within our study. Further, we identified four laboratory factors that were hypothesized to impact both the nature and quality of student arguments across the eight experiments in the laboratory curriculum, mainly experiment order, experiment type, the amount and types of data sources, and scaffolding of student argumentation. Although the undergraduate laboratory offers a ripe opportunity for students to engage in argument from evidence, laboratory activity involves a complex web of components each with the potential to affect productive and quality sensemaking. Our findings highlight the importance of explicit consideration of various laboratory factors and their impact on how students express their chemical reasoning through written argumentation.

## Research study

### Research goals and questions

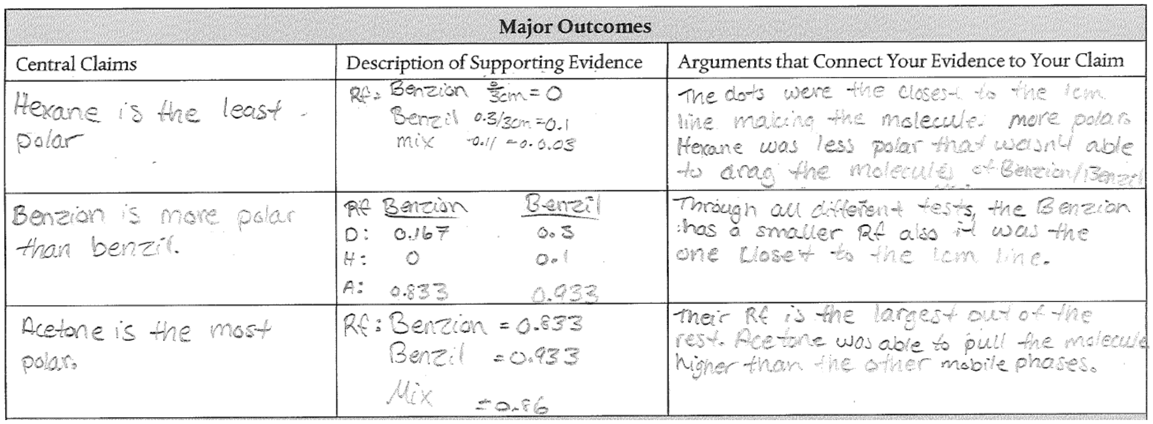
The main goal of this research study was to identify and characterize different factors that could affect the nature and quality of student written argumentation in the undergraduate organic chemistry laboratory. To accomplish this goal, this research project followed eight laboratory experiments in the first semester of a 200-level organic chemistry laboratory in a U.S. university. Specifically, we aimed to characterize various components of student arguments to better understand the factors that impacted the nature and quality of student claims, evidence, and rationales as they made sense of their laboratory findings. We used the following research questions to guide our analysis and presentation of results:

(1) In what ways are the eight experiments similar and different in terms of the nature and quality of student argumentation?

(2) Which factors are most impactful and how do these factors affect the nature and quality of student argumentation across each of the eight experiments of interest?

### Laboratory setting

The research study described herein was conducted during the first semester of a two-semester sequence of organic chemistry laboratory courses at a public research-intensive



| Major Outcomes | | |
|---|---|---|
| Central Claims | Description of Supporting Evidence | Arguments that Connect Your Evidence to Your Claim |
| Hexane is the least polar | Rf: Benzion 0/3cm = 0<br>Benzil 0.3/3cm = 0.1<br>mix 0.1/ = 0.0.03 | The dots were the closest to the 1cm line making the molecule more polar. Hexane was less polar that wasn't able to drag the molecules of Benzion/Benz |
| Benzion is more polar than benzil. | Rf Benzion Benzil<br>D: 0.167 0.3<br>H: 0 0.1<br>A: 0.833 0.933 | Through all different tests, the Benzion has a smaller Rf also it was the one closest to the 1cm line. |
| Acetone is the most polar. | Rf: Benzion = 0.833<br>Benzil = 0.933<br>Mix = 0.86 | Their Rf is the largest out of the rest. Acetone was able to pull the molecule higher than the other mobile phases. |

**Figure 2** Example of a post-lab argument from the thin-layer chromatography (TLC) laboratory experiment. For their arguments, students were asked to compare the relative polarities of the compounds they separated from a mixture.

university in the southwestern US. Students enrolled in the course attended a three-hour laboratory section every week. Each weekly session began with a pre-laboratory lecture and discussion designed to introduce students to the laboratory techniques and core chemical concepts and ideas relevant to each experiment. The next hour to two hours were spent conducting experiments as described by Kelley's (2019) laboratory workbook. Students used the remaining time during each laboratory session to analyze data and construct arguments with a claims-evidence-rationale scaffold for argumentation. Following this framework, students were guided to make claims that "describe what happened", provide evidence in the form of "data to support [their] claim", and a rationale to "connect [their] evidence to [their] claim" (Kelley, 2019). Figure 2 shows an example of a post-lab argument constructed following this claims-evidence-rationale framework. Student reports from the eight experiments were analyzed to characterize similarities and differences in the arguments built by study participants. These eight experiments were categorized into five types based on the goals and tasks presented to students in the laboratory, including a) data collection, analysis, and interpretation; b) isolation and characterization; c) identification and structure elucidation; d) prediction-verification; and e) synthesis and characterization. Basic characteristics of the experiments in each of these categories are summarized in Table 10. Experiments were completed in lab in the same order as presented in this table.

**Table 12** Laboratory experiments performed in the targeted organic chemistry laboratory course and the number of study participants, post-lab reports, and arguments collected and analyzed. Experiments were completed in lab in the order presented in this table.

| Category | Experiment | Description | Student N | Post-lab Report N | Argument N |
|---|---|---|---|---|---|
| **Data collection, analysis, and interpretation experiments (DC)** | Thin-layer Chromatography (TLC) | Students were presented two pure substances, a mixture with these two components, and three laboratory solvents to separate and identify the two substances in the mixture. They collected data involving the movement of their substances on a silica gel TLC plate and analyzed these data by calculating retention factor ($R_f$) values. Interpretation of these data allowed students to investigate the relative polarities of the substances and observe differences in behavior. | 59 | 59 | 177 |
| | Infrared (IR) Spectroscopy | Students were tasked with preparing a sample for analysis using IR spectroscopy. Students collected spectroscopic data in the form of IR spectra and analyzed these data by identifying the wavenumbers of the peaks shown. Interpretation of these data allowed students to identify various bond types and functional groups in molecules of interest. | 68 | 68 | 208 |
| **Extraction and characterization experiment (EC)** | Column Chromatography (CC) | Students were given a hexane extract from ground, raw spinach leaves. This extract contained a mixture of several organic compounds, including the β-carotene compound that students were tasked to isolate using column chromatographic techniques. Students assessed the identity and purity of their isolate using thin-layer chromatography (comparing to a pure standard). | 99 | 99 | 198 |
| **Identification and structure elucidation experiments (ISE)** | Gas Chromatography (GC) | Students performed a transesterification reaction to convert an unknown triglyceride into its component fatty acid methyl esters. They recorded qualitative data about the reactions and collected chromatographic data involving the composition of their product mixture. These data were then used to craft arguments that identified their unknown starting material based on comparisons to known GC standards. | 88 | 59 | 161 |
| | Nuclear Magnetic Resonance (NMR) Spectroscopy | Students were assigned an unknown compound, prepared their own NMR sample, and collected both $^1$H NMR data and an IR spectrum. These data were accompanied by the molecular formula and $^{13}$C NMR data that were provided to each student for their respective unknown compound. Using the provided and collected spectroscopic data, students were tasked with elucidating the structure of their unknown compound and developing an argument that rationalized their structural choice based on their data. | 170 | 170 | 170 |
| **Prediction-verification experiments (PV)** | Substitution Reactions (SR) | Students explored the behavior of eight known alkyl halide starting materials under two sets of reaction conditions: $S_N1$-favorable conditions ($AgNO_3$ in ethanol) and $S_N2$-favorable conditions (NaI in acetone). They used background information about solvent environment (protic vs. aprotic) and molecular structure (methyl, 1°, 2°, or 3° compounds) to predict the reactivity of their eight starting materials under each set of reaction conditions. | 126 | 56 | 162 |
| | Elimination Reactions (ER) | Students performed two separate elimination reactions using known starting materials: 1) the acid-catalyzed dehydration of 2-butanol (E1), and 2) the base-catalyzed dehydrohalogenation of 2-bromobutane (E2). They used background information about solvent environment, reaction mechanisms, and the molecular structure of their expected products to predict the major and minor products in these reactions. | 140 | 70 | 210 |
| **Synthesis and characterization experiment (SC)** | Synthesis of Esters (SE) | Students performed a Fischer esterification reaction in which they refluxed acetic acid, a catalytic amount of sulfuric acid, and an alcohol starting material of their choice to produce a fragrant ester product. Students recorded qualitative data regarding the fragrance of their reaction in addition to collecting both $^1$H NMR and IR spectroscopic data for their observed product. Arguments were then constructed to characterize their synthesized product. | 176 | 69 | 207 |

## Data collection and participants

Most students enrolled in the targeted laboratory course were non-chemistry, science majors (*e.g.*, biology, biochemistry, engineering, physiology) who were concurrently enrolled in the first semester of a two-semester organic chemistry sequence. The course was structured to have up to 24 students divided into individual laboratory sections, each of which was led by a Graduate Student Instructor (GSI). A total of 13 laboratory sections (totaling 11 different GSIs) were randomly selected and consented for participation between the Spring 2019 and Fall 2019 semesters. Recruitment of students in each of these laboratory sections was done verbally at the start of the semester before any laboratory experiments were conducted. In accordance with Institutional Review Board (IRB, protocol #1901297974) policies, all participants consented to participate in the study and data collection, storage, and analysis were conducted following approved IRB guidelines.

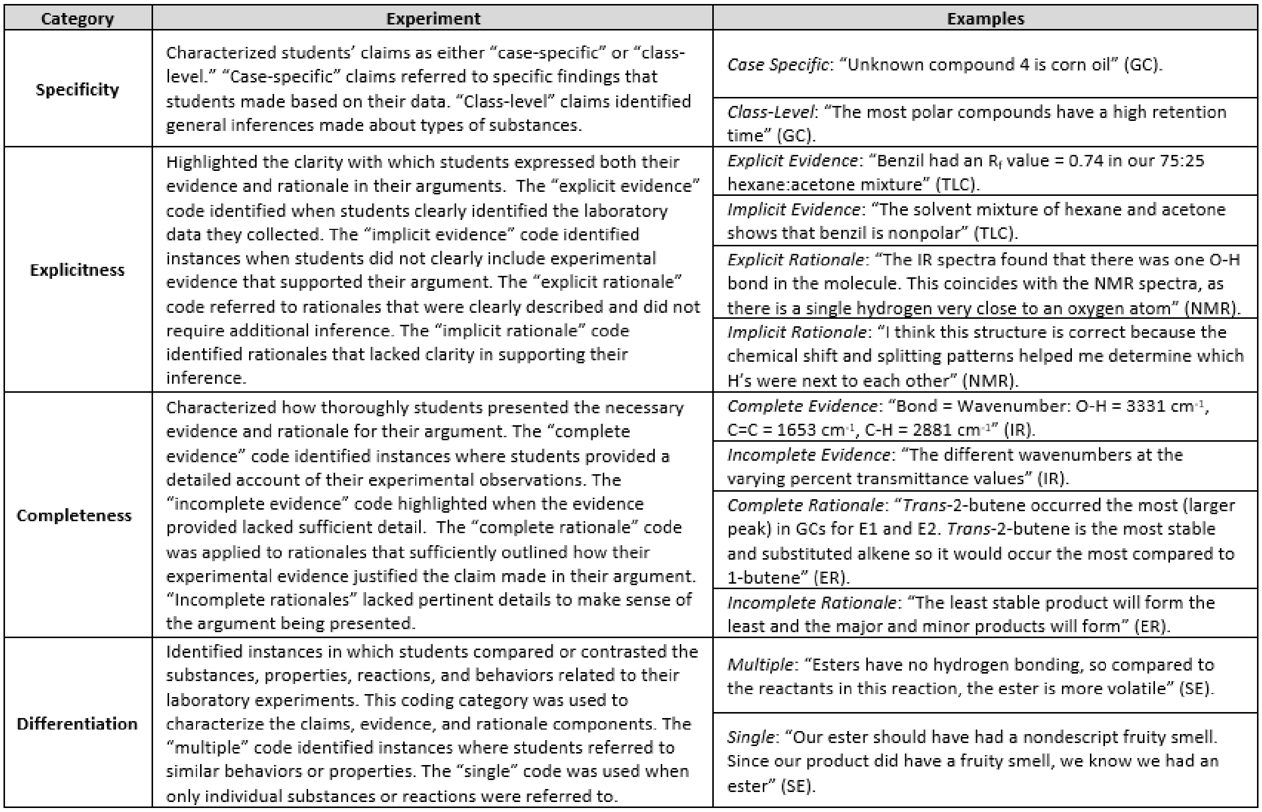
Students completed each laboratory experiment following procedures described in their laboratory workbook and by the instructions of their respective GSI. After completion of the in-lab portion of each experiment, students wrote a post-laboratory report in which they constructed arguments following a claim-evidence-rationale framework. Post-lab arguments were constructed individually, in pairs, or groups of three to five depending on the instructional decisions made by each GSI. Regardless of whether post-lab arguments were written by individual or multiple students, each argument was counted only once during the data collection and analysis processes. Each argument contained an individual claim, evidence, and rationale component as identified by the student(s) and were handwritten into the post-lab argumentation scaffold available to each student in their laboratory workbook (Figure 2). Arguments collected from each of the eight laboratory experiments of interest served as the primary source of data for this research study and analysis of these arguments is presented herein.

Table 10 summarizes student participation, post-lab reports collection, and total arguments analyzed for each laboratory experiment. Post-lab arguments produced by the research participants were collected following each laboratory session. Each post-lab report was de-identified, scanned, and immediately returned to each respective GSI. Post-lab arguments were then transcribed and used for qualitative coding analysis and quantitative characterization of argument quality.

### Qualitative coding analysis

Qualitative data analysis was conducted by adapting a previously established coding scheme (Petritis *et al.*, 2021). This coding scheme focused on domain-general components of

**Table 13** Our previously established qualitative scheme utilized for domain-general coding categories and examples of each code from different experiments.



| Category | Experiment | Examples |
|---|---|---|
| **Specificity** | Characterized students' claims as either "case-specific" or "class-level." "Case-specific" claims referred to specific findings that students made based on their data. "Class-level" claims identified general inferences made about types of substances. | *Case Specific*: "Unknown compound 4 is corn oil" (GC). |
| | | *Class-Level*: "The most polar compounds have a high retention time" (GC). |
| **Explicitness** | Highlighted the clarity with which students expressed both their evidence and rationale in their arguments. The "explicit evidence" code identified when students clearly identified the laboratory data they collected. The "implicit evidence" code identified instances when students did not clearly include experimental evidence that supported their argument. The "explicit rationale" code referred to rationales that were clearly described and did not require additional inference. The "implicit rationale" code identified rationales that lacked clarity in supporting their inference. | *Explicit Evidence*: "Benzil had an $R_f$ value = 0.74 in our 75:25 hexane:acetone mixture" (TLC). |
| | | *Implicit Evidence*: "The solvent mixture of hexane and acetone shows that benzil is nonpolar" (TLC). |
| | | *Explicit Rationale*: "The IR spectra found that there was one O-H bond in the molecule. This coincides with the NMR spectra, as there is a single hydrogen very close to an oxygen atom" (NMR). |
| | | *Implicit Rationale*: "I think this structure is correct because the chemical shift and splitting patterns helped me determine which H's were next to each other" (NMR). |
| **Completeness** | Characterized how thoroughly students presented the necessary evidence and rationale for their argument. The "complete evidence" code identified instances where students provided a detailed account of their experimental observations. The "incomplete evidence" code highlighted when the evidence provided lacked sufficient detail. The "complete rationale" code was applied to rationales that sufficiently outlined how their experimental evidence justified the claim made in their argument. "Incomplete rationales" lacked pertinent details to make sense of the argument being presented. | *Complete Evidence*: "Bond = Wavenumber: O-H = 3331 $cm^{-1}$, C=C = 1653 $cm^{-1}$, C-H = 2881 $cm^{-1}$" (IR). |
| | | *Incomplete Evidence*: "The different wavenumbers at the varying percent transmittance values" (IR). |
| | | *Complete Rationale*: "*Trans*-2-butene occurred the most (larger peak) in GCs for E1 and E2. *Trans*-2-butene is the most stable and substituted alkene so it would occur the most compared to 1-butene" (ER). |
| | | *Incomplete Rationale*: "The least stable product will form the least and the major and minor products will form" (ER). |
| **Differentiation** | Identified instances in which students compared or contrasted the substances, properties, reactions, and behaviors related to their laboratory experiments. This coding category was used to characterize the claims, evidence, and rationale components. The "multiple" code identified instances where students referred to similar behaviors or properties. The "single" code was used when only individual substances or reactions were referred to. | *Multiple*: "Esters have no hydrogen bonding, so compared to the reactants in this reaction, the ester is more volatile" (SE). |
| | | *Single*: "Our ester should have had a nondescript fruity smell. Since our product did have a fruity smell, we know we had an ester" (SE). |

arguments that illustrated the focus and connectivity of student claims, evidence, and rationale components (see Table 11 and Table 12). The same previously reported domain-general coding categories were used to characterize the arguments produced by students from each of the eight laboratory experiments analyzed in this study. They included specificity, explicitness, completeness, differentiation, integration, alignment, and approach to reasoning. These domain-general codes served as our primary method for comparing and contrasting the arguments crafted during each experiment.

Qualitative analysis began with the first author (a graduate student) randomly selecting post-lab reports for the eight laboratory experiments of interest. A qualitative codebook was developed for each of these experiments which included all domain-general coding categories. Coding categories and codes were first identified and applied by the graduate student researcher and discussed with the second author. The two researchers independently coded each argument

**Table 14** Our previously established qualitative scheme utilized for domain-general coding categories and examples of each code from different experiments.

| Category | Experiment | Examples |
|---|---|---|
| **Integration** | Characterized the level of coordination of chemical concepts and experimental observations in student rationales. The "integrated" code applied to rationales that connected student ideas and laboratory findings. The "fragmented" code highlighted when students separately discussed their experimental observations and chemical knowledge without attempt to connect their ideas. | *Integrated*: "As shown on the TLC plate, there is only one spot for the isolated carotenes that matches the $\beta$-carotene standard. Also, the $R_f$ value for each isolated carotene matches the $R_f$ value for the $\beta$-carotene standard." (CC). |
| | | *Fragmented*: "Because the $R_f$ value for our isolate was the same $R_f$ value for the $\beta$-carotene structure" (CC). |
| **Alignment** | Characterized arguments that had a consistent focus between the claims and rationale components. "Aligned" arguments demonstrated instances where student claims and rationale components presented a coherent focus. "Misaligned" arguments failed to demonstrate a coherent focus between the claims and rationale. | *Aligned*: "Claim: $S_N2$ reactions happened faster than $S_N1$ reactions. Rationale: In general, the $S_N2$ reactions were observed to be faster than the $S_N1$ reactions due to the fact that it's a one-step reaction, also because $S_N2$ usually occurs with primary carbons, which is less hindered" (SR). |
| | | *Misaligned*: "Claim: *Tert*-butyl chloride did not react as expected. Rationale: The tertiary carbon favors an $S_N1$ reaction because it is the most stable which means it has the lowest transition state energy" (SR). |
| **Approach to Reasoning** | Identified the line of reasoning employed by students as they rationalized their claims and was characterized as "deductive", "inductive", or "hybrid". The "deductive" code highlighted students applying general chemical principles. The "inductive" code was used when students rationalized general claims involving their experiments. The "hybrid" code identified claims supported by both experimental data and general chemical rules and principles. | *Deductive*: "When using TLC, molecules travel further in solvents with similar polarities. It is known that methanol is polar, so benzoin is polar" (TLC). |
| | | *Inductive*: "The $R_f$ value for $\beta$-carotene was very close to one, meaning that the vitamin moved closely along with the nonpolar mobile phase, which means the $\beta$-carotene is also nonpolar" (TLC). |
| | | *Hybrid*: "Nonpolar components would result in it traveling further up the plate, which it did with an $R_f$ value of 0.92….The pure $\beta$-carotene has C=C and C-H bonds, which makes it very hydrophobic" (TLC). |

component from the selected reports for a given laboratory experiment and subsequently met to discuss their respective coding decisions. Discussions were had until both researchers agreed on their coding choices for each analyzed argument. This iterative process was followed until consensus was reached on at least 25% of the collected arguments for each of the eight laboratory experiments of interest. Remaining arguments were qualitatively coded by the graduate student researcher.

### Quantitative analysis of argument quality

Following our qualitative coding analysis, we sought to characterize the quality of student arguments for each of the eight laboratory experiments. We identified four of our seven domain-general coding categories as indicative of the overall quality of an argument, specifically the explicitness, completeness, integration, and alignment coding categories. For each of these four coding categories, one of the two possible codes was assigned a higher-quality argument score. For example, in the integration coding category for the rationale component, the code "integrated" was assigned a higher numerical value than "fragmented". Similarly, the explicitness, completeness, and alignment coding categories all contain a more highly valued code. Arguments coded as "explicit" (evidence and rationale), "complete" (evidence and rationale), "integrated" (rationale), and "aligned" (rationale) were judged to lead to a higher quality argument.

To quantitatively characterize the quality of arguments from each experiment, we assigned a value of "1" to argument components characterized as "explicit", "complete", "integrated", and "aligned", and assigned a "0" to argument components coded as "implicit", "incomplete", "fragmented", and "misaligned". The explicitness and completeness coding categories were both counted twice in this analysis as these categories were applied to both the evidence and rationale components in our qualitative coding analysis. Thus, in total each argument got scores for each of the six components and an added quality score in the range 0 to 6. Additional domain-general codes not deemed indicative of argument quality (*i.e.*, specificity, differentiation, and approach to reasoning) were given an arbitrary value. For example, the approach to reasoning coding category had three possible codes, “deductive”, “inductive”, and “hybrid”, which were assigned a “2”, “1”, and “0”, respectively. A total of 1,493 arguments (each consisting of one claim, one evidence, and one rationale component) were analyzed and categorized in this manner.

The R statistical software (Windows, version 4.0.3) was used to run all statistical analysis for this research study. The frequency of occurrence of each qualitative code were calculated for each of the eight laboratory experiments in our study. We used the Chi-square ($\chi2$) test for independence to investigate the association between our qualitative codes and each laboratory experiment as well as to compare overall argument quality across our eight experiments. Statistically significant associations between each code and laboratory experiment were investigated at the $\alpha = .05$ level. In addition, we used the R statistical software to calculate the standardized Chi-square residual values for the association between each combination of the categorical variables (qualitative codes and laboratory experiment).

## Results

In the following subsections, we present results on: (1) the characterization of similarities and differences of written arguments for the eight experiments we explored in this research study, and (2) the analysis of several factors and their impact on the nature and quality of laboratory arguments.

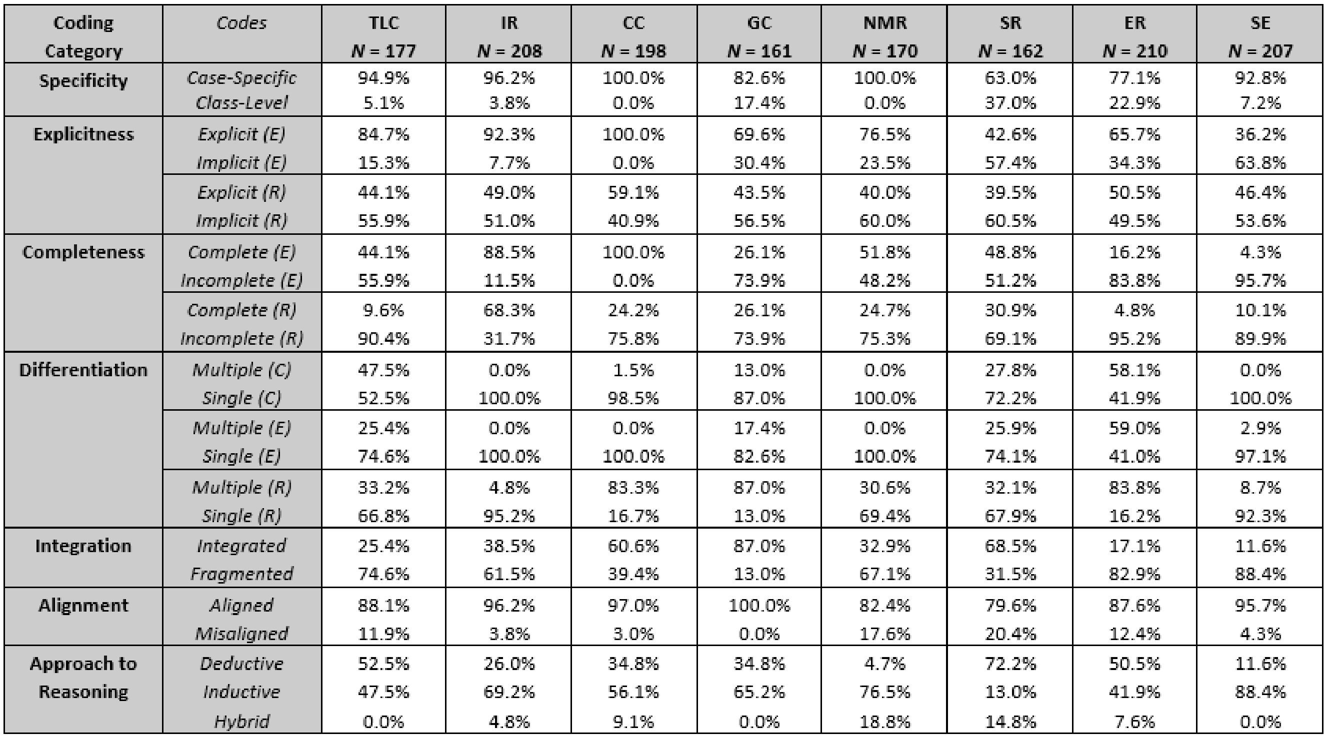
### In what ways are the eight experiments similar and different in terms of the nature and quality of student argumentation?

Students' arguments in the eight laboratory experiments varied in structure depending on the coding category of interest. Our qualitative analysis of the nature of student argumentation focused on the domain-general categories described in Table 11. Table 12 displays the relative frequency of each code broken down by laboratory experiment. Table 13 summarizes the Chi-square residual values for each coding category across the eight experiments.

### *Specificity*

The majority of student claims across all eight experiments were characterized as case-specific, including the column chromatography and NMR experiments that contained solely case-specific claims. For example, the column chromatography experiment example in Table 14 has a case-specific claim referring to the "greatly pure" isolated "carotenes" with which the students worked. Class-level claims were sparse for most experiments with the GC (17.4%), elimination

**Table 15** Relative frequency of domain-general codes for the claims, evidence, and rationale components of students arguments for each experiment.



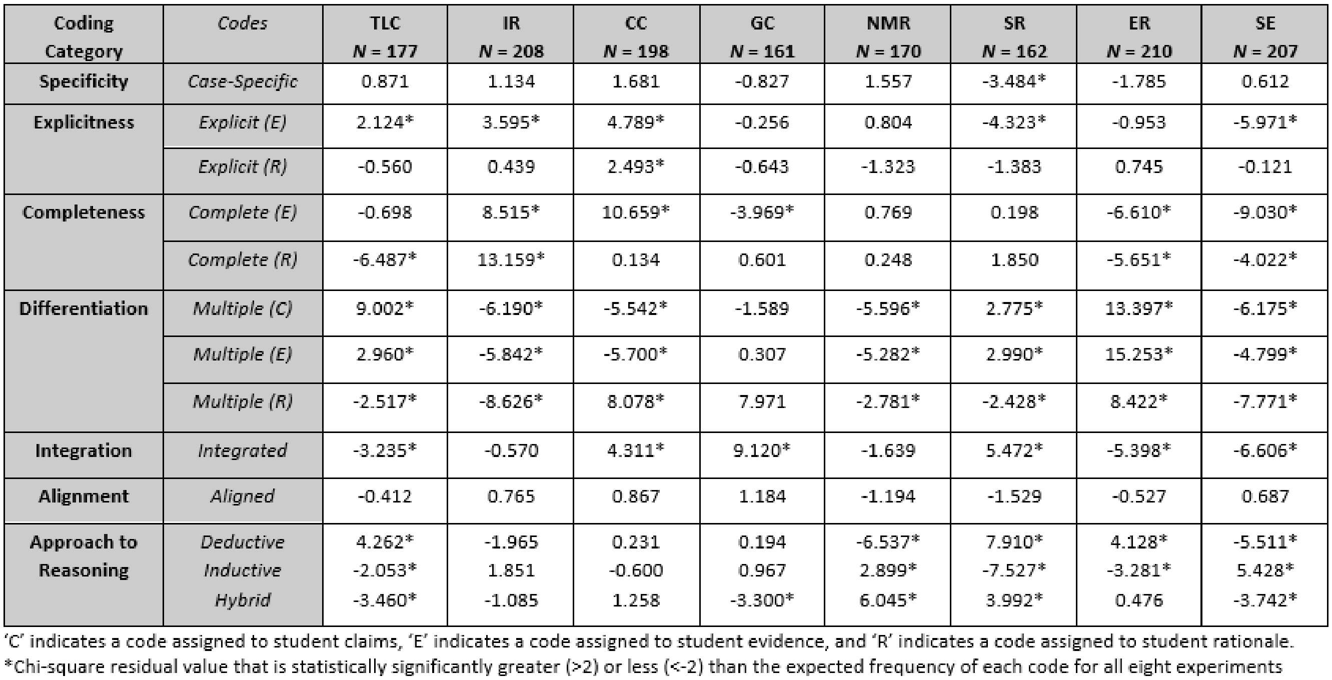
| Coding Category | *Codes* | TLC<br>*N* = 177 | IR<br>*N* = 208 | CC<br>*N* = 198 | GC<br>*N* = 161 | NMR<br>*N* = 170 | SR<br>*N* = 162 | ER<br>*N* = 210 | SE<br>*N* = 207 |
|---|---|---|---|---|---|---|---|---|---|
| **Specificity** | *Case-Specific* | 94.9% | 96.2% | 100.0% | 82.6% | 100.0% | 63.0% | 77.1% | 92.8% |
| | *Class-Level* | 5.1% | 3.8% | 0.0% | 17.4% | 0.0% | 37.0% | 22.9% | 7.2% |
| **Explicitness** | *Explicit (E)* | 84.7% | 92.3% | 100.0% | 69.6% | 76.5% | 42.6% | 65.7% | 36.2% |
| | *Implicit (E)* | 15.3% | 7.7% | 0.0% | 30.4% | 23.5% | 57.4% | 34.3% | 63.8% |
| | *Explicit (R)* | 44.1% | 49.0% | 59.1% | 43.5% | 40.0% | 39.5% | 50.5% | 46.4% |
| | *Implicit (R)* | 55.9% | 51.0% | 40.9% | 56.5% | 60.0% | 60.5% | 49.5% | 53.6% |
| **Completeness** | *Complete (E)* | 44.1% | 88.5% | 100.0% | 26.1% | 51.8% | 48.8% | 16.2% | 4.3% |
| | *Incomplete (E)* | 55.9% | 11.5% | 0.0% | 73.9% | 48.2% | 51.2% | 83.8% | 95.7% |
| | *Complete (R)* | 9.6% | 68.3% | 24.2% | 26.1% | 24.7% | 30.9% | 4.8% | 10.1% |
| | *Incomplete (R)* | 90.4% | 31.7% | 75.8% | 73.9% | 75.3% | 69.1% | 95.2% | 89.9% |
| **Differentiation** | *Multiple (C)* | 47.5% | 0.0% | 1.5% | 13.0% | 0.0% | 27.8% | 58.1% | 0.0% |
| | *Single (C)* | 52.5% | 100.0% | 98.5% | 87.0% | 100.0% | 72.2% | 41.9% | 100.0% |
| | *Multiple (E)* | 25.4% | 0.0% | 0.0% | 17.4% | 0.0% | 25.9% | 59.0% | 2.9% |
| | *Single (E)* | 74.6% | 100.0% | 100.0% | 82.6% | 100.0% | 74.1% | 41.0% | 97.1% |
| | *Multiple (R)* | 33.2% | 4.8% | 83.3% | 87.0% | 30.6% | 32.1% | 83.8% | 8.7% |
| | *Single (R)* | 66.8% | 95.2% | 16.7% | 13.0% | 69.4% | 67.9% | 16.2% | 92.3% |
| **Integration** | *Integrated* | 25.4% | 38.5% | 60.6% | 87.0% | 32.9% | 68.5% | 17.1% | 11.6% |
| | *Fragmented* | 74.6% | 61.5% | 39.4% | 13.0% | 67.1% | 31.5% | 82.9% | 88.4% |
| **Alignment** | *Aligned* | 88.1% | 96.2% | 97.0% | 100.0% | 82.4% | 79.6% | 87.6% | 95.7% |
| | *Misaligned* | 11.9% | 3.8% | 3.0% | 0.0% | 17.6% | 20.4% | 12.4% | 4.3% |
| **Approach to Reasoning** | *Deductive* | 52.5% | 26.0% | 34.8% | 34.8% | 4.7% | 72.2% | 50.5% | 11.6% |
| | *Inductive* | 47.5% | 69.2% | 56.1% | 65.2% | 76.5% | 13.0% | 41.9% | 88.4% |
| | *Hybrid* | 0.0% | 4.8% | 9.1% | 0.0% | 18.8% | 14.8% | 7.6% | 0.0% |

'C' indicates a code assigned to student claims, 'E' indicates a code assigned to student evidence, and 'R' indicates a code assigned to student rationale.

reaction (22.9%), and substitution reaction (37.0%) experiments exhibiting the greatest frequency of such claims. Class-level claims, as exemplified by the substitution reaction experiment claim in Table 14, were inferences about general classes of molecules, reactions, or substances ("Carbons not sp3 hybridized"). The two prediction-verification experiments were characterized by the most class-level claims, the majority of which were made about reaction pathway for both the substitution reaction (90.1%) and elimination reaction (61.9%) experiments. However, Chi-square analysis, as summarized in Table 13, demonstrated that only the substitution reaction experiment was characterized as having fewer case-specific claims (more class-level claims) compared to what was expected of the claims for all eight experiments. Thus, the experiments did not vary greatly with regards to the specificity of the claims students identified in their arguments.

### *Explicitness*

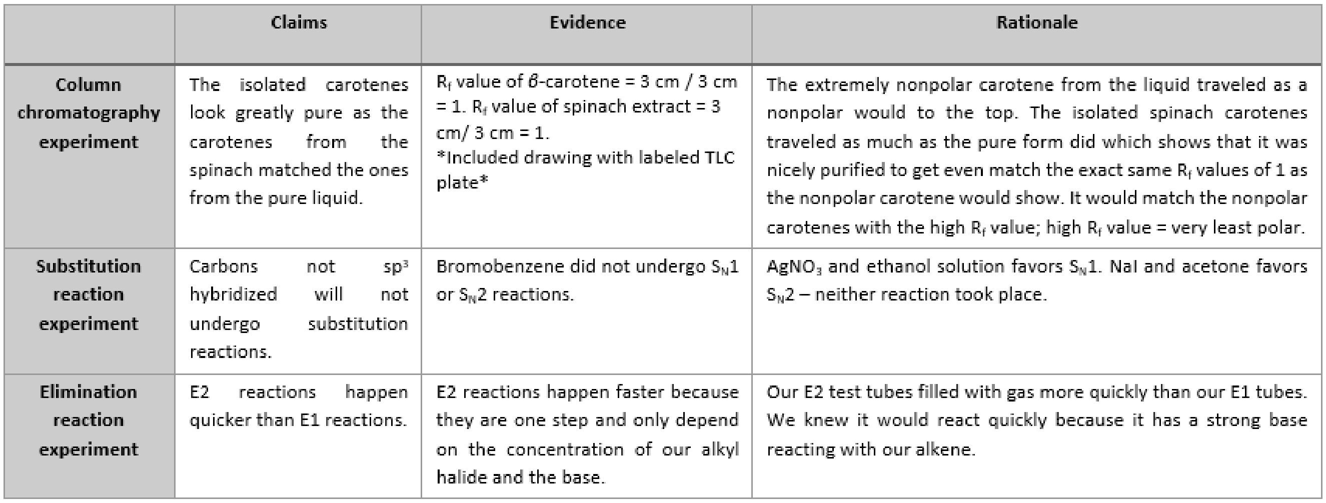
**Table 16** Chi-square residuals of domain-general codes for the claims, evidence, and rationale components of students arguments for each experiment.



| Coding Category | *Codes* | TLC *N* = 177 | IR *N* = 208 | CC *N* = 198 | GC *N* = 161 | NMR *N* = 170 | SR *N* = 162 | ER *N* = 210 | SE *N* = 207 |
|---|---|---|---|---|---|---|---|---|---|
| **Specificity** | *Case-Specific* | 0.871 | 1.134 | 1.681 | -0.827 | 1.557 | -3.484* | -1.785 | 0.612 |
| **Explicitness** | *Explicit (E)* | 2.124* | 3.595* | 4.789* | -0.256 | 0.804 | -4.323* | -0.953 | -5.971* |
| | *Explicit (R)* | -0.560 | 0.439 | 2.493* | -0.643 | -1.323 | -1.383 | 0.745 | -0.121 |
| **Completeness** | *Complete (E)* | -0.698 | 8.515* | 10.659* | -3.969* | 0.769 | 0.198 | -6.610* | -9.030* |
| | *Complete (R)* | -6.487* | 13.159* | 0.134 | 0.601 | 0.248 | 1.850 | -5.651* | -4.022* |
| **Differentiation** | *Multiple (C)* | 9.002* | -6.190* | -5.542* | -1.589 | -5.596* | 2.775* | 13.397* | -6.175* |
| | *Multiple (E)* | 2.960* | -5.842* | -5.700* | 0.307 | -5.282* | 2.990* | 15.253* | -4.799* |
| | *Multiple (R)* | -2.517* | -8.626* | 8.078* | 7.971 | -2.781* | -2.428* | 8.422* | -7.771* |
| **Integration** | *Integrated* | -3.235* | -0.570 | 4.311* | 9.120* | -1.639 | 5.472* | -5.398* | -6.606* |
| **Alignment** | *Aligned* | -0.412 | 0.765 | 0.867 | 1.184 | -1.194 | -1.529 | -0.527 | 0.687 |
| **Approach to Reasoning** | *Deductive*<br>*Inductive*<br>*Hybrid* | 4.262*<br>-2.053*<br>-3.460* | -1.965<br>1.851<br>-1.085 | 0.231<br>-0.600<br>1.258 | 0.194<br>0.967<br>-3.300* | -6.537*<br>2.899*<br>6.045* | 7.910*<br>-7.527*<br>3.992* | 4.128*<br>-3.281*<br>0.476 | -5.511*<br>5.428*<br>-3.742* |

'C' indicates a code assigned to student claims, 'E' indicates a code assigned to student evidence, and 'R' indicates a code assigned to student rationale.
*Chi-square residual value that is statistically significantly greater (>2) or less (<-2) than the expected frequency of each code for all eight experiments

**Table 17** Examples of three student written arguments: one from the column chromatography experiment, one from the substitution reaction experiment, and one from the elimination reaction experiment.



| | **Claims** | **Evidence** | **Rationale** |
|---|---|---|---|
| **Column chromatography experiment** | The isolated carotenes look greatly pure as the carotenes from the spinach matched the ones from the pure liquid. | $R_f$ value of *β*-carotene = 3 cm / 3 cm = 1. $R_f$ value of spinach extract = 3 cm/ 3 cm = 1.<br>*Included drawing with labeled TLC plate* | The extremely nonpolar carotene from the liquid traveled as a nonpolar would to the top. The isolated spinach carotenes traveled as much as the pure form did which shows that it was nicely purified to get even match the exact same $R_f$ values of 1 as the nonpolar carotene would show. It would match the nonpolar carotenes with the high $R_f$ value; high $R_f$ value = very least polar. |
| **Substitution reaction experiment** | Carbons not $sp^3$ hybridized will not undergo substitution reactions. | Bromobenzene did not undergo $S_N1$ or $S_N2$ reactions. | $AgNO_3$ and ethanol solution favors $S_N1$. NaI and acetone favors $S_N2$ – neither reaction took place. |
| **Elimination reaction experiment** | E2 reactions happen quicker than E1 reactions. | E2 reactions happen faster because they are one step and only depend on the concentration of our alkyl halide and the base. | Our E2 test tubes filled with gas more quickly than our E1 tubes. We knew it would react quickly because it has a strong base reacting with our alkene. |

Six of the eight experiments were characterized by a majority of arguments coded as having an explicit evidence component. Consider the evidence component of the column chromatography experiment argument listed in Table 14. The student clearly identified their TLC plate measurements, presented their calculation of the corresponding $R_f$ values for both the *β*-carotene standard and the spinach extract, drew a labeled TLC plate as a visual guide, leading to the assigned explicit evidence code. The TLC (84.7%), IR (92.3%), and column chromatography (100.0%) experiments were most often characterized by arguments with explicit evidence components, each of which were more frequently coded as explicit compared to what was expected across all eight laboratory experiments. Conversely, the explicit evidence code appeared least often for the substitution reaction (42.6%) and synthesis of esters (36.2%) experiments, both falling below the expected frequency of the explicit evidence code. The arguments from these experiments more often included implicit evidence, as highlighted by evidence provided for the substitution reaction experiment example in Table 14. In case, the student identified that "Bromobenzene did not undergo $S_N1$ or $S_N2$ reactions" without clearly detailing specific observations made while running these separate sets of reaction. Because one has to imply that no precipitation, color change, or cloudiness was observed throughout either of the $S_N1$ and $S_N2$ reactions described, this evidence component was coded as implicit for this student's argument.

Chemical rationales across each experiment did not vary greatly with regards to how explicitly these ideas were supported in student arguments. Arguments across most experiments were either about evenly split between the frequency of the explicit and implicit rationale codes (IR and elimination reaction experiments) or showed a slight preference for an implicit rationale (TLC, GC, NMR, and synthesis of esters experiments). Consider the following argument from the GC experiment:

**Claim:** *"Different length fatty acid carbon chains exhibit different retention times."*

**Evidence:** *"Carbon chains with 12, 14, 16, and 18 carbons show different peaks on GC spectra."*

**Rationale:** *"We had peaks showing carbon chains of 12, 14, 16, and 18 carbons in our unknown fat."*

In this argument, the student supported their class-level claim about fatty acid carbon chain length and retention time with a rationale that did not clearly connect the chain lengths observed in their "unknown fat" to the "different peaks on the GC spectra" they collected. Thus, this chemical rationale implicitly rationalized the presence of "different length fatty acid carbon chains" without using their data to complete the inference. On the other hand, only the column chromatography (59.1%) experiment was characterized by chemical rationales that were more often coded as explicit than expected across all experiments. Rationales produced during the column chromatography experiment were often clearly supported by coordinating ideas related to observed reaction conditions (72.7%), chemical properties (75.8%) and molecular structures (37.9%) of the molecules with which they worked, and the chromatographic data (87.9%) collected from their experiment. As exemplified by the rationale in Table 14, the student clearly rationalized the TLC plate behavior of each *β*-carotene sample, as indicated by $R_f$ values, and connected these ideas to the expected polarity of the compound.

***Completeness***

The degree of completeness of student evidence varied greatly depending on which experiment students performed. The IR (88.5%) and column chromatography (100.0%) experiments displayed the greatest frequency of complete evidence components, in which students identified all the necessary supporting evidence to justify their proposed claim. Again, consider the evidence component of the column chromatography experiment argument shown in Table 14. This student identified all the relevant pieces of empirical data associated with their experiment and provided an illustration of their TLC plate as support. Similarly, reflect on this argument from the IR experiment:

> **Claim:** *"Limonene has alkane and alkene functional groups."*
>
> **Evidence:** *"The wavenumber values from the IR spectroscopy: C-H was 3010 cm-1 and C=C was 1644 cm-1."*
>
> **Rationale:** *"The known values for these two bonds are 3100-2900 cm-1, which matches our result of 3010 cm-1. The 1650 cm-1 for C=C compares to 1644 cm-1 experimental value."*

In this argument, the student provided evidence regarding their observed IR peaks, making specific reference to the wavenumbers collected in their spectra and the bond types identified by each peak related to their compound of interest. Conversely, the GC (73.9%), elimination reaction (83.8%), and synthesis of esters (95.7%) experiments shown a major preference for incomplete evidence components. Arguments coded with incomplete evidence for these experiments often included individual pieces of data in their argument and, thus, excluded key pieces of information need to support the proposed chemical inference. For example, consider the elimination reaction experiment argument shown below.

**Claim:** *"E1 reaction produced a greater percentage of 2-butene for its products."*

**Evidence:** *"E1 reaction was completed with 0.4 mL of 2-butanol and 0.6 mL of an acid mixture."*

**Rationale:** *"E1 2-butene total area was greater than 1-butene. This was not true for E2 reaction."*

In this argument, the student made a claim comparing the amount of products produced in E1 reactions implicitly to E2 reactions. The evidence referenced the reaction conditions of the E1 reaction, but failed to provide experimental data that could support the claim regarding the "percentage of 2-butene." Although arguments from the elimination reaction experiment often relied on chromatographic data (85.7% of arguments) as evidence, these arguments unsuccessfully detailed all the necessary evidence needed to support the proposed claim.

Arguments across each experiment (except the IR experiment) were most frequently characterized as having incomplete rationale components as well. Incomplete rationales excluded key features of either laboratory data or conceptual knowledge needed to support the claims students made. Again, consider the elimination reaction example shown in the previous paragraph. This student rationalized that the "E1 2-butene total area was greater than 1-butene" without making specific reference to the area of the peaks that would support this inference. Similarly, they stated that "This was not true for E2 reactions" without detailing the basis for the rationale they employed in that argument. Incomplete rationales were most common in the synthesis of esters (89.9%), TLC (90.4%), and elimination reaction (95.2%) experiments while the IR experiment remained the only experiment in which students more frequently provided complete rationales (68.3%) in their arguments. Looking at the IR experiment argument shown above, the student rationalized the claim “Limonene has alkane and alkene functional groups” by saying “The known values for these two bonds are 3100-2900 cm-1, which matches our result of 3010 cm-1. The 1650 cm-1 for C=C compares to 1644 cm-1 experimental value.” In this rationale, the student provided a rationale that included the wavenumbers of the data collected and the theoretical wavenumbers for each peak observed in their molecule.

### *Differentiation*

Students compared and contrasted the substances and reactions observed in each of the eight experiments. Unlike the specificity coding category, there was greater variability observed for differentiation across the claims, evidence, and rationale components of their arguments. Several experiments included arguments which were characterized by very few or no differentiations across all argument components. For instance, consider the following example from the IR experiment:

**Claim:** *“Cis-3-hexen-1-ol has an O-H functional group.”*

**Evidence:** *“The IR spectroscopy results for cis-3-hexen-1-ol show a wavenumber 3335.20 cm-1 with a curved dip.”*

**Rationale:** *“On IR spectra, O-H bond is shown as a curved dip with a broad wavenumber range between 3600-2900 cm-1 (theoretically).”*

In this argument, the student made a claim about the presence of the O-H functional group in their *cis*-3-hexen-1-ol substance, reported the wavenumber (in cm-1) of their observed peak as evidence, and conferred this bond type with what was theoretically expected for the O-H bond for their rationale. The focus of this argument was a single substance, a single observed peak, and a single bond type and, thus, each component of the argument was coded as “single” for the differentiation coding category. Similarly, arguments from the IR, NMR, and synthesis of esters experiments all demonstrated lower than expected frequency of the “multiple” code of differentiation for student claims, evidence, and rationale components (Table 13). Conversely, several experiments were relatively comparison-rich. For instance, the majority of student claims, evidence, and rationale components of the elimination reaction experiment were characterized as “multiple” differentiation, including the example shown in Table 14. This student supported their claim about the difference in reaction rate between the E1 and E2 reaction pathways with evidence and rationale that further differentiated their observations and theoretical knowledge concerning the two reaction mechanisms. Additionally, both the column chromatography (83.3%) and GC (87.0%) experiments included rationale components containing predominately “multiple” differentiation, including the column chromatography example in Table 14, which compared “the same $R_f$ values of 1” for the isolate and pure $\beta$-carotenes observed by that student.

### *Integration*

Laboratory experiments also displayed great variability in the level of integration of core chemical concepts and ideas with the data/observations students detailed in their rationales. Three experiments were characterized as having more highly integrated rationales, including the column chromatography (60.6%), substitution reaction (68.5%), and GC (87.0%) experiments. Arguments that were characterized by integrated rationales can be seen in the example below from the GC experiment.

**Claim:** *“The unknown sample was linseed oil.”*

**Evidence:** *“The 16 carbon peak: 7.6% (GC). The sp3 C-H bond: 3008.31 cm-1 (IR). The C=C and C=O bonds are absent from the IR spectrum.”*

**Rationale:** *"The unknown was determined as linseed oil based on the peak for 16-*carbon *showing ~6% (7.6%) as expected for linseed oil (palmitic acid) and the rest being composed of other carbons. The data for IR are also coherent with the structure of the fatty acid (only sp3 C-H bonds)."*

In this argument, the student made a claim about the identity of their unknown fat starting material in the GC experiment and supported this claim with the presence and absence of IR and GC peaks. They rationalized their claim by describing how both sets of data match was "expected for linseed oil (palmitic acid)" based on the materials provided to them about each possible unknown compound. The student coordinated the data they collected with what is known about each of their possible compounds and, thus, their rationale was coded as integrated. In contrast, the majority of laboratory experiments contained arguments whose rationales were characterized as fragmented, in which students failed to adequately coordinate data and observations with background knowledge in support of their proposed claims. Fragmented rationales were most common in the synthesis of esters (88.4%), elimination reaction (82.4%), and thin-layer chromatography (74.6%) experiments. Consider the example shown below from the synthesis of esters experiment.

**Claim:** *"The combination of 1-hexanol (an alcohol) with acetic acid and sulfuric acid results in an ester."*

**Evidence:** *"After the reflux reaction, we wafted the scent of our product."*

**Rationale:** *"What we smelled was characteristic of what we would expect to smell in the ester, propyl acetate."*

In this argument, the students claimed that their Fischer esterification reaction yielded them an ester product and support this claim with the qualitative data, related to the product's smell, they collected from their reaction. The student rationalized this claim by describing how their product characteristically smelled of propyl acetate. This student relied solely on the smell of their product to support their claim despite having collected both IR and NMR spectral data on their ester product. Due to their inability to coordinate these pieces of evidence in support of their claim, this rationale was coded as fragmented. Additionally, the student made two other arguments that focused independently on the IR and NMR data as they compartmentalized their data analysis and interpretation in support of their proposed claims about the identity of their ester product.

### *Alignment*

This characteristic was fairly consistent across each of our eight experiments with the "aligned" code ranging in frequency of occurrence between 79.6% for the substitution reaction experiment and 100.0% for the GC experiment. Arguments coded as "aligned" has a consistent focus from claim to rationale, as seen in the column chromatography experiment example in Table 16. In this example, the student makes a claim about the "very pure" $\beta$-carotene compound they isolate. The student then rationalizes this claim by discussing that the "$R_f$ value was 0.92" and that "both of them have reached the same height," referring to the standard that was compared to their isolated $\beta$-carotene sample. Similarly, the synthesis of esters example argument in Table 16 was also coded as "aligned" as it referred to the synthesized ester in both the claims and rationale components. Arguments coded as "misaligned" were far less common across each experiment we investigated. Consider the following example from the IR experiment:

**Claim:** *"Cis-3-hexen-1-ol has alkene, alcohol, and alkane groups."*

**Evidence:** *"O-H bond at 3333 cm-1, C=C bond at 1654 cm-1, and C-H bond at 3008 cm-1. These values match the known values for the functional group wavenumbers."*

**Rationale:** *"The peaks at different wavenumbers vary with the percent transmittance values."*

In this argument, the student made a claim about the functional groups present in their *cis*-3-hexen-1-ol substance and listed the wavenumber values obtained from their IR spectrum as their evidence. For their rationale, the student described one aspect of IR spectral data analysis. Although their rationale was related to IR analysis, the general statement about IR spectral interpretation ("different numbers vary with the percent transmittance values") was misaligned from their claim about the functional groups present in their substance of interest. Misaligned arguments were fairly uncommon across all eight experiments with none of the experiments having arguments coded as "misaligned" more or less frequently than expected.

### *Approach to Reasoning*

Students' arguments varied also greatly in this area for each of the eight experiments. The deductive approach to reasoning, in which students supported their specific claims by drawing from general chemical principles, appeared as a slight preference for both the elimination reaction (50.5%) and TLC (52.5%) experiments compared to the inductive approach (41.9% and 47.5%, respectively), but was most frequently employed in arguments from the substitution reaction (72.2%) experiments. Considering the following example from the substitution reaction experiment:

**Claim:** *"Tert-butyl chloride reacted very readily through $S_N1$ but not through $S_N2$."*

**Evidence:** *"When tert-butyl chloride was added to the solvent, a precipitate formed as a sediment at the bottom of the vial."*

**Rationale:** *"The molecular structure of this substrate contains a tertiary carbon, which react only by* $S_N1$*. This explains why the reaction occurred quickly in* $S_N1$ *but didn't react through* $S_N2$*."*

In this argument, the student rationalized their specific claim about the reactivity of *tert*-butyl chloride with conceptual knowledge about the reactivity of substrates containing tertiary carbons. Many of these arguments (76.5%) were rationalized with chemical principles related to molecular structure in lieu of emphasizing their observations of reaction conditions (46.3%) as rationale. Subsequently, along with showing the greatest preference for deductive reasoning, arguments from the substitution reaction experiment were characterized by the lowest frequency of the inductive approach to reasoning (13.0%). Arguments such as the substitution reaction example in Table 14 demonstrate the inductive approach to reasoning in which students support their general claims ("Carbons not $sp^3$ hybridized will not undergo substitution reactions") with their laboratory findings ("neither reaction took place"). This approach to reasoning was much more commonly characteristic of arguments from all other experiments, including the IR (69.2%), NMR (76.5%), and synthesis of esters (88.4%) experiments. Consider the synthesis of esters experiment example below:

**Claim:** *"Characteristic peaks show us our verified ester product structure."*

**Evidence:** *"There was a three hydrogen singlet peak downfield."*

**Rationale:** *"The three hydrogen singlet peak matches the structure of isopentyl alcohol confirming our product structure connectivity."*

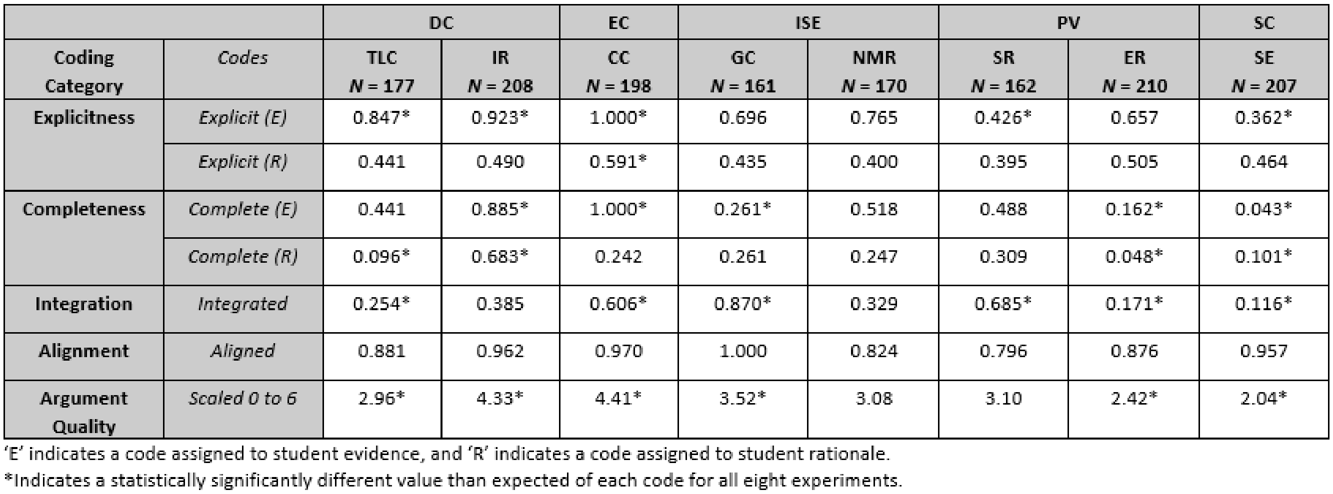
In this argument representative of those in the synthesis of esters experiment, the student inductively supported their claim about the successful synthesis of their "ester product" with a rationale that emphasized the NMR spectroscopic data the student collected during the course of their experiment. In addition to showing a highest frequency for inductive reasoning, arguments from the IR, NMR, and synthesis of esters experiments showed the greatest frequencies of rationale support with spectroscopic data (98.1%, 82.4%, and 59.4%, respectively) as well as the lowest rationale frequencies for reaction conditions (2.9%, 0.0%, and 33.3%, respectively).

The hybrid approach to reasoning, in which students supported their findings with both chemical concepts and observable data, was far less frequent across all laboratory experiment. The NMR (18.8%) and substitution reaction (14.8%) were demonstrated as the only experiments to have a higher frequency of this approach to reasoning compared to expectations across all experiments. Consider the NMR experiment rationale example shown below:

> **Rationale:** *"Since we know there are three neighbors (since it was a quartet peak), that means there cannot be 4 hydrogens bonded to a carbon (that would be methane). So, that indicates this is including more than one identical group (the integration). Then, since the IR spectra indicates the presence of a C=O bond, it is the next logical step to assume it is what connects the two groups previously determined from the NMR. We also know that since the triple is the most shielded, which is indicated by chemical shift, we know it is furthest away from the oxygen (most electronegative atom). "*

In this example, the student made a claim (not shown) in which they proposed the structure for their unknown substance. In support of this case-specific claim, the student drew from both NMR conceptual background and analysis of their own spectroscopic data. By identifying "the most shielded" peak as "indicated by chemical shift" the student applied NMR principles regarding the position of peaks in reference to their own data about "the triple" peak they observed on their spectra. Additionally, the student pieced together their proposed structure on the basis of their spectroscopic data in saying "there are three neighbors (since it was a quartet peak)." These

**Table 18** Average quality (relative frequency) of the six domain-general codes associated with student argument quality for each of our eight experiments.



| | | DC | | EC | ISE | | PV | | SC |
|---|---|---|---|---|---|---|---|---|---|
| **Coding Category** | *Codes* | **TLC** *N* **= 177** | **IR** *N* **= 208** | **CC** *N* **= 198** | **GC** *N* **= 161** | **NMR** *N* **= 170** | **SR** *N* **= 162** | **ER** *N* **= 210** | **SE** *N* **= 207** |
| **Explicitness** | *Explicit (E)* | 0.847* | 0.923* | 1.000* | 0.696 | 0.765 | 0.426* | 0.657 | 0.362* |
| | *Explicit (R)* | 0.441 | 0.490 | 0.591* | 0.435 | 0.400 | 0.395 | 0.505 | 0.464 |
| **Completeness** | *Complete (E)* | 0.441 | 0.885* | 1.000* | 0.261* | 0.518 | 0.488 | 0.162* | 0.043* |
| | *Complete (R)* | 0.096* | 0.683* | 0.242 | 0.261 | 0.247 | 0.309 | 0.048* | 0.101* |
| **Integration** | *Integrated* | 0.254* | 0.385 | 0.606* | 0.870* | 0.329 | 0.685* | 0.171* | 0.116* |
| **Alignment** | *Aligned* | 0.881 | 0.962 | 0.970 | 1.000 | 0.824 | 0.796 | 0.876 | 0.957 |
| **Argument Quality** | *Scaled 0 to 6* | 2.96* | 4.33* | 4.41* | 3.52* | 3.08 | 3.10 | 2.42* | 2.04* |

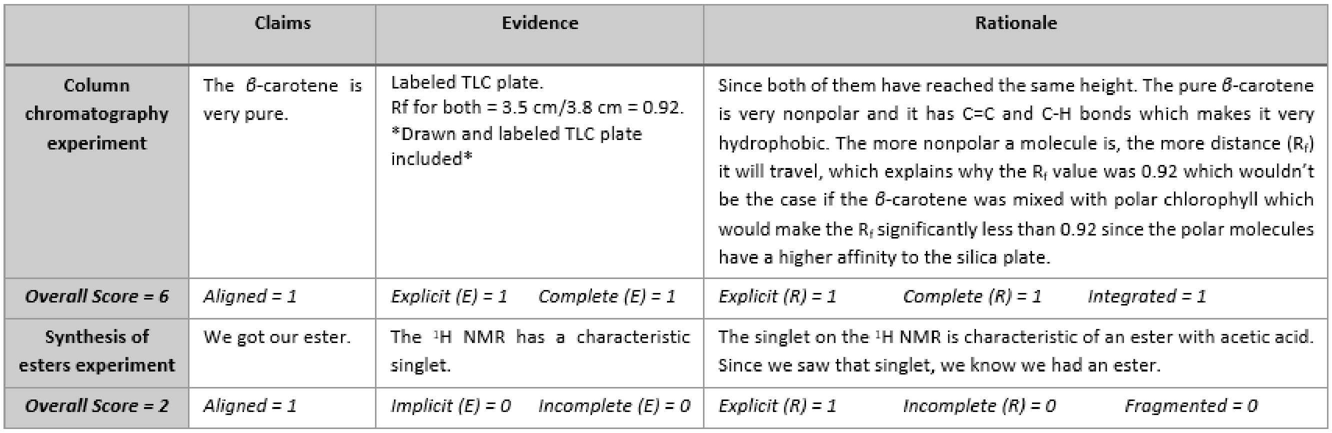
'E' indicates a code assigned to student evidence, and 'R' indicates a code assigned to student rationale.
*Indicates a statistically significantly different value than expected of each code for all eight experiments.

simultaneous references to the principles guiding analysis of an NMR spectrum and presentation of their own data analysis demonstrate the hybrid approach to reasoning that was most commonly associated with arguments produced from the NMR experiment. Similarly in the substitution reaction experiment, students produced a notable proportion of arguments following the hybrid approach to reasoning (14.8% of arguments) in which they supported their claims with a rationale referencing molecular structure (76.5% of rationale components) and their observed reaction conditions (46.3% of rationale components).

### *Quality of student argumentation*

Through the analysis of arguments from our eight experiments, we identified several domain-general coding categories that were associated with argument quality: explicitness, completeness, alignment, and integration. Quality scores included in Table 15 were calculated using the procedure described in the data analysis section. This table includes the average score (relative frequency) of each quality-based argument code broken down by laboratory experiment as well as the average overall argument score (from 0 to 6) of each experiment.

**Table 19** Argument quality of two student arguments: one from the column chromatography experiment and one from the synthesis of esters lab.



| | Claims | Evidence | Rationale |
|---|---|---|---|
| **Column chromatography experiment** | The *β*-carotene is very pure. | Labeled TLC plate.<br>Rf for both = 3.5 cm/3.8 cm = 0.92.<br>*Drawn and labeled TLC plate included* | Since both of them have reached the same height. The pure *β*-carotene is very nonpolar and it has C=C and C-H bonds which makes it very hydrophobic. The more nonpolar a molecule is, the more distance ($R_f$) it will travel, which explains why the $R_f$ value was 0.92 which wouldn't be the case if the *β*-carotene was mixed with polar chlorophyll which would make the $R_f$ significantly less than 0.92 since the polar molecules have a higher affinity to the silica plate. |
| ***Overall Score = 6*** | *Aligned = 1* | *Explicit (E) = 1* *Complete (E) = 1* | *Explicit (R) = 1* *Complete (R) = 1* *Integrated = 1* |
| **Synthesis of esters experiment** | We got our ester. | The $^1$H NMR has a characteristic singlet. | The singlet on the $^1$H NMR is characteristic of an ester with acetic acid. Since we saw that singlet, we know we had an ester. |
| ***Overall Score = 2*** | *Aligned = 1* | *Implicit (E) = 0* *Incomplete (E) = 0* | *Explicit (R) = 1* *Incomplete (R) = 0* *Fragmented = 0* |

The overall average argument score of the 1,493 arguments analyzed in this research study was 3.23 out of a possible 6. The column chromatography experiment contained arguments of the highest quality (4.41 out of 6). An example from the column chromatography experiment is shown in Table 16 to demonstrate the assignment of argument quality scores for each experiment. In this example, the argument presented a clear evidence component that fully detailed the TLC data collected (as well as a drawing of the TLC plate – not shown) and how these data were used to calculate the $R_f$ value of the observed *β*-carotene compound. This argument was also coded as "explicit" and "complete" for the rationale component, "integrated", and "aligned" and, thus, received a total quality score of 6 out of 6. The column chromatography experiment led to arguments that, on average, had the highest quality of the eight analyzed in this study, including having the greatest quality for the explicitness of evidence, explicitness of rationale, and completeness of evidence as well as the second highest quality of alignment and third highest quality of integration.

The IR and GC experiments also led to arguments with higher-than-expected quality. The IR experiment was characterized by the highest quality of completeness of student rationale components as well as the second highest explicitness and completeness of evidence. Similarly, the GC experiment was associated with arguments with the highest quality in alignment and integration.

Arguments from the substitution reaction experiment (3.10) had average quality. Despite having the second highest quality in completeness of rationale and integration, this experiment included the lowest quality arguments in terms of explicitness of rationale and alignment, as well as the second lowest quality in explicitness of evidence. Similarly, the NMR experiment (3.08) was linked to arguments with an average quality in most categories.

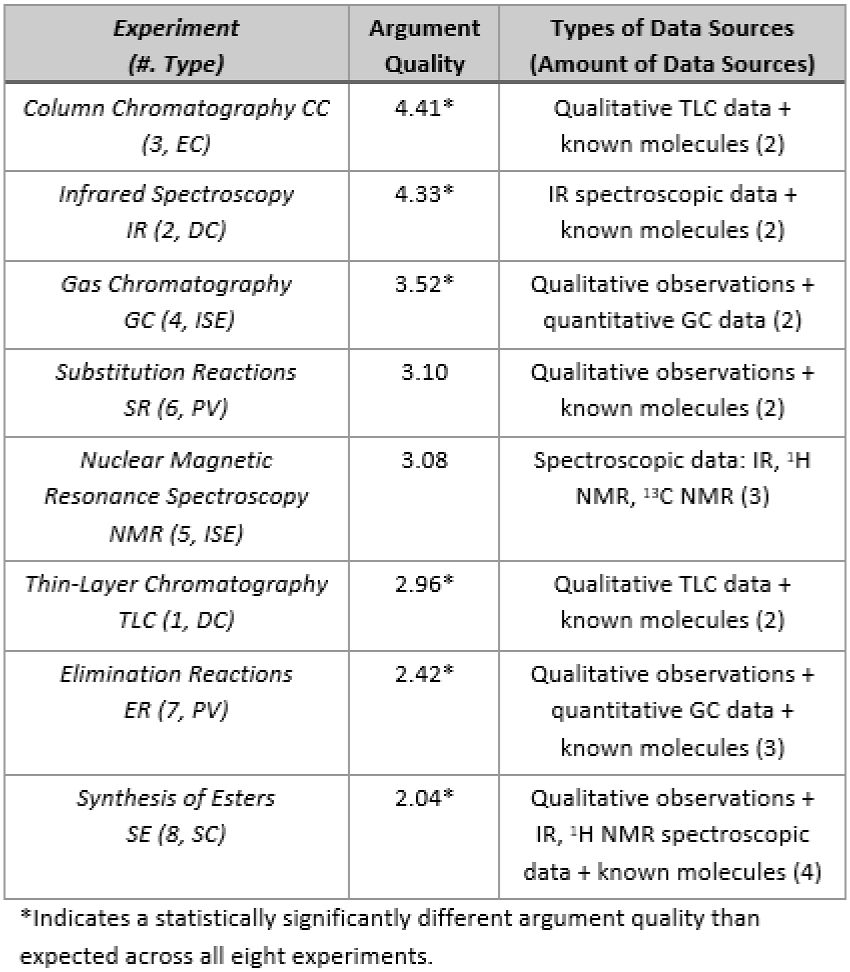
Arguments from the TLC, elimination reaction, and synthesis of esters experiments were of lower quality than expected. The TLC experiment was the first experiment performed by students and was characterized by arguments with the third lowest overall quality (2.96). These arguments had the second lowest completeness of rationale and the third lowest integration quality. The elimination reaction experiment (2.42) had arguments with the second lowest overall quality, which included several coding categories with lower-than-expected quality compared to the other experiments.

The final experiment in the semester was the synthesis of esters experiment and resulted in arguments of the lowest average quality. Table 16 includes an example of argument quality characterization for the synthesis of esters experiment. In this example, the student provided an aligned argument in which the focus of the claim and the rationale are both "our ester" product that "we know we had." The unclear description of only the "characteristic singlet" observed in their NMR spectrum neglected to include description of other peaks as well as the other types of data collected about their ester product. Thus, the argument was rated as both "implicit" and "incomplete" quality of evidence. In their chemical rationale, although the student explicitly tied the presence of "the singlet on the $^1$H NMR" to the ester product they synthesized, they included an incomplete description of how their evidence supported their claim . The student produced a fragmented rationale in which they identified only NMR data and failed to coordinate other observations, data, and structural information about their ester product.

### Which factors are most impactful and how do these factors affect the nature and quality of student argumentation across each of our eight experiments of interest?

To address our second research question, we identified several factors that we thought would impact the nature and quality of student argumentation across the eight laboratory experiments. These factors included experiment order in the semester, experiment type, the amount and types of data sources available to students, and the level of scaffolding for student argumentation present in each experiment. Below we describe and exemplify the impact of these factors on the nature and quality of student arguments.

**Table 20** Average argument quality (ranked from high to low), amount of data sources, and types of data sources for each laboratory experiment.



| *Experiment (#. Type)* | Argument Quality | Types of Data Sources (Amount of Data Sources) |
|---|---|---|
| *Column Chromatography CC (3, EC)* | 4.41* | Qualitative TLC data + known molecules (2) |
| *Infrared Spectroscopy IR (2, DC)* | 4.33* | IR spectroscopic data + known molecules (2) |
| *Gas Chromatography GC (4, ISE)* | 3.52* | Qualitative observations + quantitative GC data (2) |
| *Substitution Reactions SR (6, PV)* | 3.10 | Qualitative observations + known molecules (2) |
| *Nuclear Magnetic Resonance Spectroscopy NMR (5, ISE)* | 3.08 | Spectroscopic data: IR, $^{1}$H NMR, $^{13}$C NMR (3) |
| *Thin-Layer Chromatography TLC (1, DC)* | 2.96* | Qualitative TLC data + known molecules (2) |
| *Elimination Reactions ER (7, PV)* | 2.42* | Qualitative observations + quantitative GC data + known molecules (3) |
| *Synthesis of Esters SE (8, SC)* | 2.04* | Qualitative observations + IR, $^{1}$H NMR spectroscopic data + known molecules (4) |

*Indicates a statistically significantly different argument quality than expected across all eight experiments.

#### ***Experiment order***

We hypothesized that as students became increasingly familiar with the practice of post-lab argumentation, argument quality would increase throughout the course of a semester. As shown in Table 17, where experiments are listed in order of decreasing argument quality, that was not the case. Students' arguments from the last two experiments of the semester, ER and SE, exhibited the lowest quality of all, while arguments built for the second (IR) and third (CC) experiments in the sequence were found to have the highest quality as characterized in our study. We also analyzed the quality of each domain-general coding category in relation to experiment placement during the semester. Notably, the elimination reaction and synthesis of esters experiments were characterized by arguments of the lowest quality for the explicitness of evidence (SE), completeness of evidence (SE), completeness of rationale (ER), and integration (SE) coding categories compared to the other experiments. Our data indicated that experiment order did not have an impact on overall argument quality nor on the quality of any of the domain-general coding categories of interest.

### ***Experiment type***

We also hypothesized that the type of laboratory experiment (*e.g.*, prediction-verification versus synthesis and characterization) would be an impactful factor on the nature and quality of student argumentation. We speculated that different types of experiments would exert varied cognitive challenges and tap on diverse conceptual resources. However, our data suggests that experiment type was not a major determinant of argument quality. For example, although the TLC and IR experiments were both characterized as data collection, analysis, and interpretation (DC) experiments, students' arguments from these two labs significantly differed in both overall quality and specific quality for several domain-general categories of analysis. The overall quality of arguments from the IR experiment was the second highest (4.33), while arguments from the TLC lab were characterized as having the third lowest quality (2.96). Students' arguments from the IR experiment exhibited higher completeness of both the evidence and rationale components, while arguments from the TLC were characterized by a greater degree of differentiation. Additionally, arguments from the IR lab were more frequently based on an inductive approach to reasoning using spectroscopic data compared to arguments from the TLC lab. Equivalent disparities were observed between arguments from other experiments in the same category, while experiments classified as different types led to arguments of similar overall and specific quality. This is exemplified by the IR and CC labs in the group of experiments with highest overall argument quality and the ER and SE labs on the bottom of the list in Table 17.

### ***Amount and types of data sources***

Each of our eight experiments presented students with different pieces of data to collect, analyze, and utilize as they made sense of their laboratory findings as summarized by Table 17. We speculated that more data-intensive experiments would present a greater challenge for students as they had to coordinate more pieces and types of data to construct their post-lab arguments. Our analysis showed that the amount and types of data presented to students did have an impact on both the nature and quality of their post-lab arguments.

For example, access to both chromatographic data and spectroscopic data during an experiment (*e.g.*, IR, NMR, and SE labs) often led students to engage in an inductive approach to reasoning, while the absence of these types of data was more frequently linked to arguments built using deductive reasoning (based on theoretical knowledge). This latter case is exemplified by the substitution reaction (SR) lab in which students had access to only observational data and their deductive arguments were strongly based on their content knowledge about $S_N1$ and $S_N2$ mechanistic paths. In contrast, in the elimination reactions (ER) experiment, in the same category (PV) as the substitution reactions lab, students gathered chromatographic data and their reliance on inductive reasoning when building arguments was significantly larger.

The amount and types of data available to students also impacted the quality of their argumentation. For example, the CC and IR experiments had the fewest data sources for students to utilize and resulted in the highest quality arguments in our analysis. On the other hand, arguments from experiments in which several types of data were available to students led to lower quality arguments, as highlighted by the elimination reaction and synthesis of esters experiments. Lower argument quality suggested that students had difficulty coordinating multiple pieces of evidence. As shown in Table 12, students' arguments from these two labs were characterized as highly incomplete and fragmented.

### *Scaffolding of student argumentation*

Finally, we hypothesized that the level of scaffolding of student work could impact argument quality. We characterized level of scaffolding in terms of the amount of guidance provided to students in building their arguments through available resources, such as the laboratory workbook. We speculated that more explicit guidance would enhance the quality of students' arguments and our results seemed to confirm it.

For example, in the CC experiment, the workbook explicitly asked students to make claims regarding the "purity of your isolated $\beta$-carotene" and the relative non-polarity of $\beta$-carotene related to "what you would predict based on the structures." In this case, the students in our sample constructed arguments with case-specific claims that addressed the explicit guidance given to them. Explicit directions were also given on the evidence to be provided by asking students to "draw the results from your TLC" and "calculate the $R_f$ value of each spot." As a result, all analyzed arguments included the requested elements and were coded as both explicit and complete in the evidence component. This explicit guidance also impacted student rationales, which were characterized by a high frequency of multiple differentiation and a slight preference for an inductive approach to reasoning. In contrast, explicit scaffolding for argumentation was missing in the TLC lab in which students collected and analyzed the same amount and types of data sources as the CC lab (see Table 8) in addition to considering the same chemical concepts of structure, purity, and polarity. In this case, student arguments were of characteristically lower quality in their explicitness and completeness of both the evidence and rationale components. In their rationales, students relied less heavily on comparisons between their observed molecules and more on using theoretical constructs, which led to a slight preference for deductively framed arguments. Despite several striking similarities between the resources available in these experiments, the explicit scaffolding for argumentation in the CC lab correlated with arguments of the highest quality (4.41) while a lack of scaffolding was linked to TLC arguments whose quality (2.96) was below the average for the experiments analyzed in this study.

## Discussion

Throughout an academic semester in a college chemistry laboratory students work on different experiments that often provide diverse scenarios from which they are expected to construct meaning. After collecting and analyzing their own data, students are likely to follow varied approaches to reasoning as they analyze and communicate their laboratory findings. In our work, we identified similarities and differences in the nature and quality of the arguments generated by students working on eight distinct experiments in an organic chemistry lab. In particular, we characterized the claims, evidence, and rationale components in students' arguments along seven domain-general dimensions of analysis as summarized in Table 12.

Our analysis revealed very little variability in the specificity of student claims and the alignment of argument components. The majority of collected arguments across the eight experiments had a case-specific focus on laboratory observations and were aligned in terms of the concepts and ideas used when making claims, providing evidence, and constructing rationales. Nevertheless, the explicitness and completeness of student evidence and rationale components varied greatly across each of the eight experiments. Although the evidence presented by students in their arguments was largely characterized as explicit across most labs, the evidence component for two experiments (substitution reaction and synthesis of esters) was mostly implicit. In general, student rationales were split between being explicit and implicit in nature. While some experiments (elimination reaction and synthesis of esters) were characterized by highly incomplete evidence and rationale components alike, the column chromatography experiment was uniquely coded as having unanimously complete evidence and highly incomplete rationales amongst student arguments. Several experiments (notably IR, NMR, and synthesis of esters) were characterized by very few if any differentiation between substances and reactivity whereas some experiments (elimination reaction) or rationale components (column chromatography and GC) emphasized comparisons between substances and their properties. The level of integration of chemical concepts and ideas with data/observations was also highly variable within student rationales with some experiments (column chromatography, GC, and substitution reaction) more frequently characterized by integrated rationales while the remaining experiments displayed a majority of rationales coded as fragmented. Student approach to reasoning was often either deductive or inductive depending on the experiment, and very few instances of a hybrid approach were observed. Experiments that led to the collection of chromatographic and spectroscopic data (SR) tended to result in arguments that more often followed and inductive approach to reasoning, while experiments in which these types of data were not generated or provided led to arguments that were more frequently characterized as deductive.

We also explored the impact (or lack thereof) of various laboratory factors on the nature and quality of student reasoning as manifested in their written arguments. These factors included experiment order, experiment type, amount and types of data sources available to students, and the level scaffolding provided. Several prior studies have reported increased argument quality over time in an academic course (Hosbein *et al*., 2021; Çetin and Eymur, 2017; Walker and Sampson, 2013a; Sampson *et al*., 2010). However, in our case we observed no noticeable trend in argument quality over the course of the academic semester. We also speculated that experiments of the same type would be characterized by similar argument quality. However, we found no relationship between type of experiment and argument quality. In our study, student arguments built while working on experiments classified within the same category (*e.g.*, data collection, analysis, and interpretation) often exhibited very different qualities.

On the other hand, our analysis suggested that the amount and type of data sources available to students in an experiment as well as the degree of scaffolding provided greatly impacted both the nature and quality of the arguments they constructed. Less data-intensive experiments presented students with an easier task of analyzing their data to use in constructing well-aligned arguments. Although at times these arguments still lacked clarity and key details in supporting their claims, experiments that involved fewer data sources and types of data also trended towards higher argument quality. Conversely, experiments that were more data-intensive and included a greater quantity of data sources as well as different types of data tended to be characterized by arguments of lower quality, such as both the ER and SE experiments. Our findings are in alignment with previous studies suggesting that more data-intensive experiments that demand the consideration of different types of evidence often lead to student arguments that lack sophistication in the coordination of both empirical and theoretical pieces (Havdala and Ashkenazi, 2007;

Sandoval, 2003). Recent research studies in chemistry education have shown that although students may be able to collect and analyze spectroscopic evidence, they often struggle to construct arguments that are consistent with the entirety of their data (Stowe and Cooper, 2019).

As part of our analysis, we also elucidated the effect of argument scaffolding on the quality of student reasoning. We identified experiments that, despite being similar in terms of the amount and types of data available to make sense of laboratory findings, led to student arguments of quite different quality. In these, cases the amount of explicit scaffolding provided for the construction of arguments seemed to be responsible for the difference. For example, students in the column chromatography experiment were explicitly prompted to make specific types of claims as well as present distinct pieces of evidence. Meanwhile, the TLC experiment involved collection of the same data and an identical conceptual focus but included no prompting for student argument components. Overall, CC experiment arguments were characterized as having the highest quality while the TLC experiment was characterized by arguments of the third lowest quality. Previous studies have presented conflicting findings as to the impact of scaffolding on students' arguments. In one case, simplifying instructional contexts through scaffolding facilitated more complex argumentation (Berland and McNeill, 2010), while another recent investigation showed that explicit prompt scaffolding had no significant impact on students' data-based inferences (Stowe and Cooper, 2009). These conflicting results suggest that more research is needed to better understand how scaffolding may affect student argumentation.

## Implications

Our first investigation of student argumentation in college organic chemistry labs (Petritis *et al.*, 2021) allowed us to develop an analytical framework to characterize the nature and quality of student arguments at domain-general and domain-specific levels. That study showed the major impact that the framing of laboratory activity can have on student argumentation. In this contribution, we expanded upon this work to better understand how other factors may affect the nature and quality of students' arguments. The results of both studies provide information and insights that can be used by designers of laboratory curricula, laboratory managers, and trainers of laboratory instructors to enhance the quality of the arguments that students build individually or collectively in the laboratory.

In our study, the quality of written arguments was closely related to the nature and amount of the data with which students had to grapple. This result suggests that laboratory designers, managers, and instructors should pay close attention to the types of data students are expected to collect and analyze in any given experiment, and purposefully select and sequence experimental activity to gradually increase the variety and complexity of the data to be analyzed. Our findings indicate that, in terms of supporting the development of argumentation abilities, the type of experiment that students conduct is likely less important than the nature and amount of data they are expected to analyze and integrate to make sense of their results.

Our results also point to the need for purposeful support of student argumentation as learners become more familiar with this epistemic practice. Various studies suggest that students may not know how to argue in the laboratory context or may lack clarity of the goal of their argumentative task (Berland and Hammer, 2012; Garcia-Mila *et al*., 2013). Thus, chemistry students are likely to benefit from explicit guidance on how to coordinate experimental data with core chemical concepts and ideas in developing high-quality arguments. McNeill *et al*. (2006) suggested that fading written instructional scaffolds for argumentation in chemistry better prepared students to produce higher quality arguments compared to students lacking this support. Related studies propose developing learning progression for the epistemic practice of argumentation to improve student performance across the semester (Berland and McNeill, 2010; Smith *et al*., 2006). These recommendations along with our findings highlight the need for students to be trained in argumentation and have opportunities to learn how to argue from evidence, especially early on in their laboratory experience.

While our findings inform us on several key factors that affect student argumentation in the organic chemistry laboratory, more research is needed to characterize the effect of other variables on the development of this scientific practice, such as the nature of laboratory instruction and the interaction with peers. If we are to foster the ability of students to productively engage in scientific argumentation, it is critical that chemistry education researchers and practitioners better understand the various factors that affect students' engagement in this epistemic practice and the mechanisms through which these factors affect student reasoning.

## Limitations

Our findings emerged from the analysis of arguments generated by students in a single organic chemistry laboratory course at our institution and may thus not be generalizable to other contexts. Although laboratory instructors can be expected to play an important role in the laboratory environment, our analysis did not consider the effect that differences in instruction may have had on the written arguments that were collected. Additionally, some of the written arguments we analyzed were produced individually while others were constructed in groups of two to five. Although other studies have investigated the role of student collaboration on the outcome of argumentation (Sampson and Clark, 2009), we did not analyze differences that could exist between individually-constructed and group-constructed post-lab arguments. Lastly, we acknowledge that the laboratory learning environment entails an intricate web of factors each of which has the potential to impact student arguments. Thus, further study is warranted to investigate the interplay of these laboratory components and the impact they have on the nature and quality of student reasoning.

The content of this chapter has been reproduced, with permission from the Royal Society of Chemistry, from Petritis, S. J., Kelley, C., and Talanquer, V., (2022), Analysis of factors that affect the nature and quality of student laboratory argumentation, *Chem. Educ. Res. Pract.*, advance *article*.

# Chapter 5: Conclusions and Future Directions

The research described herein characterized the nature and quality of student arguments and attributed similarities and differences in argumentation to the various factors associated with students' laboratory experiences. Throughout the course of reading and analyzing the student arguments collected for these research projects, we were able to develop a robust framework for characterizing both the domain-specific chemical concepts and ideas invoked by students in their arguments as well as domain-general features that illustrate how students constructed and manifested their written chemical reasoning. Our framework is summarized by the domain-specific codes (see Chapter 3) and domain-general coding categories (see Table 11 in Chapter 4) presented in the preceding chapters. This framework was emergent from the student argument data we collected and allowed us to characterize the nature of student argumentation across an entire semester of organic chemistry laboratory course experiments, including one experiment that was framed in two distinct ways. We were also able to devise a quantitative measure of student argument quality as noted by the explicitness (evidence and rationale), completeness (evidence and rationale), alignment (claim to rationale), and integration (rationale) coding categories in our framework. Ultimately, the framework helped us describe the impact of various laboratory factors in the nature and quality of student arguments.

Our initial work focused on the impact of activity framing in the organic chemistry laboratory, utilizing arguments constructed as a result of a single laboratory experiment (substitution reaction) framed in two ways: predict-verify and observe-infer (Chapter 3). In the predict-verify frame, students had knowledge of the molecular structures of the compounds they were tasked with reacting under both $S_N1$-favored and $S_N2$-favored reaction conditions while observe-infer students were presented with unknown molecular structures and tasked with performing the same sets of reactions. Not surprisingly, students who worked under the predict-verify frame more frequently included references to molecular structure in their arguments whereas observe-infer students emphasized observations about the reaction conditions of their experiment. However, in addition to making more class-level claims, students who worked under the predict-verify frame more often lacked a complete description of their supporting evidence as compared to observe-infer students. We speculated that having access to the molecular structure of their compounds discouraged students from completely describing their evidence and rationale under the predict-verify frame. Predict-verify students often conflated theoretical constructs and the empirical data they collected from their series of reactions, which coincided with arguments that were characteristically more fragmented and misaligned. Observe-infer students, on the other hand, produced more highly integrated, explicit, and complete rationale components that more frequently relied on descriptions of their empirical data as support for their claims. As for their approach to reasoning, observe-infer students produced arguments that were highly inductive in which they grounded their findings in their experimental data. In contrast, predict-verify students produced arguments that were characteristically deductive in their approach to reasoning, drawing from general chemical principles in support of their findings.

Our findings from this initial study demonstrated the importance of activity framing and the necessity that this laboratory factor be a great consideration to laboratory instructors and managers when designing, implementing, and assessing their laboratory curricula. Such a simple switch, such as concealing the identity of reaction starting materials was shown to have a major impact on the way students manifested their chemical reasoning, specifically with regards to the theoretical concepts and empirical data they observed, utilized, and coordinated in their post-lab written arguments. Additionally, this study led our research team to consider the effect of other laboratory factors, including those described in our second project (Chapter 4).

This subsequent project involved expanding the use of our analytical framework to the eight experiments that included post-lab arguments in the first semester of the organic chemistry laboratory curriculum. In this study, we solely utilized the domain-general coding categories to characterize both the nature and quality of student arguments produced from these eight experiments. We hypothesized that arguments from experiments of similar type (*i.e.*, data collection, interpretation, and analysis vs. prediction-verification) would exhibit similar characteristics of both nature and quality. We found this laboratory factor to not be particularly impactful for the experiments we investigated. Similarly, there were observed no trends between the experimental order in the semester and the quality of student arguments. For example, the seventh and eighth experiments of the semester (elimination reaction and synthesis of esters) were characterized to have arguments of the lowest quality, respectively. However, we did note that both the amount and types of data sources and the degree of explicit scaffolding for argumentation were impactful factors for student argument quality. For instance, arguments from experiments (elimination reaction and synthesis of esters) that involved greater number and types of experimental evidence had lower argument quality scores while experiments with fewer data pieces (column chromatography, infrared spectroscopy, and gas chromatography) were characterized by arguments of higher quality overall. Experiments, such as the column chromatography experiment, in which argument components were given to students or when there was explicit prompting also coincided with higher argument quality compared to less scaffolded experiments.

## Implications

Although providing ripe opportunity to engage in science practices (such as argument from evidence), the undergraduate chemistry laboratory is an entangled web of factors each of which has the potential to impact how students make sense of their findings. The findings from our two major research studies involving student argumentation can serve several benefits to both research and practice within chemistry education. For chemistry education research, this work contributes the development of a detailed framework designed to characterize the nature and quality of undergraduate student written argumentation in the laboratory. This analytical framework allowed us to discover the diversity of student arguments and to begin to unravel the complex web of factors that make up students' experiences in the teaching laboratory. Even though we already characterized several impactful factors (activity framing, amount and types of data sources, and scaffolding for argumentation), we have identified several other aspects of the laboratory environment that may also prove impactful towards student argumentation (explained in greater detail below), including the role of the graduate student instructors (GSIs) and effect of student collaboration on their chemical reasoning. These factors must be studied further to help researchers better understand how interpersonal interactions impact student reasoning in the chemistry laboratory and, ultimately, how these interactions foster or hinder students' meaningfully engagement with chemistry concepts and ideas. Our analytical tool could also be harnessed to investigate a wide variety of instructional settings both within chemistry and in other laboratory-based disciplines. The domain-specific codes we identified could easily be expanded to include other areas of chemistry to aid researchers in understanding how students discuss code chemistry concepts and ideas across various chemical disciplines. Meanwhile, our domain-general coding categories can be customized to other science education research disciplines that include argument-focused laboratory curricula.

Further, our characterization of argument quality also has the potential to be used as both a pedagogical and research tool within the chemistry laboratory. Graduate student instructors (GSIs) can receive professional development training on the coding categories we utilized to characterized students arguments. GSIs can be taught how to code arguments themselves and these codes could serve a dual purpose as research data and for evaluation of laboratory course artifacts. In this way, research data can be collected and analyzed by GSIs while the research team has quicker access to an expanded data set that helps investigate other laboratory factors as well as develop evidence-based pedagogical tools to help GSIs develop greater expertise with facilitating classroom argumentation with their students.

With respect to chemistry education practice, the findings from our studies suggest the need for instructional support and scaffolding for students as they become familiar with argumentation. Although it is somewhat surprising that students demonstrated the lowest argument quality at the end of the semester (elimination reaction and synthesis of esters experiments), students received no formal training on argumentation involving the claims, evidence, and rationale framework. However, we did see promising results as the column chromatography experiment was both the most scaffolded and was characterized by arguments of the highest quality. To address this issue of scaffolding and epistemic training of argumentation, I would suggest for instructional managers to be deliberate in how they structure their laboratory curricula around argument from evidence. For example, having more structured instructional scaffolds in the early weeks of a semester could help to familiarize students with the components of an argument as well as how to construct arguments that meet our expectations as instructors. Further, along with this instructional support, students must be more thoroughly educated on how to construct an argument as well as see a purpose for argumentation in the first place. Students need to see examples of how to integrate experimental data and theoretical concepts and ideas if they stand a chance at developing their own sense of how to construct chemical reasoning in the laboratory. Over the course of a semester, these instructional supports could be lessened just as experiments increase in complexity (*i.e.*, increase the amount of data for students analyze and coordinate in their arguments). Between the previous experience with formulating arguments and explicit feedback on how to coordinate different pieces of data and consider opposing ideas in their arguments, students are more likely to craft higher quality arguments and propose more sophisticated involving the experiments they observed.

## Additional interventions and argumentation research

Our findings have begun to uncover and characterize some factors within the chemistry laboratory. However, there remain many aspects of the chemistry laboratory that have yet to be characterized in terms of their impact on how students engage in argumentation. Although not included in the scope of the two major projects associate with this dissertation, I have collected additional data and consider other potentially impactful laboratory factors. Below, I describe some of these additional interventions, present some of my anecdotal findings, and discuss the potential these areas have for future research opportunities.

### *Additional framing research*

In addition to collecting student argument data from the substitution reaction experiment involving two laboratory frames, we also collected “re-framed” argument data from the thin-layer chromatography (TLC), infrared (IR) spectroscopy, and elimination reaction experiments. In each of these cases, experiments were also framed to remove the identity of the compounds that students observed during their experiments. For example, in the TLC experiment, students were tasked with comparing the relative polarity of their compounds of interest without knowing their molecular structures. Initial analysis of these arguments suggests that students more frequently relied on theoretical constructs (*i.e.*, discussion of intermolecular forces) as they compared the observed TLC behavior of their unknown compounds. Although this finding is opposite that found for the substitution reaction experiment in which students relied more heavily on their empirical in the absence of compound identity, it provides an interesting contrast to the substitution reaction experiment that warrants future study. Specifically, these anecdotal findings emphasize the interplay of various laboratory factors (besides just activity framing) and suggest an implicit hierarchy with which students are impacted by aspects of their laboratory environment.

### *Modifications to chemical reasoning*

Following the argumentation session conducted by students involved in the observe-infer frame of the substitution reaction experiment, the identity of their eight unknown starting materials were revealed to the students. Students working under this frame were then given the chance to conduct a second argumentation session during which they could now explicitly consider the molecular structure of the compounds they had just reacted. Initial analysis of the arguments produced during this session suggests that students more frequently coordinate their empirical data (collected about the previously unknown compounds) with theoretical concepts related to the specific structure of the molecules. However, some students decided to revise their previously constructed arguments in light of new evidence and produced novel claims to reflect the additional evidence they now had in the form of a molecular structure. Future research is warranted to investigate the nature of claim change and modification to argumentation as new evidence is obtained, both in the context of this experiment and in other experiments in which new evidence can be revealed to students or collected through another analytical method.

### *Graduate student instructors (GSIs)*

One major factor that has yet to be characterized through our research studies is the impact of the graduate student instructor (GSI) on student chemical reasoning. To our knowledge, there are no related studies that explicitly consider and characterize the role laboratory instructors play in guiding student argumentation, especially in chemistry education research. This is certainly a challenging venture as GSIs each present their own strengths and limitations as instructors (*i.e.*, years of experience, confidence in laboratory techniques, comfort with chemical concepts and ideas, inter-personal skills, etc.), which are hard to characterize following methods that have been previously been described by our studies and others' in the field of argumentation research. However, through my qualitative observations of various GSIs instructing students in the organic chemistry laboratory, it is quite apparent that different instructional styles has the potential to impact student reasoning whether in the content students focus on in their arguments or other aspects of the nature and quality of their written argumentation. Thus, future research study would explicitly characterize how the laboratory learning environment fostered by various GSIs impacts the domain-specific, domain-general, and quality-based features of student laboratory argumentation. Such a study would make ample use of qualitative measures that involve both argument analysis as we have previously reported and video/audio recordings of the GSI and students as they navigate the chemistry laboratory environment.

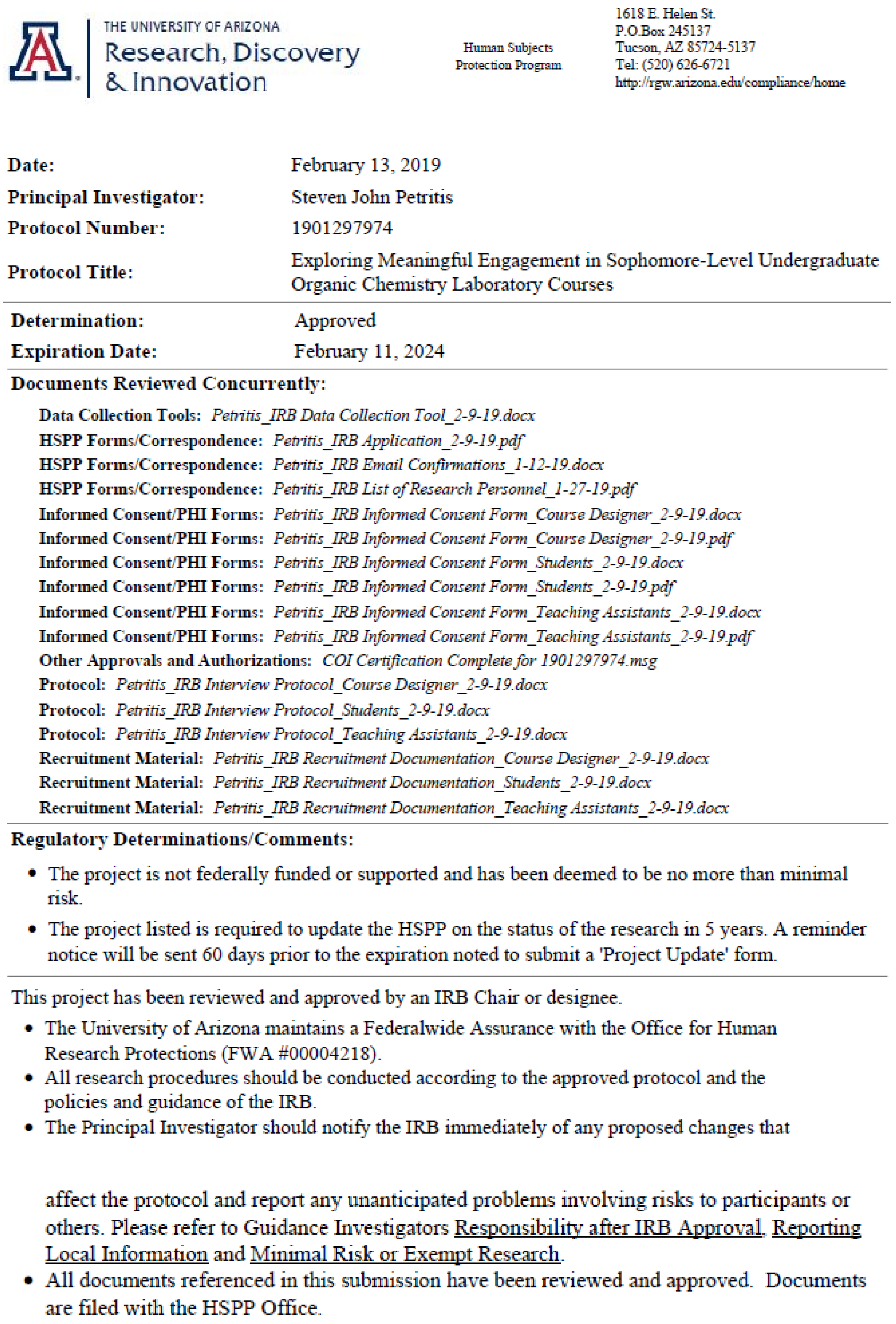
### *Individual versus collaborative argumentation*

As noted both in Chapter 3 and 4, another factor that remains unexplored by our research studies is the impact of collaboration on the nature and quality of student argumentation. Each GSI is typically given the freedom to decide if they want students to produce arguments individual, in pairs, or in groups of three to five students. As a result, each combination of students presents a unique environment for laboratory argumentation. For example, while some GSIs are comfortable allowing their students to work in small groups throughout the semester, others work predominately individually and experience rare instances of group argumentation, which often happens towards the end of the semester in the laboratory. Inevitably, the ability or inability to engage in a productive exchange of chemical ideas and observations has the potential to impact how students manifest their reasoning as they make sense of their laboratory findings (either individually or with their peers). Anecdotally from our argumentation data, we have observed many instances in which individually-constructed arguments tend to appear more sophisticated, of higher quality, and more frequently coordinate theory and empirical data. Conversely, there have been many noted instances in which group-produced arguments appear more simplistic in nature and less frequently coordinate their ideas in support of their claims. These initial findings suggest inefficiencies in the process of group argumentation that warrant further study. Thus, I propose performing a follow-up qualitative study to record group interactions as they construct their laboratory arguments and analyze these findings in reference to the written arguments students submit at the end of this group argumentation session.

### ***Domain-specific chemistry concepts and ideas***

As a part of our second major study, we also analyzed student arguments across each of our eight experiments using our previously established domain-specific coding scheme. Although we expanded this coding scheme to account for all the chemistry concepts, ideas, and types of data described in student arguments, we chose not to include these data in our manuscript for this work due to the sheer volume of data associated with this analysis. These data serve an important role in analyzing how students discuss chemistry ideas in their arguments and, thus, we propose they could be used for future work focused on the domain-specific knowledge in student written arguments. Appendix E includes these data, broken down by argument component and the domain-specific code present in each laboratory experiment. Work in this area could help to better gauge which chemical concepts, ideas, and data are more challenging for students to consider in their arguments, as noted by arguments of lower quality. Thus, additional support and instruction could, be focused on these topics to ensure students are better equipped to reason with these conceptual areas.

# Appendix A – Human Subject Approval



THE UNIVERSITY OF ARIZONA
Research, Discovery
& Innovation

Human Subjects
Protection Program

1618 E. Helen St.
P.O.Box 245137
Tucson, AZ 85724-5137
Tel: (520) 626-6721
http://rgw.arizona.edu/compliance/home

| | |
|---|---|
| **Date:** | February 13, 2019 |
| **Principal Investigator:** | Steven John Petritis |
| **Protocol Number:** | 1901297974 |
| **Protocol Title:** | Exploring Meaningful Engagement in Sophomore-Level Undergraduate Organic Chemistry Laboratory Courses |
| **Determination:** | Approved |
| **Expiration Date:** | February 11, 2024 |

**Documents Reviewed Concurrently:**

**Data Collection Tools:** *Petritis_IRB Data Collection Tool_2-9-19.docx*
**HSPP Forms/Correspondence:** *Petritis_IRB Application_2-9-19.pdf*
**HSPP Forms/Correspondence:** *Petritis_IRB Email Confirmations_1-12-19.docx*
**HSPP Forms/Correspondence:** *Petritis_IRB List of Research Personnel_1-27-19.pdf*
**Informed Consent/PHI Forms:** *Petritis_IRB Informed Consent Form_Course Designer_2-9-19.docx*
**Informed Consent/PHI Forms:** *Petritis_IRB Informed Consent Form_Course Designer_2-9-19.pdf*
**Informed Consent/PHI Forms:** *Petritis_IRB Informed Consent Form_Students_2-9-19.docx*
**Informed Consent/PHI Forms:** *Petritis_IRB Informed Consent Form_Students_2-9-19.pdf*
**Informed Consent/PHI Forms:** *Petritis_IRB Informed Consent Form_Teaching Assistants_2-9-19.docx*
**Informed Consent/PHI Forms:** *Petritis_IRB Informed Consent Form_Teaching Assistants_2-9-19.pdf*
**Other Approvals and Authorizations:** *COI Certification Complete for 1901297974.msg*
**Protocol:** *Petritis_IRB Interview Protocol_Course Designer_2-9-19.docx*
**Protocol:** *Petritis_IRB Interview Protocol_Students_2-9-19.docx*
**Protocol:** *Petritis_IRB Interview Protocol_Teaching Assistants_2-9-19.docx*
**Recruitment Material:** *Petritis_IRB Recruitment Documentation_Course Designer_2-9-19.docx*
**Recruitment Material:** *Petritis_IRB Recruitment Documentation_Students_2-9-19.docx*
**Recruitment Material:** *Petritis_IRB Recruitment Documentation_Teaching Assistants_2-9-19.docx*

**Regulatory Determinations/Comments:**

- The project is not federally funded or supported and has been deemed to be no more than minimal risk.
- The project listed is required to update the HSPP on the status of the research in 5 years. A reminder notice will be sent 60 days prior to the expiration noted to submit a 'Project Update' form.

This project has been reviewed and approved by an IRB Chair or designee.

- The University of Arizona maintains a Federalwide Assurance with the Office for Human Research Protections (FWA #00004218).
- All research procedures should be conducted according to the approved protocol and the policies and guidance of the IRB.
- The Principal Investigator should notify the IRB immediately of any proposed changes that affect the protocol and report any unanticipated problems involving risks to participants or others. Please refer to Guidance Investigators Responsibility after IRB Approval, Reporting Local Information and Minimal Risk or Exempt Research.
- All documents referenced in this submission have been reviewed and approved. Documents are filed with the HSPP Office.

# Appendix B – Institutional Review Board (IRB) Consent Form

**University of Arizona**
**Student Consent to Participate in Research**

**Study Title: Exploring Meaningful Engagement in Sophomore-Level Undergraduate Organic Chemistry Laboratory Courses**

**Principal Investigator: Steven Petritis**

**You are being asked to participate in a research study.** Your participation in this research study is voluntary and you do not have to participate. This document contains important information about this study and what to expect if you decide to participate. Please consider the information carefully. Feel free to ask questions before making your decision whether or not to participate.

- The purpose of this research study is to investigate how weekly in-lab activities in an undergraduate organic chemistry laboratory course are effective at meaningfully engaging students with the central ideas of chemistry.
- The research study will last for the entirety of this semester's Chemistry 243A/B laboratory courses (and will continue each semester, including the summer, through the end of the Spring 2022 semester).
- The only requirements of the study are that you complete the normally assigned written and oral components of the Chemistry 243A/B laboratory courses.
- Video and audio recordings will be conducted of post-lab round-up discussion sessions.
- The educational records (i.e., written products) from each laboratory experiment (i.e., quizzes, laboratory workbooks, post-lab round-up summaries) will be collected, de-identified, photocopied/scanned, and returned to your TA promptly following the completion of each week's laboratory session.
- Written products from each laboratory experiment will not be used to publish research data from this study until after final grades are submitted each semester.
- The typical time commitment for participating is the same time required for attendance at weekly laboratory sessions for the Chemistry 243A/B laboratory courses.
- The only additional time commitment is that you may be asked to complete brief interviews, outside of in-lab time, regarding your responses to the in-lab activities conducted throughout the semester of the Chemistry 243A/B laboratory courses.
- With your permission, I would like to audiotape these interviews so that I can make an accurate transcript. Once I have made the transcript, I will erase the recordings. Your name will not be in the transcript or my notes.
- There are no expected risks to you as a result of participating in this study.
- You will not benefit directly from participating in this study.
- There will be no compensation for participation in this research study.

- If you choose not to participate, you are free to do so and your assessment in the Chemistry 243A/B laboratory courses will not be affected.
- Your responses will be assigned a code number. The list connecting your name to this code will be kept in an encrypted and password protected file. Only the research team will have access to the file. When the study is completed and the data have been analyzed, the list will be destroyed.
- Information collected about you will not be used or shared for future research studies.
- The information that you provide in the study will be handled confidentially. However, there may be circumstances where this information must be released or shared as required by law. The University of Arizona Institutional Review Board may review the research records for monitoring purposes.
- The educational records disclosed for this research study are the written products from each laboratory session throughout the semester (i.e., quizzes, laboratory workbooks, post-lab round-up summaries).
- These educational records will be disclosed in order for the researcher to investigate the extent of meaningful engagement of undergraduate organic chemistry students in the central ideas of chemistry.
- The disclosure of the educational records associated with this project can be made to Steven Petritis, the principal investigator of this research study.

Education records used by this research project are education records as defined and protected by Family Educational Rights and Privacy Act (FERPA). FERPA is a federal law that protects the privacy of student education records. Your consent gives the researcher permission to access the records identified above for research purposes.

For questions, concerns, or complaints about the study you may contact **Steven Petritis** by email **(petritis@email.arizona.edu)**.

For questions about your rights as a participant in this study or to discuss other study-related concerns or complaints with someone who is not part of the research team, you may contact the Human Subjects Protection Program at 520-626-6721 or online at http://rgw.arizona.edu/compliance/human-subjects-protection-program.

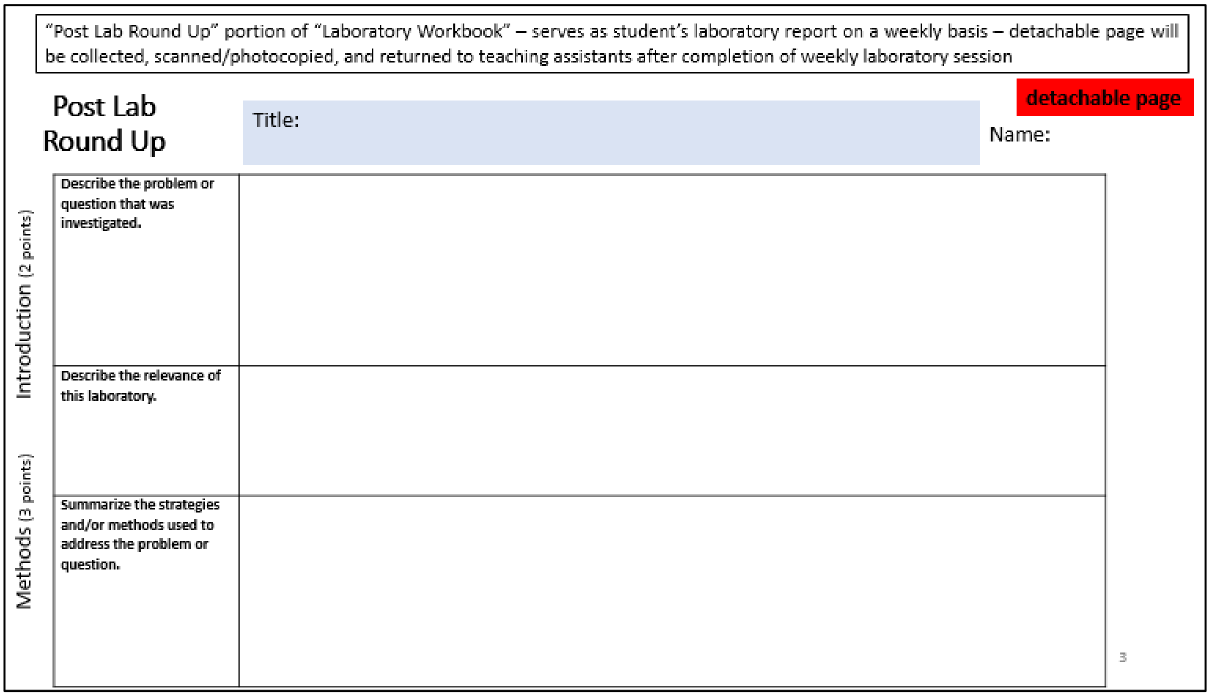
## Signing the consent form

I have read (or someone has read to me) this form, and I am aware that I am being asked to participate in a research study. I have had the opportunity to ask questions and have had them answered to my satisfaction. I voluntarily agree to participate in this study.

I am not giving up any legal rights by signing this form. I will be given a copy of this form.

| | | |
|---|---|---|
| **Printed name of subject** | **Signature of subject** | **Date** |

# Appendix C – Post-Lab Argument Data Collection Tool



"Post Lab Round Up" portion of "Laboratory Workbook" – serves as student's laboratory report on a weekly basis – detachable page will be collected, scanned/photocopied, and returned to teaching assistants after completion of weekly laboratory session

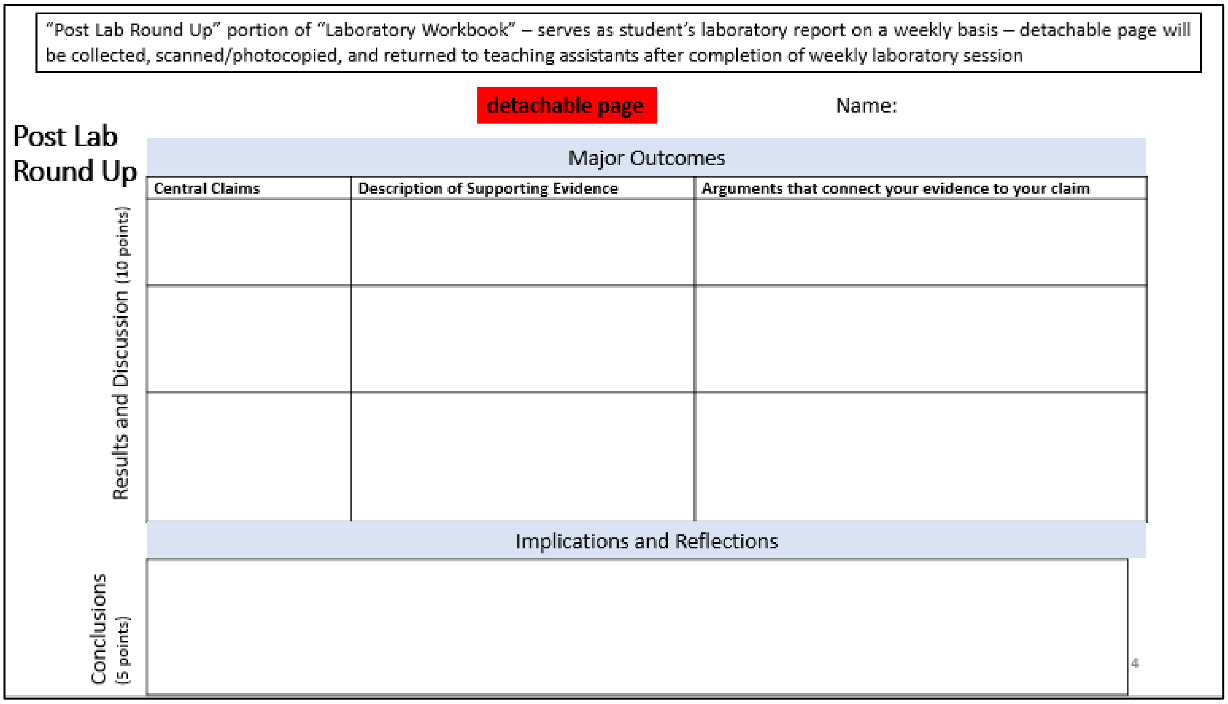
**detachable page**

Post Lab Round Up

Title:

Name:

| | | |
|---|---|---|
| Introduction (2 points) | Describe the problem or question that was investigated. | |
| | Describe the relevance of this laboratory. | |
| Methods (3 points) | Summarize the strategies and/or methods used to address the problem or question. | |



"Post Lab Round Up" portion of "Laboratory Workbook" – serves as student's laboratory report on a weekly basis – detachable page will be collected, scanned/photocopied, and returned to teaching assistants after completion of weekly laboratory session

**detachable page**

Name:

Post Lab Round Up

Major Outcomes

| | Central Claims | Description of Supporting Evidence | Arguments that connect your evidence to your claim |
|---|---|---|---|
| Results and Discussion (10 points) | | | |
| | | | |
| | | | |

Implications and Reflections

| | |
|---|---|
| Conclusions (5 points) | |

# Appendix D – Student Interview Protocol

**University of Arizona**
**Student Interview Protocol**

**Study Title: Exploring Meaningful Engagement in Sophomore-Level Undergraduate Organic Chemistry Laboratory Courses**

**Principal Investigator: Steven Petritis**

A total of 50 in-person interviews will be conducted each semester for the duration of this research study. These in-person interviews, between the principal investigator (Steven Petritis) and randomly selected student participants from each laboratory section observed in this research study (with their voluntary consent), will be ~15 minutes in duration and will serve as the only additional commitment for the student participants besides the normally assigned work for the Chemistry 243A/B laboratory courses. Interview sessions will be audio and video recorded within 48 hours of the conclusion of each week's laboratory session and will be immediately transcribed by the principal investigator upon completion of the interview session. Student participant identifiers (both educational records and interview recordings/transcriptions) will be kept confidential throughout the duration of this research study and their participation or lack thereof will not impact their assessment in the Chemistry 243A/B courses. Student participants who decline to consent specifically for interview sessions will still have their educational records used for the duration of the research study, assuming they still offer their voluntary consent to the research study.

The focus of these interview sessions will be on the educational records turned in by the student participants for their Chemistry 243A/B laboratory courses and the nature of their group-based discussion sessions during the post-lab portion of their weekly three hour laboratory session. During the interviews, the interviewer (principal investigator – Steven Petritis) will ask the interviewees (voluntary student participants) open-ended questions regarding their responses to the educational records turned in (see attached Data Collection Tool) and the group-based discussion they participate in each week during their Chemistry 243A/B courses. Additionally, student participants may also be shown video/audio clips of their group-based discussions from their laboratory sessions and asked similar open-ended question related to these experiences. Open-ended questions for the interview sessions will include, but not be limited to:

- What were you thinking when you wrote this?
- What were you doing at this point?
- How do you feel about what you wrote here?
- What were you doing here and why?
- Why did you do this here?
- How did you decide on this as the main problem in this laboratory experiment?

- What was your approach in addressing the problem/question for this laboratory experiment?
- What were the major claims, supporting evidence, and arguments/conclusions for this experiment and how did you decide on these?
- What are the implications of this laboratory experiment?

# Appendix E – Activity Framing Coding Data

**Table 21** Relative frequencies of domain-specific codes for the claims component of the predict-verify and observe-infer laboratory frames.

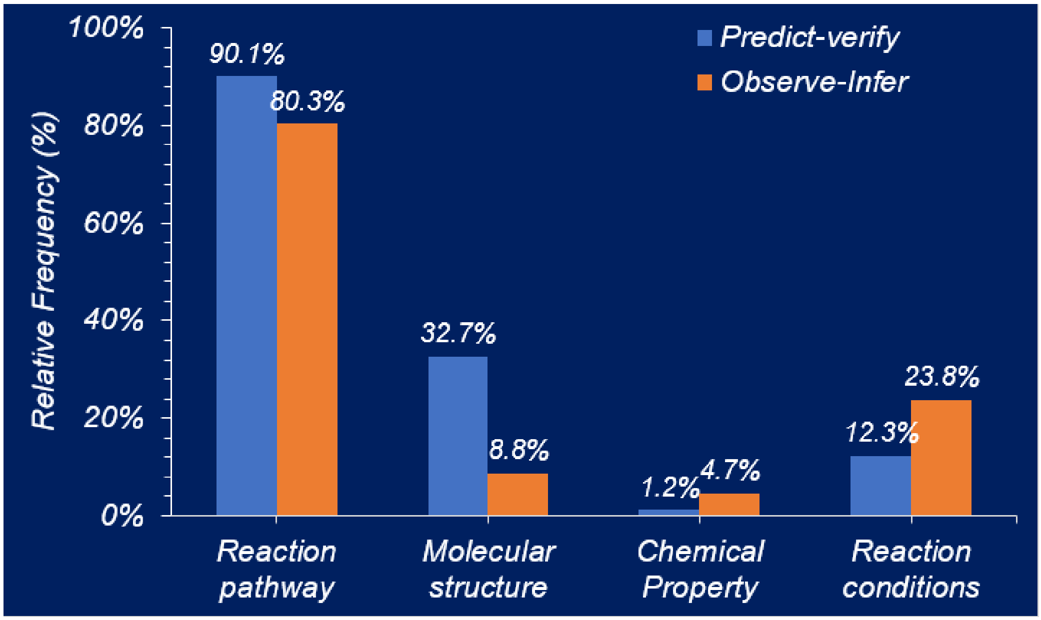


**Table 22** Relative frequencies of domain-specific codes for the evidence component of the predict-verify and observe-infer laboratory frames.

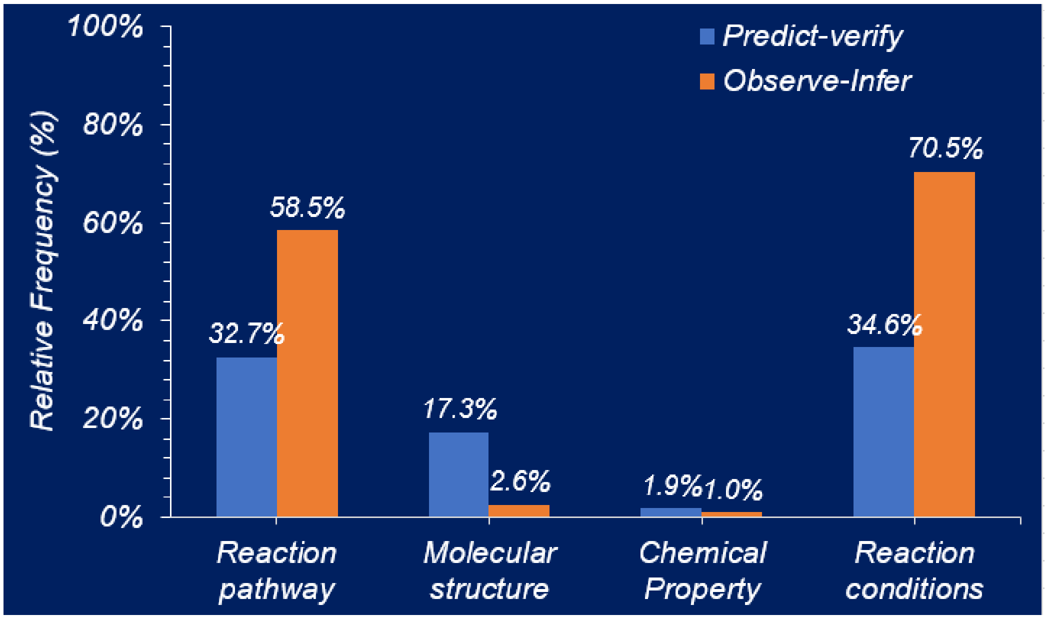


**Table 23** Relative frequencies of domain-specific codes for the rationale component of the predict-verify and observe-infer laboratory frames.

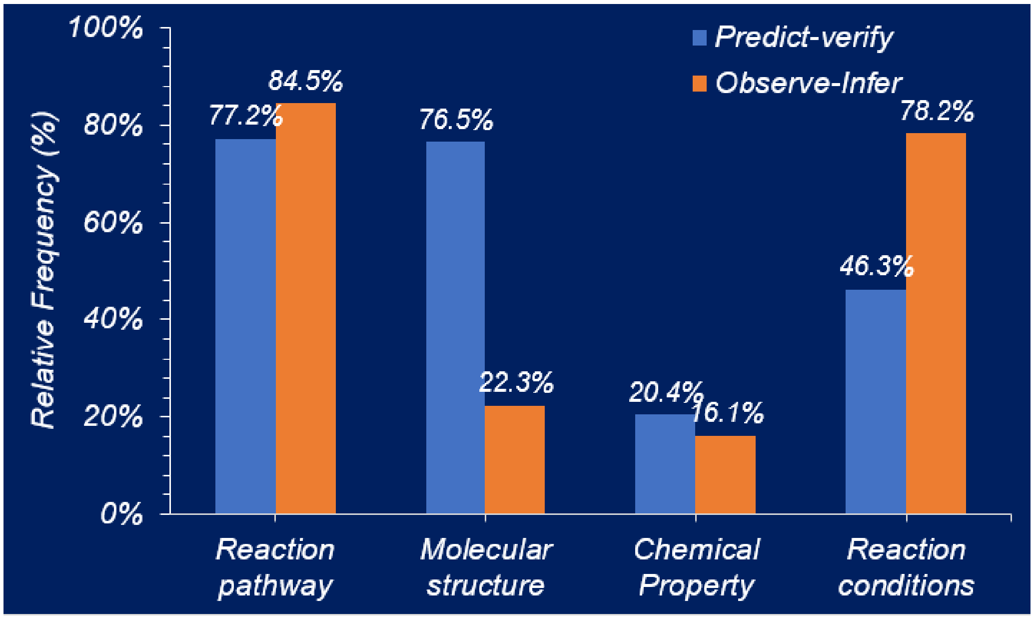


**Table 24** Relative frequencies of domain-general coding categories of the predict-verify and observe-infer laboratory frames.

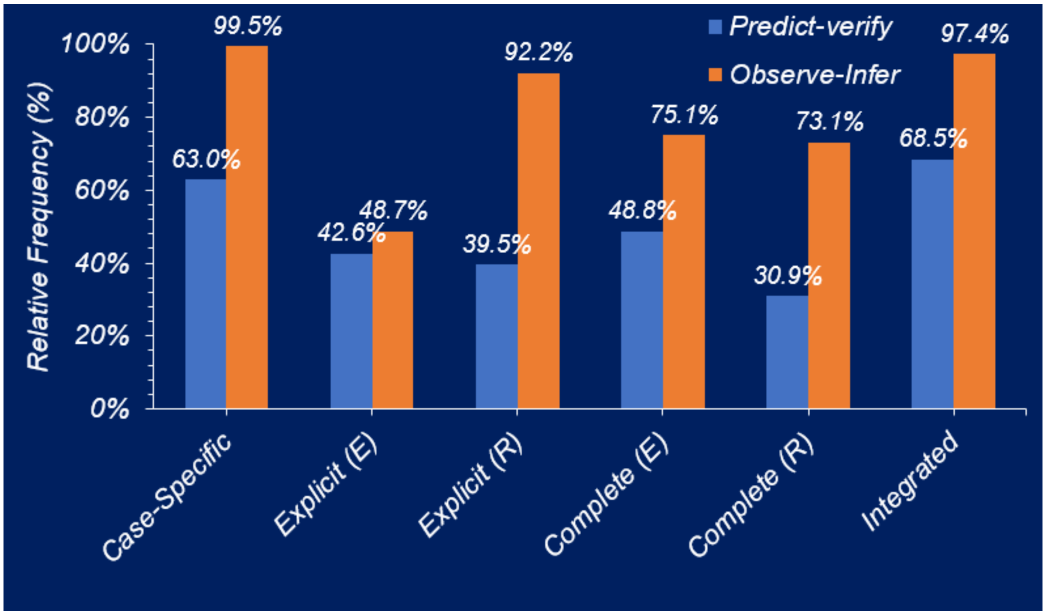
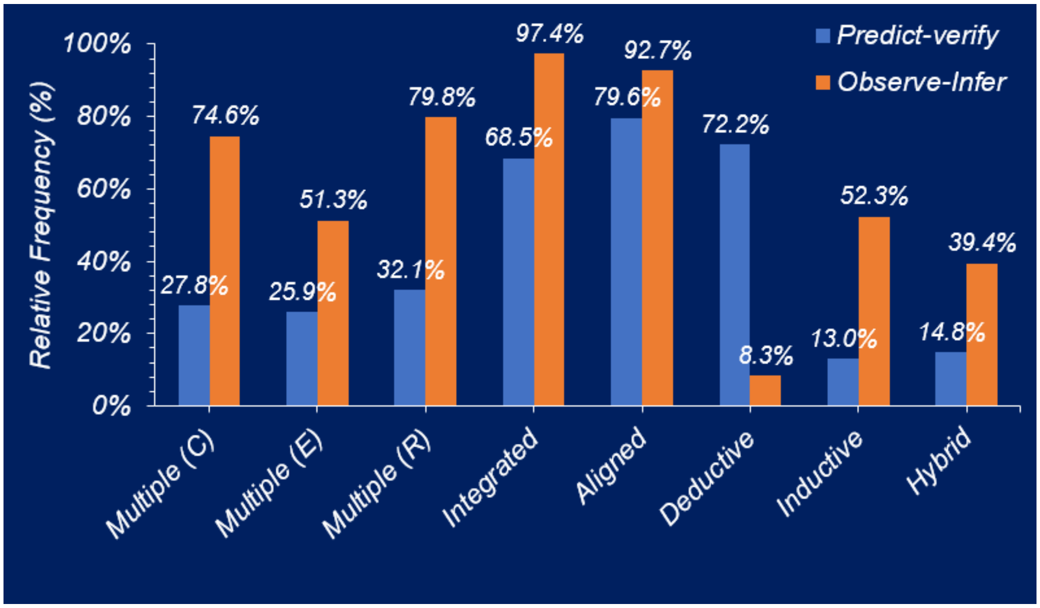


**Table 25** Relative frequencies of domain-general coding categories of the predict-verify and observe-infer laboratory frames.



# Appendix F – Domain-Specific Laboratory Factors Data

**Table 26** Domain-specific qualitative coding results (absolute and relative frequencies) for the claims, evidence, and rationale components of student arguments for each of the eight experiments.

| ***Claims Codes*** | **TLC** ***N* = 177** | **IR** ***N* = 208** | **Column** ***N* = 198** | **GC** ***N* = 161** | **NMR** ***N* = 170** | **Substitution** ***N* = 162** | **Elimination** ***N* = 210** | **Synthesis of Esters** ***N* = 207** |
|---|---|---|---|---|---|---|---|---|
| *Reaction Pathway* | - | - | - | - | - | 146 (90.1%) | 130 (61.9%) | - |
| *Reaction Conditions* | 27 (15.3%) | 6 (2.9%) | 3 (1.5%) | 133 (82.6%) | - | 20 (12.3%) | 104 (49.5%) | 57 (27.5%) |
| *Chemical Property* | - | 8 (3.8%) | 198 (100.0%) | 35 (21.7%) | - | 2 (1.2%) | 30 (14.3%) | 21 (10.1%) |
| *Molecular Structure* | - | 208 (100.0%) | 99 (50.0%) | 35 (21.7%) | 170 (100.0%) | 53 (32.7%) | 40 (19.0%) | 27 (13.0%) |
| *Chromatographic Data* | 39 (22.0%) | - | - | 63 (39.1%) | - | - | 20 (9.5%) | - |
| *Spectroscopic Data* | - | 8 (3.8%) | - | 7 (4.3%) | 170 (100.0%) | - | - | 9 (4.3%) |
| ***Evidence Codes*** | **TLC** ***N* = 177** | **IR** ***N* = 208** | **Column** ***N* = 198** | **GC** ***N* = 161** | **NMR** ***N* = 170** | **Substitution** ***N* = 162** | **Elimination** ***N* = 210** | **Synthesis of Esters** ***N* = 207** |
| *Reaction Pathway* | - | - | - | - | - | 53 (32.7%) | 74 (35.2%) | - |
| *Reaction Conditions* | 117 (66.1%) | 8 (3.8%) | - | 42 (26.1%) | - | 56 (34.6%) | 80 (38.1%) | 48 (23.3%) |
| *Chemical Property* | 54 (30.5%) | - | - | 14 (8.7%) | - | 3 (1.9%) | 8 (3.8%) | 60 (29.0%) |
| *Molecular Structure* | 12 (6.8%) | 122 (58.7%) | - | 70 (43.5%) | 170 (100.0%) | 28 (17.3%) | 12 (5.7%) | 54 (26.2%) |
| *Chromatographic Data* | 174 (98.3%) | - | 198 (100.0%) | 126 (78.3%) | - | - | 180 (85.7%) | - |
| *Spectroscopic Data* | - | 208 (100.0%) | - | 42 (26.1%) | 170 (100.0%) | - | - | 129 (62.3%) |
| ***Rationale Codes*** | **TLC** ***N* = 177** | **IR** ***N* = 208** | **Column** ***N* = 198** | **GC** ***N* = 161** | **NMR** ***N* = 170** | **Substitution** ***N* = 162** | **Elimination** ***N* = 210** | **Synthesis of Esters** ***N* = 207** |
| *Reaction Pathway* | - | - | - | - | - | 125 (77.2%) | 76 (36.2%) | - |
| *Reaction Conditions* | 153 (86.4%) | 6 (2.9%) | 144 (72.7%) | 119 (73.9%) | - | 75 (46.3%) | 106 (50.5%) | 69 (33.3%) |
| *Chemical Property* | 165 (93.2%) | 6 (2.9%) | 150 (75.8%) | 21 (13.0%) | 74 (43.5%) | 33 (20.4%) | 82 (39.0%) | 69 (33.3%) |
| *Molecular Structure* | 18 (10.2%) | 208 (100.0%) | 75 (37.9%) | 56 (34.8%) | 140 (82.4%) | 124 (76.5%) | 116 (55.2%) | 126 (60.9%) |
| *Chromatographic Data* | 135 (76.3%) | - | 174 (87.9%) | 126 (78.3%) | - | - | 90 (42.9%) | - |
| *Spectroscopic Data* | - | 204 (98.1%) | - | 28 (17.4%) | 140 (82.4%) | - | - | 123 (59.4%) |

‘-‘ identifies a domain-specific code that does not appear in either the claims, evidence, or rationale component of arguments for the different experiments.

# Appendix G – Domain-General Laboratory Factors Data

**Table 27** Relative frequencies of the domain-general specificity coding category for each of our eight experiments.

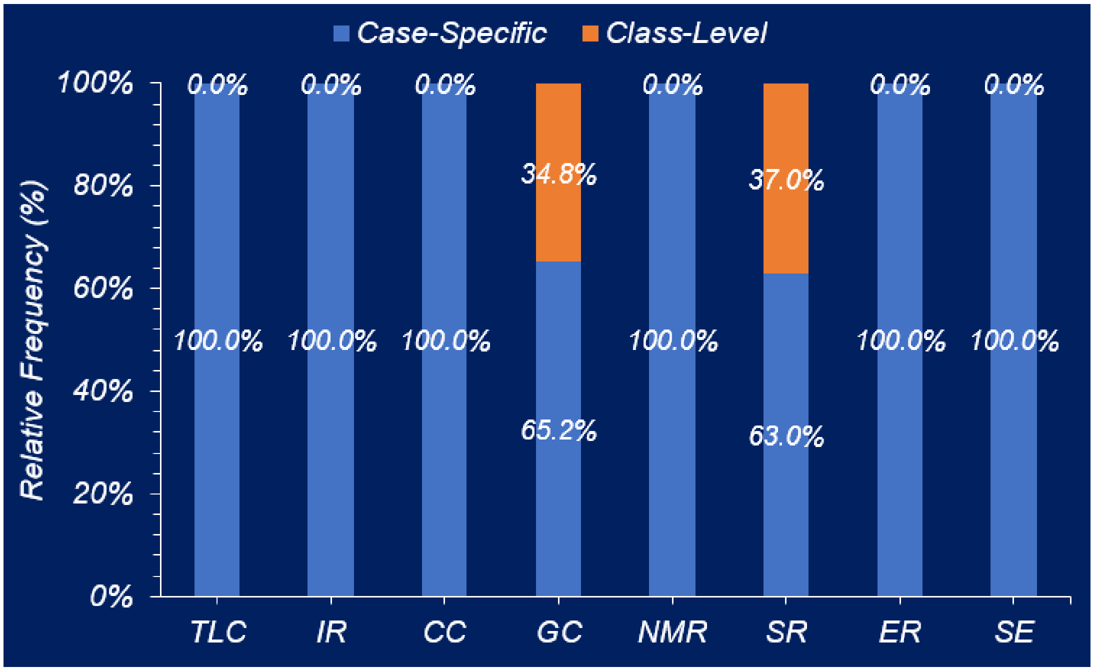


**Table 28** Relative frequencies of the domain-general explicitness of evidence coding category for each of our eight experiments.

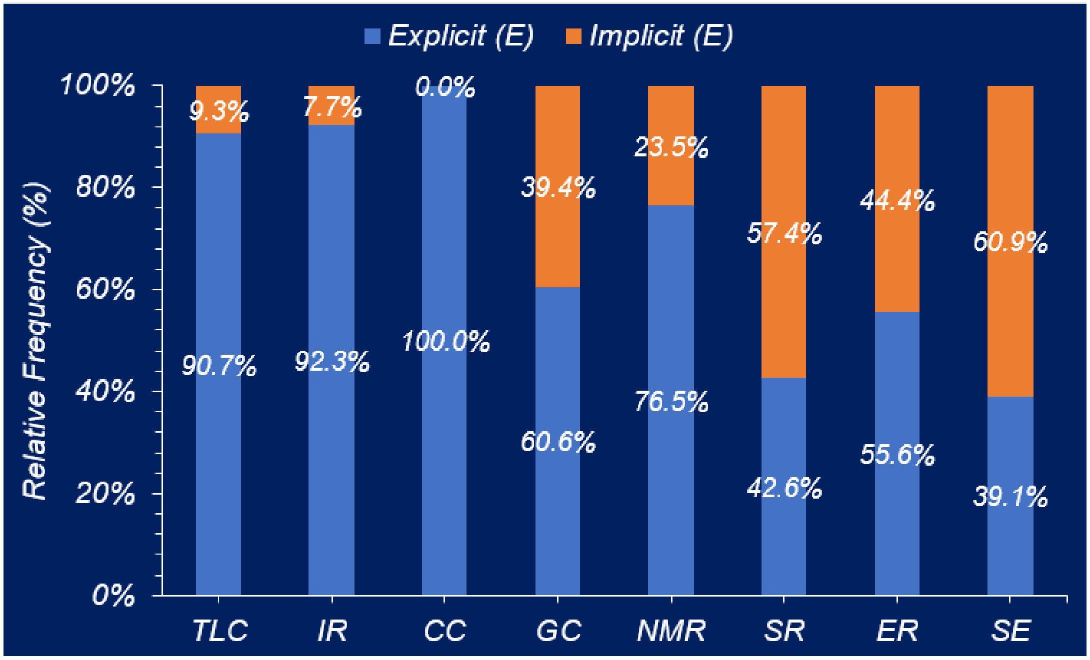


**Table 29** Relative frequencies of the domain-general explicitness of rationale coding category for each of our eight experiments.

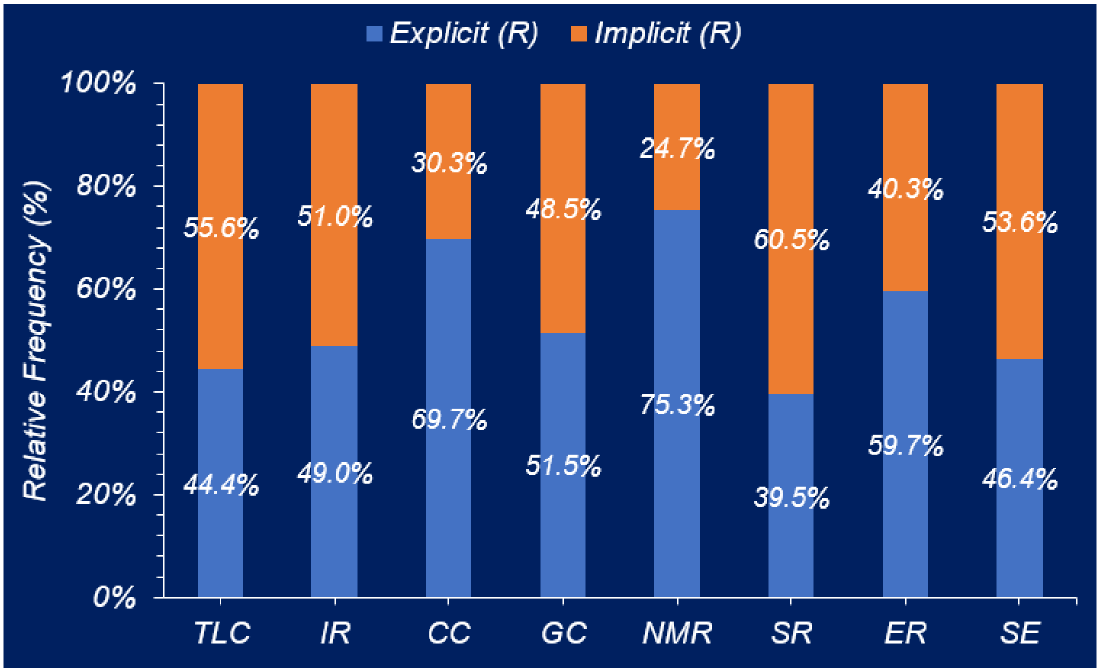


**Table 30** Relative frequencies of the domain-general completeness of evidence coding category for each of our eight experiments.

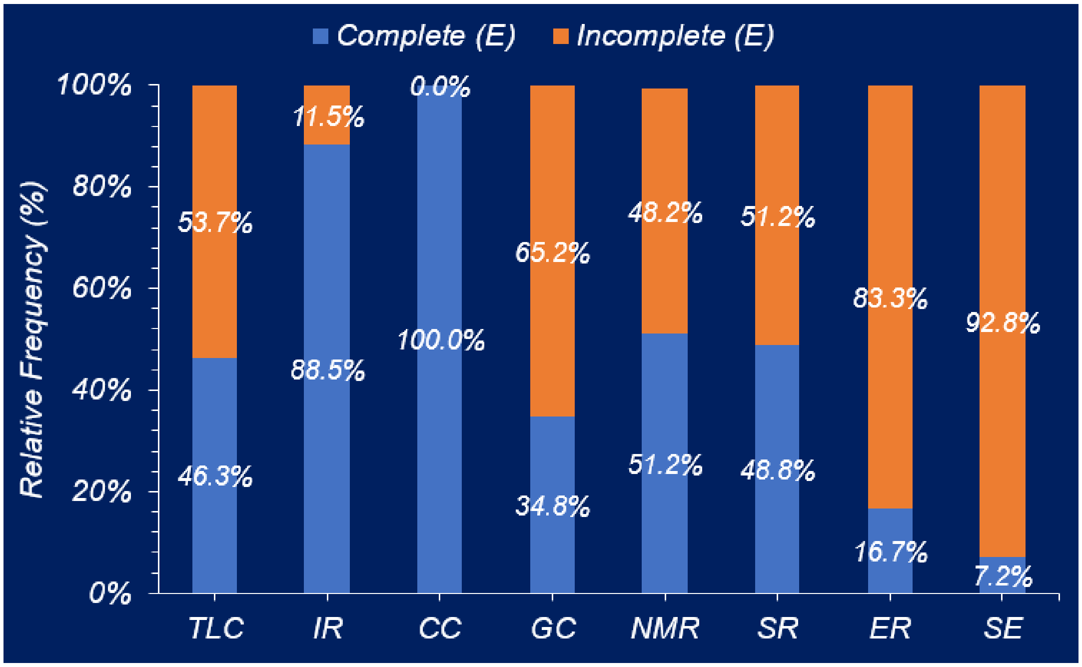


**Table 31** Relative frequencies of the domain-general completeness of rationale coding category for each of our eight experiments.

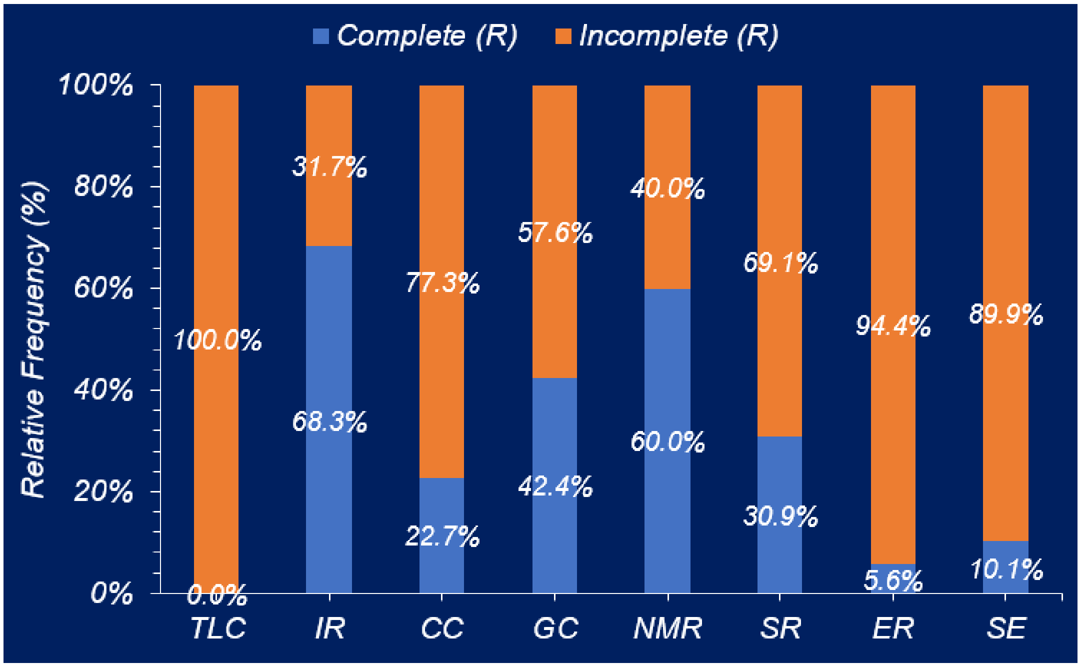


**Table 32** Relative frequencies of the domain-general differentiation of claims coding category for each of our eight experiments.

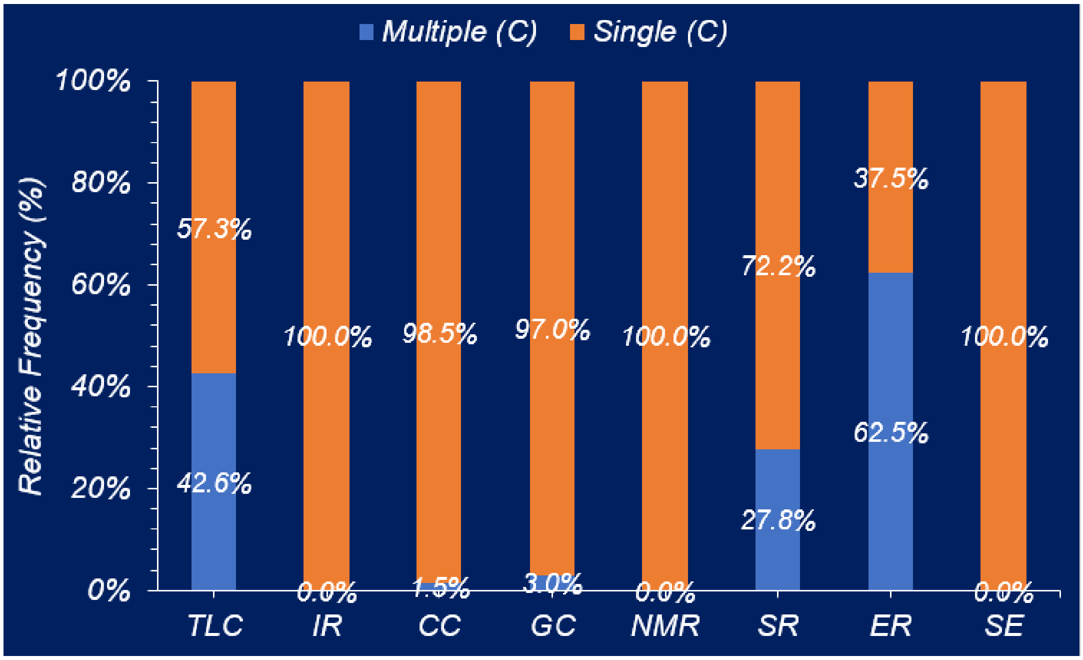


**Table 33** Relative frequencies of the domain-general differentiation of evidence coding category for each of our eight experiments.

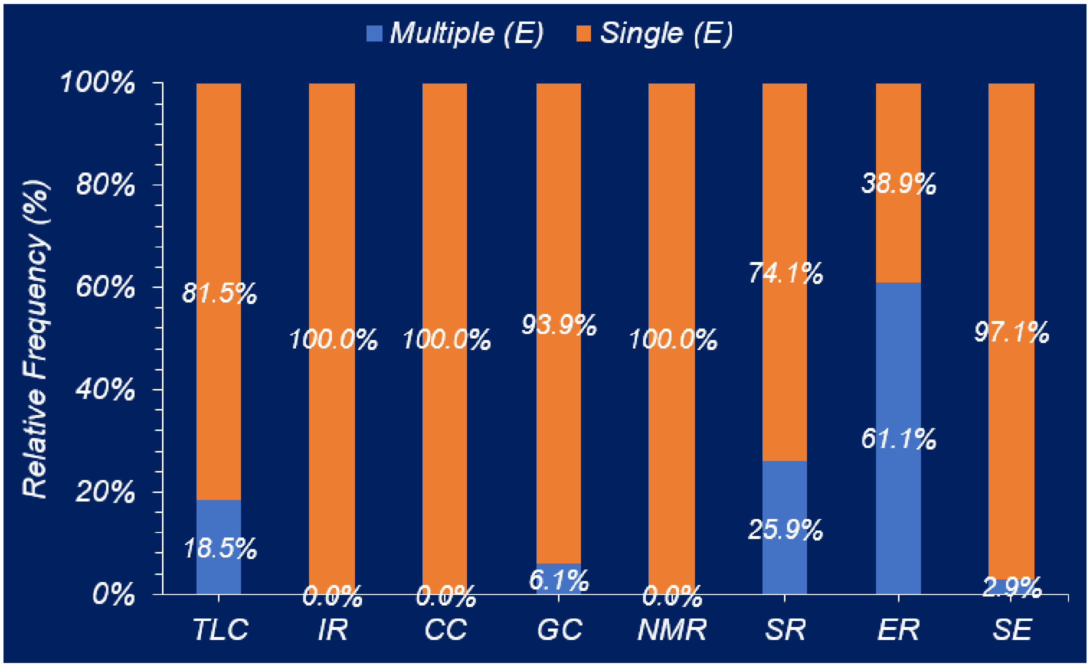


**Table 34** Relative frequencies of the domain-general differentiation of rationale coding category for each of our eight experiments.

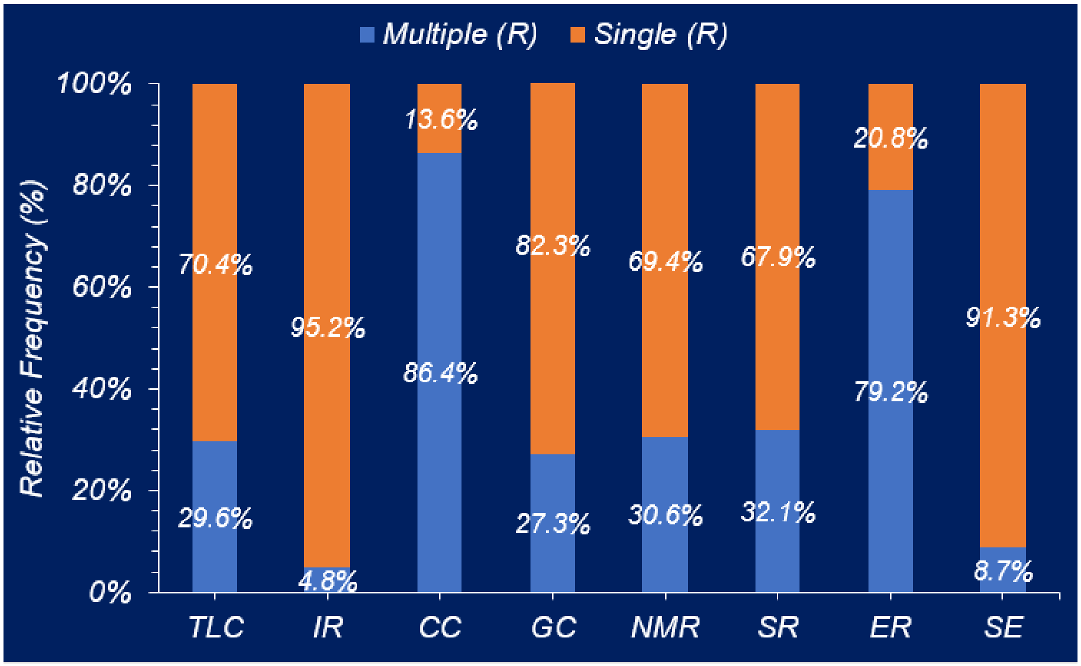


**Table 35** Relative frequencies of the domain-general integration coding category for each of our eight experiments.

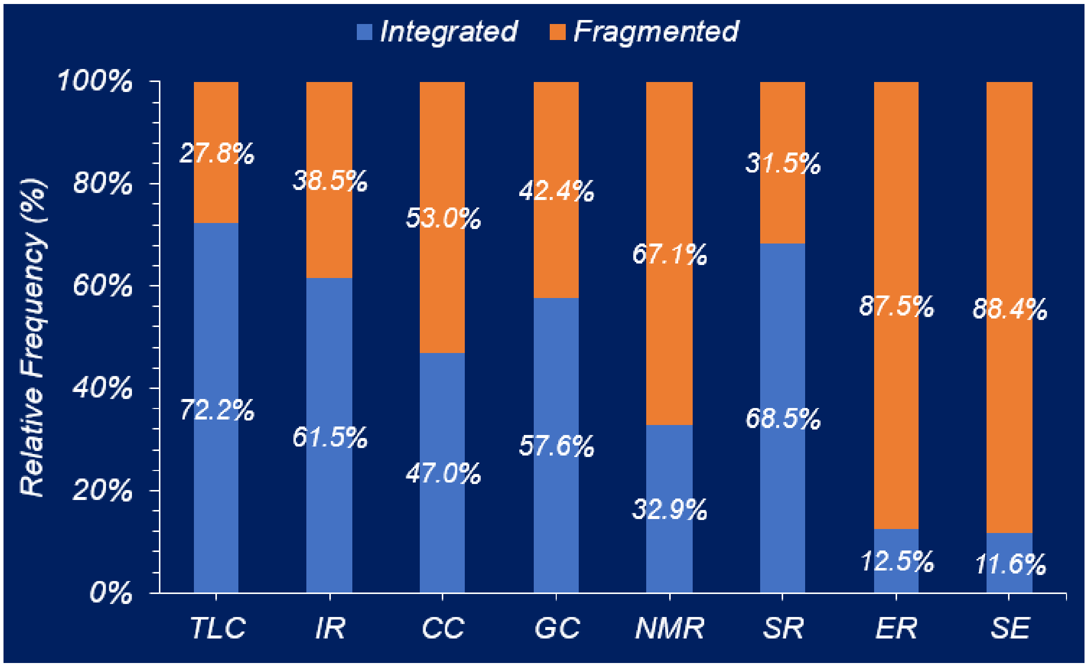


**Table 36** Relative frequencies of the domain-general alignment coding category for each of our eight experiments.

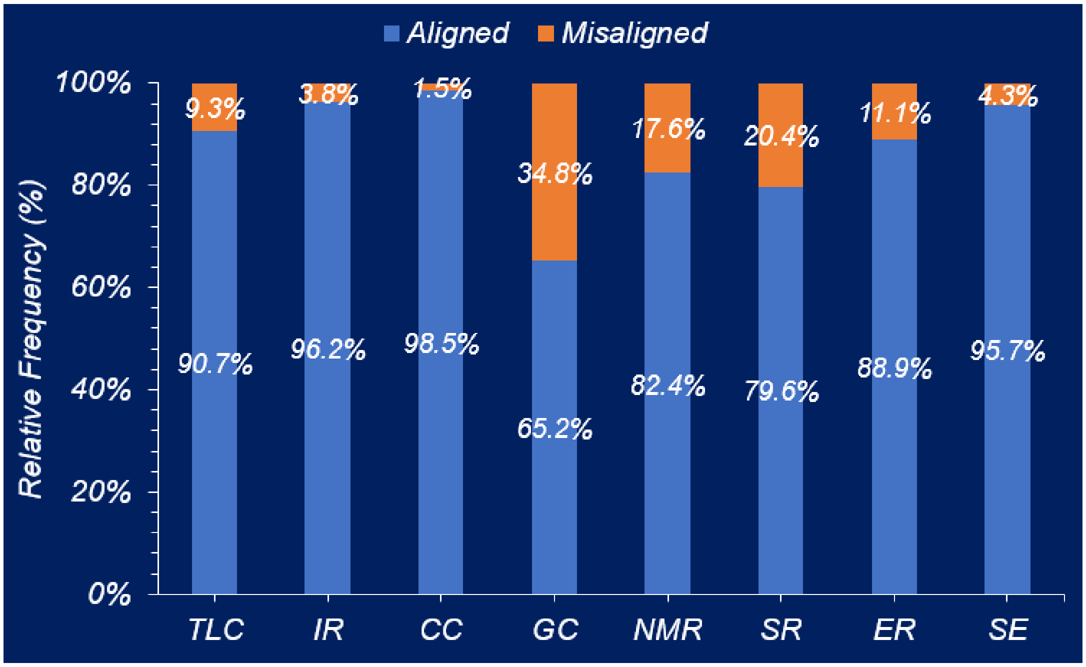
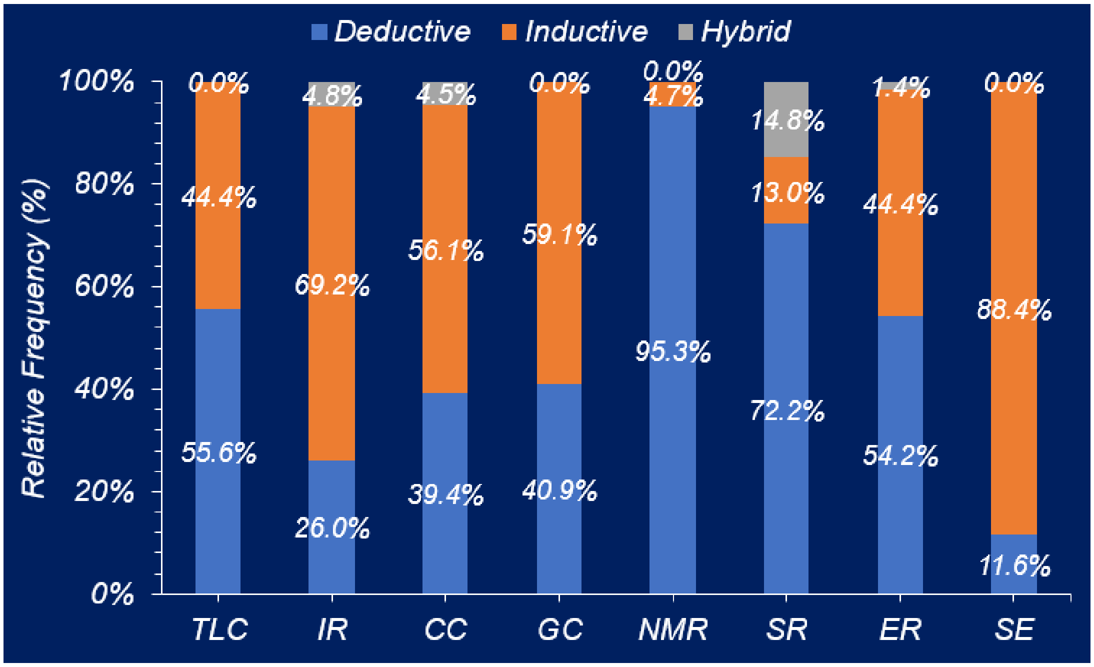


**Table 37** Relative frequencies of the domain-general approach to reasoning coding category for each of our eight experiments.



# Reference List

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 McNeill, K. L., (2010), A Learning Progression for Student Argumentation: Understanding Student Work and Designing Supportive Instructional Contexts, *Sci. Educ.*, **94**, 765-793.

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.

Bransford, J., Brown, A., and Cocking, R. (Eds.), (2000), *How people learn: Brain, mind, experience and school*, Washington, DC: National Academy Press.

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.

Çetin, P. S., (2021), Effectiveness of Inquiry Based Laboratory Instruction on Developing Secondary Students' Views on Scientific Inquiry, *J. Chem. Educ.*, **98**, 756-762.

Çetin, P. S. and Eymur, G., (2017), Developing Students' Scientific Writing and Presentation Skills through Argument-Driven Inquiry: An Exploratory Study, *J. Chem. Educ.*, **94**, 837-843.

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., (2015), Why Ask Why?, *J. Chem. Educ.*, **92**, 1273-1279.

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.

Crujeiras-Pérez, B. and Jiménez-Aleixandre, M. P., (2017), High school students' engagement in planning investigations: findings from a longitudinal study in Spain, *Chem. Educ. Res. Pract.*, **18**, 99-112.

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.

Grooms, J., Sampson, V., and Enderle, P., (2018), How concept familiarity and experience with scientific argumentation are related to the way groups participate in an episode of argumentation, *J. Res. Sci. Teach.*, **55**, 1264-1286.

Hammer, D., Elby, A., Scherr, R. E., and Redish, E. F., (2005), Resources, Framing, and Transfer. In Mestre, J. P. (Ed.), Transfer of Learning from a Modern Multidisciplinary Perspective, 89-119, Greenwich, CT: Information Age Publishing.

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.

Hogan, K. and Corey, C., (2001), Viewing classrooms as cultural contexts for fostering scientific literacy, *Anthropology & Education Quarterly*, **32** (*2*), 214-243.

Hosbein, K. N., Lower, M. A., and Walker, J. P., (2021), Tracking Student Argumentation Skills across General Chemistry through Argument-Driven Inquiry Using the Assessment of Scientific Argumentation in the Classroom Observation Protocol, *J. Chem. Educ.*, **98**, 1875-1887.

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.

Juntunen, M. K. and Aksela, M. K., (2014), Improving students' argumentation skills through a product life-cycle analysis project in chemistry education, *Chem. Educ. Res. Pract.*, **15**, 639-649.

Kadayifci, H., Atasoy, B., and Akkus, H., (2012), The correlation between the flaws students define in an argument and their creative and critical thinking abilities, *Soc. Behav. Sci.*, **47**, 802-806.

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.

Kelly, G. J. and Takao, A., (2002), Epistemic Levels in Argument: AN Analysis of University

Oceanography Students' Use of Evidence in Writing, *Sci. Educ.*, **86**, 314-342.

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.

Lemke, J. L., (1990), Talking science: Language, learning, and values, Norwood, NJ: Ablex.

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.

McNeill, K. L., Lizotte, D. J., Krajcik, J., and Marx, R. W., (2006), Supporting Students' Construction of Scientific Explanations by Fading Scaffolds in Instructional Materials, *J. Learn. Sci.*, **15**(2), 153-191.

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.

Newton, P., Driver, R., and Osborne, J., (1999), The place of argumentation in the pedagogy of school science, *Inter. J. Sci. Educ.*, **21** (*5*), 553-576.

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.

Park, J., Jung, D., Kim, G., Jun, J., and Nam, J., (2020), The Effects of Argument-Based Inquiry Activities on Elementary School Students' Claims and Evidence in Science Writing, *J. Kor. Chem. Soc.*, **64**(6), 389-400.

Petritis, S. J., Kelley, C., and Talanquer, V., (2021), Exploring the impact of the framing of a laboratory experiment on the nature of student argumentation, *Chem. Educ. Res. Pract.*, **22**, 105-121.

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. and Clark, D. B., (2008), Assessment of the Ways Students Generate Arguments in

Science Education: Current Perspectives and Recommendations for Future Directions, *Sci. Educ.*, **92**, 447-472.

Sampson, V. and Clark, D. B., (2009), The Impact of Collaboration on the Outcomes of Scientific Argumentation, *Sci. Educ.*, **93**, 448-484.

Sampson, V. and Clark, D. B., (2011), A Comparison of the Collaborative Scientific Argumentative Practices of Two High and Two Low Performing Groups, *Res. Sci. Educ.*, **41**, 63-97.

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.

Smith, C. L., Wiser, M., Anderson, C. W., and Krajcik, J., (2006), FOCUS ARTICLE: Implications of Research on Children’s Learning for Standards and Assessment: A Proposed Learning Progression for Matter and Atomic-Molecular Theory, Measurement, *Interdisciplinary Research and Perspectives*, **4**(1&2), 1-98.

Soysal, Y. and Yilmaz-Tuzun, O., (2021), Relationships Between Teacher Discursive Moves and Middle School Students’ Cognitive Contributions to Science Concepts, *Res. Sci. Educ.*, **51**(Suppl. 1), S325-S367.

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.

White, B. and Frederiksen, J., (1998), Inquiry, modeling, and metacognition: Making science accessible to all students, *Cognition and Instruction*, **16**, 3-118.

White, B. and Frederiksen, J., (2000), Metacognitive facilitation: An approach to making scientific inquiry accessible to all, In J. Minstrell & E. v. Zee (Eds.), *Inquiring into inquiry learning and teaching in science*, 331-370, Washington, DC: American Association for the Advancement of Science.

Wood, D., Bruner, J. S., and Ross, G., (1976), The role of tutoring in problem solving, *Journal of Child Psychology and Psychiatry*, **17**, 89-100.