How Assessments Influence the Conceptual Ideas Students Invoke for Explaining Chemical Phenomena

and

Modeling and Characterizing Epistemic Ideas for Teaching and Learning Chemistry

By

Kimberly S. DeGlopper

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN–MADISON

2024

Date of final oral examination: May 15, 2024

The dissertation is approved by the following members of the Final Oral Committee:
Ryan L. Stowe, Professor, Chemistry
Sam Pazicni, Professor, Chemistry
Teshik P. Yoon, Professor, Chemistry
Rosemary S. Russ, Professor, Curriculum and Instruction

Table of Contents

Acknowledgements

This work would not have been possible without the guidance, wisdom, and support from colleagues, friends, and family.

I am very grateful to Ryan Stowe for taking me on as his first graduate student and for mentoring me throughout my PhD journey. I appreciate his continued unwavering support, prompt communication, and patience during these challenging years. I also appreciate his sense of humor and tolerance of our office antics.

Thank you to the members of the Stowe group, past and present, who were always willing to discuss ideas with me and give me feedback on my projects. Thank you to Tom Kuborn, Adam Schafer, Cara Schwarz, Nicole Greco, Brie Bradshaw, Lindsay Wells, and Theodore Gierszal for making the office a fun, supportive, and entertaining place to work. Not everyone gets to work alongside their friends all day, but I did.

I would like to express my gratitude to my committee members: Rosemary Russ, Sam Pazicni, and Tehshik Yoon. Rosemary taught me so much about epistemology and that you do not have to be a boring person to study it. Sam's questions and feedback on my research always made me think more deeply about my assumptions and rationales. I am grateful to Tehshik for recognizing my early struggles and helping me find the right place for me in graduate school.

I owe Brian Esselman and Aubrey Ellison so many thanks for supporting me and mentoring me as a teaching assistant and as a graduate student. They gave me a job when I was feeling lost and insecure about my ability to succeed. They never doubted me, and their confidence and support enabled me to stay in graduate school. Thank you to Brian for pushing me to talk to Ryan about his research, for teaching me how to teach, for collaborating with me on several projects, and for advising me on my career and on life in general.

I want to thank Liana Lamont, Stephen Block, Mary Beth Anzovino, and Amanda Buchberger-Jones for the conversations on teaching and for their advice and encouragement as I prepare for the next stage of my career.

I have made many wonderful friends during my time in graduate school. In addition to those already mentioned, I want to thank Sam Wood and Ellorah James, Michelle Akana and Brett Schneider, Steven Chapman, Logan Vine, Robert Ward, and Kate Nicastri. I would not have made it through graduate school without them. Meghan Campbell and Eric Weeda were friends through both undergrad and grad school, and I look forward to the fun we will continue to have together. I am so grateful to Yukki Li for being my roommate and best friend for most of graduate school. Hopefully someday we will live close to each other again!

My therapist Jean is incredible. Without her wealth of knowledge and experience, I would not have made nearly as much progress in acquiring the mental and emotional tools I have relied upon throughout graduate school.

I will always have a special place in my heart for Jeff Johnson and the Johnson group at Hope College. I first discovered research in Jeff's lab and spent many wonderful summers there. Thank you to Jeff for giving me a great deal of independence while always being available to answer questions. More importantly, I greatly benefited from your wisdom about life. I was fortunate to overlap with many members of the Johnson lab during my time at Hope, and I have so many good memories with them. In short, the lab was my home, and I am forever grateful for the time I spent there.

My friends at Hope College were and continue to be supportive friends and science buddies. Stanna Dorn, Tiffany Oken, Kathryn Trentadue, Liz Ensink, Amanda Gibson, Lyndsy Miller, Dan

Clark, Chris Gager, and Sam VanderYacht. I wish I could have taken you all to graduate school with me, but I am grateful we have stayed connected.

Growing up, I was surrounded by a neighborhood that has become family to me. Thank you to Diane and David Tyler, Amy and Andy Levesque, Rita Blackmar, Adam Blackmar, Clayton and Jamie Blackmar, Joe Levesque, Steve Levesque, Nick Hayes, Kelly Wagner, and Amber Townsend for being a crazy fun group to return home to.

I am thankful for my lifelong friends Caitlin Vanna, Mallory Dreyer, Amber Townsend, and Kelly Wagner. Even though we often go months without talking, I know I can call them anytime, and we can pick up like no time has passed. That is true friendship.

It would not be a proper Stowe group thesis if I did not thank the pets that brought me happiness whenever I saw them. Thank you to Bursley, Nala, Abbot, and Molly for all their cuteness and affection.

My family is my reason for being. Their unconditional love and support mean more to me than I can possible say. Thank you to my brother-in-law Grant Gebben for giving me the biggest and loudest hugs I have ever received. I will forever be grateful to my mom and dad for always believing in me and for always allowing me to choose my path, even when it led me further away from home. My sisters, Annie and Jenna, are my closest friends. I have never been lonely because I don’t remember a time without them in my life. My nephew Will brings me so much joy. I look forward to teaching him chemistry as he grows older.

Abstract

The work presented in the following chapters was undertaken with the goal of understanding and improving education in organic chemistry. Initial work, described in Chapters 1 and 2, focuses on students' ability to explain and rationalize the outcomes of organic chemistry reactions. Chapter 1 explores the relationship between assessment emphasis and the structure-energy connections students made when asked to explain the outcome of a hydrobromination reaction. Using a mixed methods study design, an association was found between the kinds of tasks students were given on assessments and the ideas they utilized in their written explanations on a separate researcher-authored assessment. The relationship between assessment emphasis and conceptual ideas was probed qualitatively in Chapter 2. Students were interviewed as they proposed electron-pushing mechanisms and predicted the products for familiar and unfamiliar reactions. The ideas they drew on in their reasoning were found to be similar to the core ideas around which their course was designed, indicating that the students had internalized the utility of these ideas for problem-solving.

More recently, the focus of the work has shifted toward understanding the epistemic aspect of organic chemistry education. Chapter 2 describes our first foray into epistemology; in addition to characterizing the conceptual resources students invoked, we characterized some of the epistemic resources they utilized. This allowed us to differentiate between moments when students were recalling explanations versus moments when they were constructing them in the moment by connecting different bits of prior knowledge. Chapter 3 presents a critical review of how the field of chemistry education has studied and modeled undergraduate students' epistemic cognition and in particular, draws attention to the implications of evaluating students' epistemic cognition in a hierarchical manner that does not attend to context. In Chapter 4, we compare two ways of modeling epistemic cognition and discuss their affordances and limitations with regard to

interpreting interview data on organic chemistry instructor’s thinking for teaching and learning chemistry.

Appendix A presents some of the ongoing work aimed at understanding the influence of various aspects of organic chemistry courses have on students’ epistemic cognition.

Appendices B–E contain reprints of the published works to which I have contributed in the formats of their respective journals. Appendices B and C are reprints of Chapters 1 and 4, respectively. Appendices D and E are chemistry education works to which I contributed but was not the primary author.

Many people contributed to the work summarized above and presented in the following chapters. All of this work was conducted under the supervision of Dr. Ryan L. Stowe. In addition, the following people, listed alphabetically, contributed to various parts of the work: Niall J. Ellias, Dr. Aubrey J. Ellison, Dr. Brian J. Esselman, Dr. Nicholas J. Hill, Dr. Jeffrey D. Martell, Dr. Vanessa R. Ralph, Dr. Rosemary S. Russ, Cara E. Schwarz, and Prayas Sutar. At the beginning of each chapter, the specific people involved in the work are noted.

Chapter 1

The Impact of Assessment Emphasis on Organic Students' Explanations for an Alkene Addition Reaction

This work was conducted in collaboration with Cara E. Schwarz, Niall J. Ellias, and Ryan L. Stowe.

Abstract

To potentially engage students in "doing organic chemistry," organic chemistry courses should foreground weaving together structure- and energy-related ideas to construct causal accounts for phenomena. Here, we investigate whether enrolling in an organic chemistry course that places substantial emphasis (~50% of total points) on explaining phenomena on exams is associated with more productively justifying the outcome of a chemical process. This work occurred in the context of three learning environments that differed principally by assessment emphasis. The "explanation focused" course allotted 40-66% of points on exams to explaining phenomena while the other two enactments placed much less emphasis on connecting big ideas to how and why chemical processes occur (~0-25% of total points). Students enrolled in each course were given a prompt which asked them to draw mechanisms for a hydrobromination reaction and subsequently justify the regiochemical outcome of that reaction. We described student responses by noting the connections made between structure (of reactants, intermediates, transition states, or products) and energy. Most students described how charge or electron delocalization impacted the relative energies of the two possible intermediates or transition states. Other explanations invoked steric repulsion or differences in relative energy due to degree of carbocation substitution. Examination of the association between learning environment enrollment and explanation code distribution revealed that students who enrolled in two semesters of an explanation-focused course were substantially *less* likely to leave out charge delocalization in their explanations while students who were never enrolled in the explanation-focused learning environment were substantially *more*

likely to leave out charge delocalization. These findings suggest that changing what is assessed to better align with “doing organic chemistry” may be a promising avenue for reform.

Introduction

Organic chemists seek to understand and manipulate reactions with the goal of efficiently and selectively synthesizing molecules that possess properties suitable for particular functions (e.g., inhibition of disease-relevant macromolecules, impact resistance, conductivity). Toward that end, many organic chemists are engaged in designing new reaction systems, which involves predicting and rationalizing the possible outcomes of reactions based on knowledge of how chemical structure influences reactivity. Take, for example, the development of enantioselective organocatalysis, which recently won the Nobel prize in chemistry (The Nobel Committee for Chemistry, 2021). Enantioselectivity is achieved because the chiral catalyst alters the transition state structures leading to the two possible enantiomeric products and in turn, the energy difference between the transition states. In some cases, the catalyst lowers the energy of one transition state relative to the other via noncovalent interactions while in others, the catalyst raises the relative energy of one through increased steric repulsion. Constructing explanatory accounts of how altering system parameters might change reaction outcomes allows chemists to purposefully plan experiments directed at reaction optimization (Manz et al., 2020). Stated succinctly, figuring out how a reaction is likely to proceed via connecting chemical structure (*e.g.*, of intermediates, transition states, or products) and relative energy is a fundamental part of “doing chemistry.”

There are compelling reasons to frame chemistry learning environments as opportunities to “do chemistry.” Engaging learners in creating, refining, and communicating knowledge about how and why the world works has the potential to focus the classroom community on “figuring out” aspects of their existence rather than “learning about” disaggregated skills and facts (Schwarz et al., 2017). Emphasis on “figuring out” makes explicit that science activities and content knowledge are *always* related in the practice of science (National Research Council, 2012). Specialized skills, such as drawing electron-pushing arrows, have no inherent meaning; one can learn the rules of “mechanism drawing” without ever considering donor-acceptor interactions that might occur in a system. Likewise, one may memorize the definition for “π conjugation” or “asymmetric induction” without ever being able to use these ideas to articulate why something happened (and why other things did not). Skills and knowledge gain meaning when purposefully woven together with the aim of understanding how the world works or designing a solution to a pressing problem.

Unfortunately, it is fair to say that “doing science” is not a central focus of many organic chemistry courses. Indeed, it is common for “correct” application of a skill or recall of a reaction product to be allotted far more points on assessments than construction of causal accounts for phenomena (Stowe & Cooper, 2017). This is troublesome since one may readily apply the “rules of the game” to enact a skill without understanding the chemical system related to that skill. Perhaps the best example of an explanatory tool that is often reduced to a skill is the electron-pushing formalism (EPF). Organic chemists make use of the EPF to reason about how the movement of electrons transforms starting material into product. However, several studies have demonstrated that students commonly “decorate” starting materials and intermediates with arrows rather than using curved arrow mechanisms as predictive tools (Grove et al., 2012; Houchlei et al., 2021). When investigating how students reason about reaction mechanisms, Graulich and coworkers found that many students relied on teleological reasoning and used the favorability of a subsequent mechanistic step to justify proposing an earlier one, despite the molecule having no way of knowing a subsequent step would be productive (Caspari, Weinrich, et al., 2018). Even first-year graduate students often proposed mechanisms based on what would get them from the starting material representation to the product representation rather than based on arguments grounded in stabilization of charge or electrophilicity (Bhattacharyya & Bodner, 2005). Potential energy surfaces constitute another potentially powerful model that may not hold meaning for students (Popova & Bretz, 2018a; Popova & Bretz, 2018b; Lamichhane et al., 2018). Importantly, we argue that a tendency toward recall and decontextualized skill application is the result of inappropriate learning environment design, not a deficiency on the part of enrolled students. Students in the aforementioned studies were often quite adept at the performances emphasized and rewarded in their courses. Therefore, our focus in this study is not on documenting students' misconceptions but on examining how learning environments may be designed to support students in doing chemistry.

When we say a performance was "emphasized and rewarded" in this manuscript, we mean a task was assessed. Assessments serve two important roles in learning environments. First, they convey strong messages to students as to what is important in the course and in the discipline in general (Momsen et al., 2013; Scouller, 1998; Scouller & Prosser, 1994; Snyder, 1973; Entwistle, 1991; Crooks, 1988). Students who are frequently asked to explain why a phenomenon occurs on assessments will learn to do so, as Crandell et al. (2020) observed when comparing the explanations of students enrolled in a transformed course to those of students enrolled in a traditional organic chemistry course After a year enrolled the transformed course Organic Chemistry, Life, the Universe and Everything (or OCLUE), which prioritizes reasoning with core ideas, students were more likely to provide causal explanations for an S_N2 reaction than students in the traditional organic chemistry courses (Cooper et al., 2019). The second role of assessments is to elicit evidence of what students know and can do. Ideally, evidence elicited by assessments should be used to inform instructors and to help students chart their learning priorities. The sorts of inferences that can be made from what students write or say are powerfully influenced by the structure of the prompt as well as the model of mind adopted by the instructor or researcher (more on that shortly). Responses to an assessment item that simply asks for a claim (such as “circle the most acidic molecule”) provide no evidence of how students arrived at that claim. We argue that learning environments which foreground “doing chemistry” should consistently engage learners in constructing and critiquing causal accounts for phenomena (Stowe et al., 2021). This entails moving beyond asking solely for claims and toward expecting reasonable connections of big ideas (*e.g.*, energy, donor-acceptor interactions) to phenomena.

Despite the influential role assessments play in signaling to students “what counts,” we are aware of no scholarship exploring the impact of changing assessment emphasis (and little else) on student explanations for phenomena. Most assessment reform efforts in chemistry education have been coupled with curricular transformations. For example, the general chemistry course Chemistry, Life, the Universe and Everything (CLUE) integrates assessments focused on mechanisms underpinning phenomena and also reorganizes the curriculum around scaffolded sequences of big ideas (Cooper & Klymkowsky, 2013). Likewise, Chemical Thinking represents an overhaul of both curricular sequencing and assessment emphasis (Talanquer & Pollard, 2010). With regard to organic chemistry education, both OCLUE and Flynn’s mechanisms before reactions courses have simultaneously reformed their curricula and assessments (Cooper et al., 2019; Flynn & Ogilvie, 2015). The context for this study is thus a bit unique: organic chemistry courses at a research-intensive Midwestern university follow a unified curricular sequence informed by a commercial textbook but (as we will see) emphasize different sorts of tasks on assessments. This presented us the opportunity to probe the relationship between assessment emphasis and students’ justifications for agreeing or disagreeing with a given reaction outcome claim.

Theoretical Framework

Before describing the study in detail, it is important to clearly articulate the model of learning we are using to infer aspects of student cognition. The model of learning one adopts influences the conclusions and implications of a study by bounding the study designs that seem reasonable and the sorts of inferences one can draw from student response data (National Research Council, 2001). To aid readers in understanding the constraints and affordances of our study design, we provide here a brief description of the model of learning we adopt and discuss in broad terms how it influenced our view on the roles of instructors and students.

Our study design and implementation was informed by the resources model of cognition put forth by Hammer et al. (2005), which draws on the knowledge-in-pieces framework from diSessa (1988). This model of mind assumes that knowledge exists in pieces that are activated in the moment for a specific purpose. These small-grain “resources,” such as the primitive “more means more” (Ohm’s p-prim), are not inherently “correct” or “incorrect” but may be more or less productive depending on context (diSessa, 1988). Connections between knowledge elements elicited by a given assessment or instructional scenario are not assumed to be stable across time and place (although some may be). In contrast, a theory-theory model of learning assumes stable, coherent knowledge structures which, if incorrect, are referred to in the literature as “misconceptions,” “alternative conceptions,” or “naïve theories” (diSessa, 2006). Students possessing theory-like knowledge would be expected to offer consistent responses to questions that require use of the same naïve theory to answer. In fact, studies have shown students’ responses are often contradictory, depending on the contexts of the questions and the ways in which they are worded (Cooper et al., 2013; Rosenberg et al., 2006). This inconsistency can be better accounted for by a resources model where activation of knowledge elements is dependent upon the specific context.

These two contrasting models of cognition result in very different roles for the instructor. An instructor with a resources model of cognition would interpret student responses as momentary coalescences of knowledge elements and seek to recognize, reward, and build upon productive resource use. Adopting a resources view of learning *requires* an instructor to attend to the substance of student thinking, even if the "incorrect" vocabulary is used (Robertson et al., 2015). By contrast, an instructor with a theory-theory model of cognition would interpret "incorrect" student responses as indicative of misconceptions and seek to replace these with the "correct" conceptions. Some scholars propose a process of rationally challenging and replacing student "misconceptions" in which students are shown the inadequacies of their naïve theory and offered a canonical theory as a more useful replacement (Posner et al., 1982). Needless to say, there is no convincing evidence that a robust and useful understanding of fundamental chemistry ideas comes about through a series of rational "paradigm shifts." These models of cognition also have important implications for how we view students. From a resources perspective, students possess potentially productive knowledge elements that can be used to further their own learning. Theory-theory views align more with a deficit perspective on students, in which their prior knowledge is often problematic and needs to be rooted out rather than utilized.

We should note that "knowledge-in-pieces" and "theory-theory" models of mind represent two ends of a continuum rather than mutually exclusive theories of cognition (Brown, 2014). Resources that are consistently activated together may approach theory-like stability. However, because we are not characterizing students' knowledge over time or across multiple prompts, we do not intend to make any claims concerning consistency or stability. A resources model of cognition is thus appropriate for our aim of identifying the productive connections students make in the moment when explaining a reaction outcome.

Study Goals

As part of a broader, ongoing effort to improve introductory organic chemistry courses at a large, research-intensive, Midwestern university, we aim to characterize the knowledge elements that students call to mind and connect when asked to reason about phenomena of interest to organic chemists. In this study, we asked students enrolled in three different learning environments (described below) to consider a benzylic alkene addition reaction, construct electron-pushing mechanisms depicting formation of two possible products, and evaluate a claim regarding which would be the major product. Specifically, we sought to answer the following questions:

1) What knowledge elements do students activate and connect when asked to explain the outcome of a benzylic alkene addition reaction?
2) How do the electron-pushing mechanisms students draw relate to the explanations they provide?
3) How does the intellectual work emphasized and rewarded on assessments given in each learning environment relate to the structure-energy connections invoked in student explanations?

Our focal phenomenon (addition of HBr to a benzylic alkene) was chosen due to the importance of π conjugation in stabilizing the high energy intermediate leading to the major product. Constructive overlap of a series of p-orbitals explains a wide variety of important chemical and biochemical phenomena (e.g., protein 3D structure, stability of drug-like aromatic compounds) and should therefore be prominent among the powerful explanatory ideas students use for sensemaking.

Methods

Study Context

This study was conducted at a large, public, research-intensive university in the Midwest. The courses involved in this study were the on-sequence first and second semester of introductory organic chemistry (OChem I and OChem II) enacted during the Fall 2020-Spring 2021 academic year. These courses serve chemistry majors, biology majors, and chemical engineering majors; most students enrolled intend to pursue careers in the health fields. Both courses were administered entirely online due to the COVID-19 pandemic. As the data collection and analysis described here was undertaken primarily for the purposes of program improvement, IRB approval was not required. The assessment prompts used as the outcome measure for this work were developed as part of another, IRB-approved study (ID 2020-0684). Accordingly, participant consent was obtained for the response-process interviews conducted during instrument development.

At this institution, organic instructors agree to use the same textbook and cover roughly the same set of content in their organic chemistry courses. However, each instructor has the freedom to operate independently and full control over how they teach and assess students. Some instructors voluntarily choose to team up and write and administer common assessments. During the course of this study, four unique learning environments were enacted. We defined a learning environment to include all course sections that were taught by the same instructional team and had the same structure, assessment formats, and assessment emphases (see below for details on assessment emphases). Two learning environments (A and B) extended the full two-semester sequence. That is, Learning Environments A and B are characterized by one instructional team enacting both

OChem I and OChem II with a consistent structure and assessment emphasis. Learning Environment C was an enactment of OChem I while Learning Environment D was an enactment of OChem II (Fig. 1.1). Note that a student might not be enrolled in the same learning environment for both courses. For example, a student enrolled in Learning Environment A for OChem I could be enrolled in any of the three learning environments for OChem II. All learning environments proceeded through roughly the same sequence of topics for each course, and they all featured a mix of pre-recorded lecture videos, synchronous problem-solving sessions with the instructor, and optional TA-led synchronous discussion sections. The sixth edition of Marc Loudon's *Organic Chemistry* was used as the textbook for all learning environments under study (Loudon & Parise, 2016).

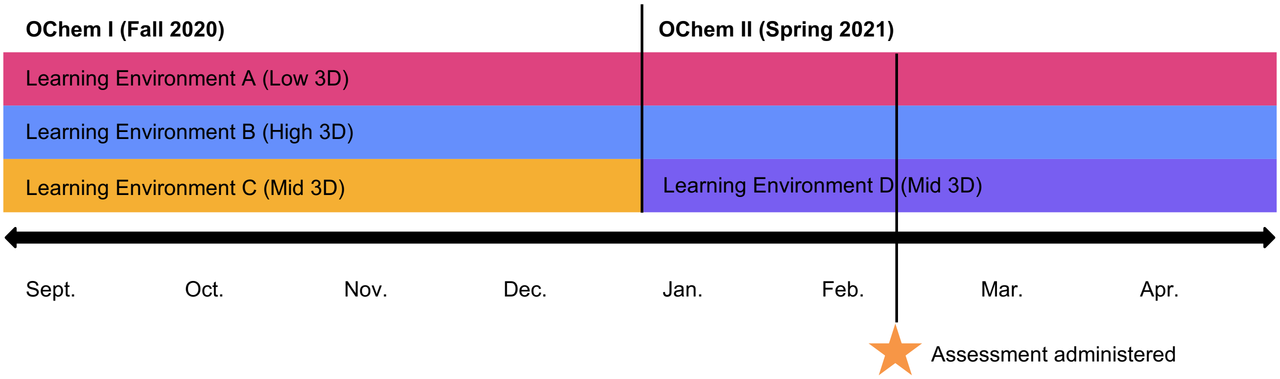


Figure 1.1. Timeline showing when each learning environment was operating during the Fall 2020-Spring 2021 academic year and when the assessment item used in this study was administered.

Since assessments convey strong messages to students regarding what sorts of performances are valued, the learning environments were characterized according to their assessment emphasis. The 3-Dimensional Learning Assessment Protocol (3D-LAP) was used to identify assessment items with the potential to elicit evidence of engagement with scientific practices, disciplinary core ideas, and crosscutting concepts (*i.e.*, 3D learning; Laverty et al., 2016). Details of our 3D-LAP coding process may be found in the SI. We agree with modern science reform efforts that K-16 STEM learning environments should support integration of activities scientists do (i.e., science practices) and content knowledge (*i.e.*, core ideas; National Research Council, 2012). Use of the 3D-LAP here is a concrete reflection of this commitment; LAP readouts help us capture the extent to which instructors emphasize and reward knowledge-in-use on the exams they give. Items which fulfill LAP criteria for potentially eliciting evidence of 3D performances require students to articulate reasoning underpinning claims made while non-3D items typically only require an answer in the form of a claim (e.g., product of a reaction, most acidic molecule) without the accompanying reasoning. It is important to emphasize that items denoted as 3D in this study have the *potential* to elicit evidence of knowledge-in-use. Meeting 3D-LAP criteria does not guarantee a given item *will* elicit evidence of engagement in a 3D performance. For examples of 3D prompts, we refer the reader to our recent publication (Stowe et al., 2020). The supporting information accompanying this publication contains dozens of examples of 3D prompts administered as homework and on exams.

In Figure 1.2, the percentage of points dedicated to assessment items with the potential to elicit evidence of 3D learning is shown for each exam given in each of the learning environments. Exams 1-3 were given throughout the semester while Exam 4 was given as the final exam and was worth more points. For OChem I, Learning Environment B gave exams with 42-59% of the points dedicated to 3D items while Learning Environment C gave exams with 18-46% of the points dedicated to 3D items. None of the exams given in Learning Environment A contained a 3D assessment item. A similar trend was observed for OChem II. Learning Environment B gave exams with 40-66% of the points dedicated to 3D items, Learning Environment D gave exams with 0-15% of the points dedicated to 3D items, and none of the exams given in Learning Environment A contained a 3D assessment item. The substantial differences among learning environments prompted us to further investigate the link between assessment emphases and student-constructed explanations for phenomena. To aid readers, we will refer to Learning Environments A and B as low and high 3D learning environments, respectively, and Learning Environments C and D as mid 3D learning environments throughout the remainder of the paper.

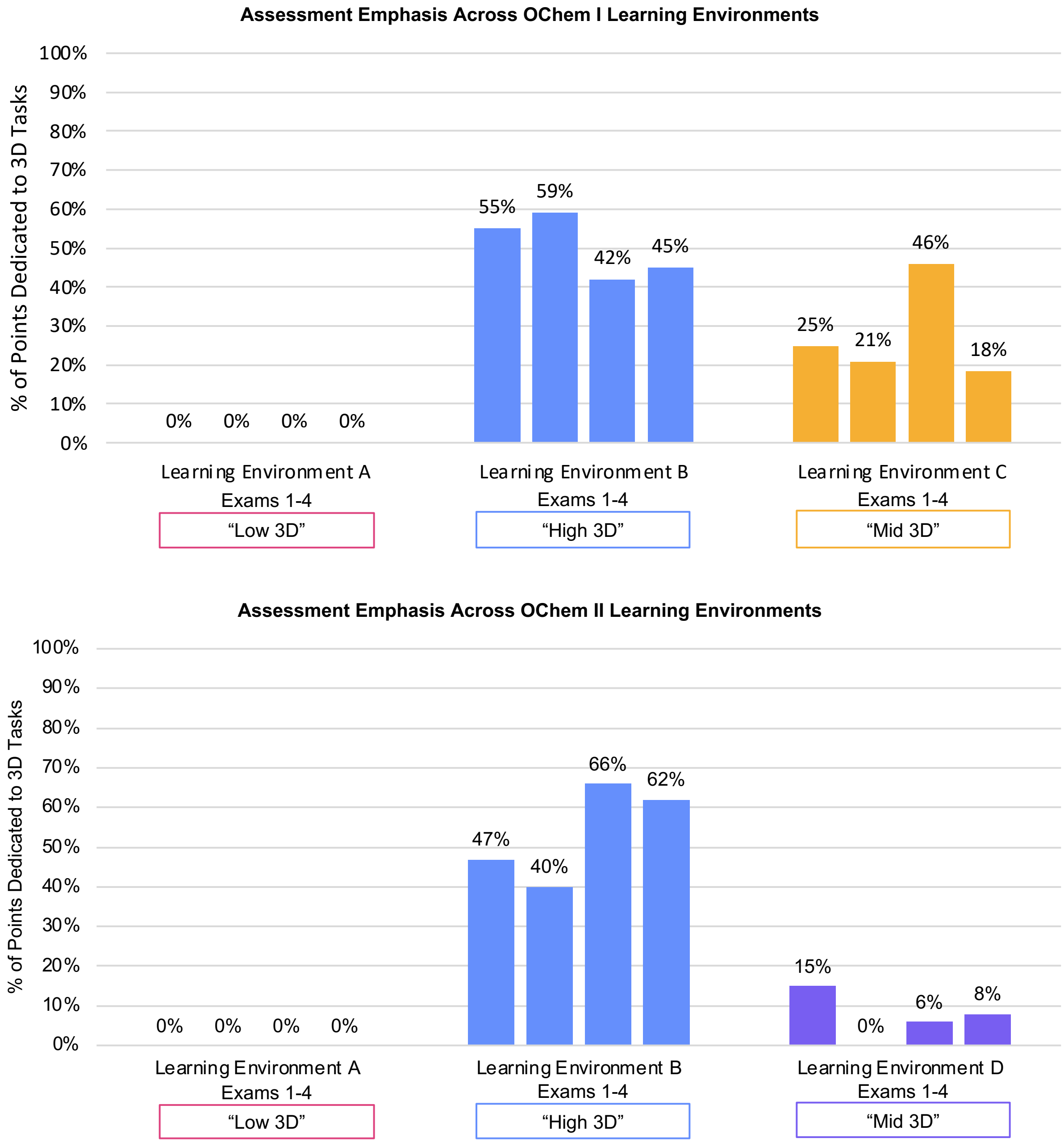
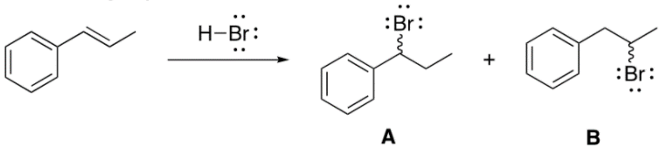


Figure 1.2. Percentage of points dedicated to 3D assessment items on exams given in each of the three learning environments during OChem I (top) and OChem II (bottom). Based on the data shown here, we designated Learning Environment A as "low 3D," Learning Environment B as "high 3D," and Learning Environments C and D as "mid 3D.

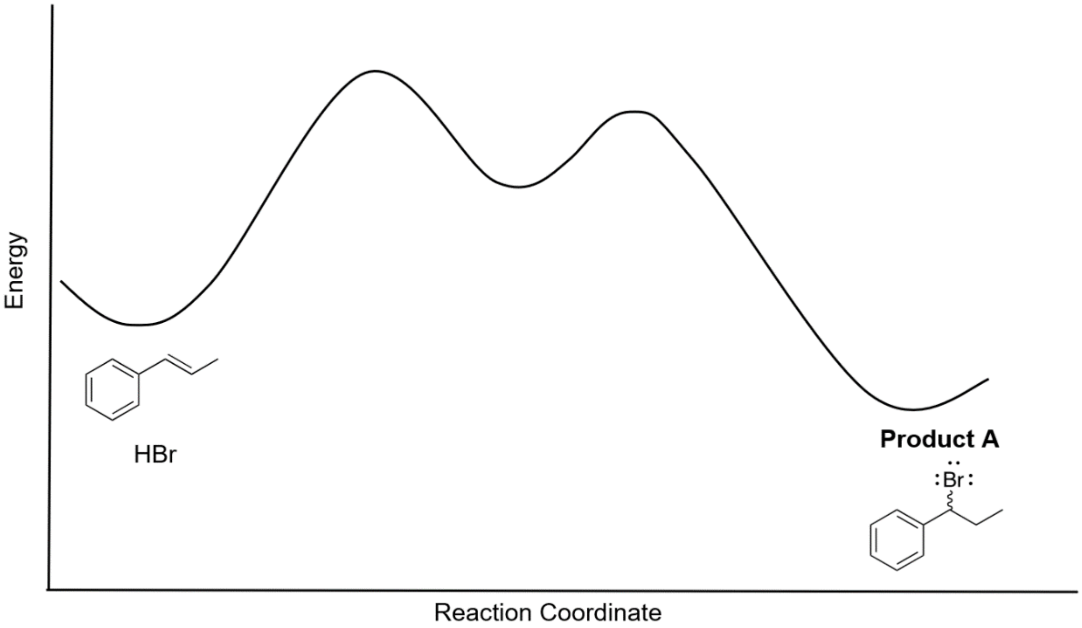
Instrument

The assessment item used in this study was created using evidence-centered design (ECD) (Mislevy & Haertel, 2006). Accordingly, a precise expectation for what organic-enrolled students should know and be able to do guided task design. The performance expectation our task was built to assess was: “Construct and use an electron pushing mechanism and/or a reaction energy profile to evaluate the validity of claims as to the outcome of a chemical process.” This performance expectation does not solely reflect what the authors deem valuable but reflects the consensus of all instructors who teach organic chemistry at the focal institution. Following unanimous approval of this performance expectation (along with several others), the authors designed an assessment to elicit evidence of student engagement in the specified performance (Fig. 1.3). Response-process validity was established via cognitive interviews conducted with eight consenting students following IRB approval (Arjoon et al., 2013). The wording was slightly altered in response to these interviews to clarify the intent of the prompt.

Consider the reaction between hydrobromic acid (HBr) and the alkene shown below. You read a claim that **Product B** is the major product of this reaction.



1. Draw a mechanism for the reaction above that leads to **Product A**.
2. Draw a mechanism for the reaction above that leads to **Product B**.
3. The potential energy surface for the pathway leading from reactants to **Product A** is drawn below. Using a dashed line, draw the potential energy surface for the pathway leading from reactants to **Product B** on the same axes. Label all intermediates and products.



4. Would you expect **Product B** to be the major product of this reaction? Explain your answer using the mechanisms and potential energy surface you drew in parts 1–3.

Figure 1.3. Assessment item used to elicit evidence of how students connect structure- and energy-related ideas to rationalize the outcome of a benzylic alkene addition reaction.

We chose to center the assessment item on a benzylic alkene addition reaction because, to explain its outcome, students need to grapple with how π conjugation influences stability and how the relative energies of species present at various points in the reaction influence the process outcome. Attending to how charge delocalization impacts the stability of species in a reaction system is key to predicting and explaining the outcomes of a wide variety of reactions in organic chemistry. This prompt contained four parts. Parts 1 and 2 asked students to provide electron-pushing mechanisms that would account for formation of Product A and Product B, respectively. In Part 3, students were given a potential energy surface with a curve depicting formation of Product A. They were asked to add a second curve representing formation of Product B. Responses to this part of the prompt were not analyzed in this study for two reasons. First, the different learning environments varied considerably in how much students were expected to construct and use potential energy surfaces whereas all of them emphasized electron-pushing mechanisms. Second, a cursory examination of responses suggested that we would not obtain much additional information beyond what the mechanisms and explanations provided. In the final part of the prompt, students were asked to use the mechanisms and potential energy surface they drew to state whether they agreed with the claim that Product B would be the major product and to explain their reasoning.

A complete, canonical answer to the prompt shown in Figure 1.3 is provided in the supporting information (Fig. 1.S1), but we include here an example of how one might productively connect knowledge of structure and energy to explain the outcome of the reaction. This reaction is kinetically-controlled (irreversible), so the product formed most quickly (i.e., the reaction path with the lowest energy barrier to the rate determining step) will predominate. Formation of a carbocation intermediate is the rate-determining step en route to both possible products. We can use the relative energies of the carbocations as a proxy for the relative energies of the transition states for the rate determining step of each process (Hammond's Postulate). As the starting system is the same for both reaction paths, the path that proceeds through the higher energy carbocation would be expected to have the higher energy barrier to the rate determining step and thus be slower than the competing path. We can determine the relative energies of the intermediates by comparing their structures. The carbocation leading to Product A (Carbocation A) is stabilized by π-conjugation with the neighboring aromatic ring, resulting in delocalization of the positive charge. The carbocation leading to Product B (Carbocation B) is not in conjugation with the aromatic ring so it is less stable (i.e., higher in energy). Since Carbocation A is lower in energy than Carbocation B, Transition State A should be lower in energy than Transition State B, which means that Product A should be the major product. The claim in the prompt is thus incorrect. In short, the relative energies of the first transition states determine the reaction outcome, and examination of structural features that act to stabilize these transition states (e.g., π conjugation) allows for prediction of these relative energies.

Data Collection and Reduction

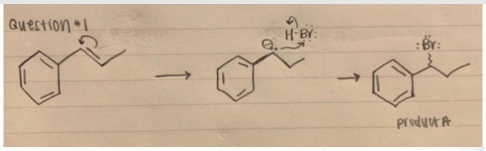
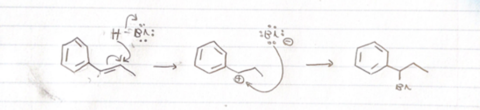
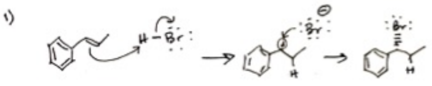
Permission to administer this assessment item was obtained from each professor. The assessment item was given in OChem II as a stand-alone homework assignment; a small amount of course credit was given to students for completing the assignment. This assessment was administered approximately halfway through the semester, directly after each learning environment had covered the chapter on the reactivity of benzylic systems. Students were given a week to complete the assignment. Student explanations from each learning environment were exported from the online learning management system into a spreadsheet and subsequently de-identified and randomized. 100 randomly selected student responses from each learning environment were compiled for the analyses described in the next section. Mechanism drawings were exported and renamed using the random ID that corresponded to the explanation.

Characterization of Student-Constructed Mechanisms

Each student submitted two mechanisms, one depicting formation of Product A and one depicting formation of Product B. These mechanisms were coded separately using a three-part coding scheme informed by prior work conducted by Grove et al. (2012), Crandell et al. (2020), and

Houchlei et al. (2021) in their studies on students' mechanistic reasoning. Responses with canonically correct arrows and intermediate structures were coded as "3," responses with canonically correct arrows or correct intermediate structures were coded as "2," and responses with neither canonically correct arrows nor correct intermediate structures were coded as "1" (Table 1.1). Note that we chose to ignore missing formal charges on bromide. All mechanisms were coded independently by the first three authors, and inter-rater agreement was calculated using Fleiss' Kappa (Fleiss, 1971). A value of 0.90 was obtained, indicating high agreement (McHugh, 2012). Any discrepancies were resolved, resulting in consensus codes for the entire data set. For the chi-square tests, codes "1" and "2" were collapsed because of their low counts, resulting in a binary mechanism variable (*i.e.*, incorrect and correct). We also elected to only use mechanism codes corresponding to formation of Product A, since A is the major product and 89% of responses earned the same code for both mechanism drawing prompts.

Table 1.1. Descriptions of mechanism codes and examples of students' responses

Code	Description	Examples
1	Incorrect arrows and incorrect intermediates	
2	Correct arrows or correct intermediates	
3	Correct arrows and correct intermediates	

Description of Productive Structure-Energy Connections in Student Explanations

Our analysis of student-constructed explanations focused on describing how ideas related to molecular structure and energy were productively activated and connected in the context of the prompt we administered. This analytic focus was borne of recognition that most phenomena of interest to organic chemists (and hopefully organic chemistry students) can be understood by connecting the structure of entities in a system (e.g., reactants, products, transition states, intermediates) to energetic changes that occur as these entities interact. For example, the energy barrier to produce a secondary non-conjugated carbocation from the benzylic alkene shown in Figure 1.3 is substantially higher than the corresponding barrier to produce a secondary benzylic carbocation due to the ability of the benzylic carbocation to delocalize charge via π conjugation. The central importance of connecting structural to energetic accounts in reasoning about organic chemistry phenomena has underpinned past scholarship by Caspari, Kranz, et al. (2018) and Bodé et al. (2019), as well as contributions by Goodwin on the philosophy of organic chemistry (Goodwin, 2003; Goodwin, 2008).

To describe invocation of structure- and energy- related knowledge elements by students responding to our task, the first three authors read through the responses and noted patterns in the structural and energetic features discussed. Comparing and contrasting these patterns allowed the team to revise, combine, and collapse descriptive codes in order that each code describes a distinct reasoning pattern. The consensus coding scheme that emerged from this dialogue can be found in Table 1.2. Explanations grounded in structural features other than π conjugation were coded as “0” as were responses that lacked both structural and energetic components. Responses that described only π conjugation or the energies of the transition states or intermediates were given a “1.” Responses that described π conjugation and related it to the energy of the intermediate were given a “2” while responses that described π conjugation and related it to the transition state energy were given a “3.” Both “2” and “3” responses represent productive connection of structural and energetic ideas to explain the outcome of the focal reaction. Note that we conceive of these codes as descriptors representing activation of knowledge elements in-the-moment, rather than judgements of durable “understanding” or “misunderstanding.”

Table 1.2. Descriptions and Examples of Explanation Codes for RQ2 and RQ3.

Code	Description	Example
0	Does not describe π conjugation or energy OR attributes energy differences to irrelevant structure feature (e.g., steric repulsion, carbocation substitution)	*I would expect Product B to be the major product in this reaction as it has less steric interaction with the attached benzene ring. Because I predicted less steric interaction, the energy threshold to reach Product B's intermediates and end product are lower than Product A's, so Product B takes less energy to create.*
1	Describes π conjugation and does not connect to energy or connects to product energy OR describes energy of intermediate and/or transition state without relating to structure	*No, I would expect A to be the major product of the reaction because it involves a more stable carbocation intermediate that which is lower energy and easy to form. Product B has an additional intermediate due to a rearrangement which is higher energy and less stable.*
2	Describes π conjugation and relates to intermediate energy	*Product B is not the major product because its higher in energy than Product A. A is more favored because the positive charge on the carbon can delocalize within the pi system of the ring which has stabilizing effects and lowers its energy.*
3	Describes π conjugation and relates to transition state energy	*No, I would not expect Product B to be the major product of this reaction. The intermediate leading to the formation of product B is less stable as the positive charge is not delocalized into the pi system by being at a benzylic position. By Hammonds postulate, the transition state energy would then be higher and product B reaction would happen slower making product A the major product instead.*

Reliably bounding each of the bins described previously involved several rounds of joint coding of responses, discussion of inconsistencies in this coding, and revision of the coding scheme. One outcome of this discussion was the decision to treat descriptions of stability as synonymous with descriptions of energy (i.e., a more stable intermediate is a lower energy intermediate.) This was done because many organic chemists use these terms interchangeably, including the organic instructors who taught the courses included in our study. We also determined that simply stating a species was benzylic did not count as describing π conjugation and that vague descriptions of energy, such as “energy pathway” were assumed to refer to the intermediate energy rather than transition state energy. Mention of product energy was ignored as this is not relevant to determining the outcome of this reaction. Our iterative, reflective coding process gave us confidence that the codebook was sufficiently detailed to characterize the data. In the end, the entire dataset was jointly coded by the first three authors, and consensus was reached on the code best describing each response. A more detailed description of the coding process is available in the SI.

Toward a Comprehensive Description of Structure- and Energy- Ideas Embedded in Student Explanations

After completing the coding process described previously, we recognized that many of the responses coded as “0” contained evidence of knowledge elements that are productive in different contexts. For example, the notion that steric repulsion can impact the energies of intermediates or transition states and influence the outcomes of reactions is a valuable resource for making sense of many organic chemistry phenomena. To get a holistic sense of the structure- and energy- ideas that students thought useful for explaining benzylic hydrobromination (the aim of RQ 1), we realized that we would need a more nuanced set of descriptors than those found in Table 1.2. Thus, we revisited our coding scheme and endeavored to more thoroughly describe the features of intermediate, product, or transition state structure students connected to relative energy. Discussion of these structural elements led to a coding scheme consisting of four categories: electron/charge delocalization, steric repulsion, carbocation substitution, and other/none (Table 1.3). We then examined which molecular species students referred to when describing relative energy or stability, leading to the development of the five energy codes shown in Table 1.3. Using this expanded coding scheme, we were able to more precisely describe students’ ideas and illustrate how they were connecting structural differences to energetic differences as part of predicting a reaction outcome. However, we acknowledge that, at the undergraduate level, constructing a correct explanation (*i.e.*, one that aligns with scientific canon) is important. We thus maintained the original 4-bin coding scheme displayed in Table 1.2 for addressing RQ2 and RQ3.

Table 1.3. Structure and energy codes, along with examples, for RQ1.

Structure Code	**Example**
Electron/charge delocalization	*No, I would not expect it to be the major product. The carbocation intermediate for product B should be less stable because the positive charge cannot be delocalized. The carbocation intermediate for A is* ***resonance stabilized by the benzylic ring adjacent. The ability to delocalize positive charge is very stabilizing****. Since the carbocation intermediate for B might be less stable, by Hammond's postulate the transition state for B is higher, so A would form faster.*
Steric repulsion	*Product A is the favored product because it has a lower energy potential than product B. This is because there is a* ***lower steric hinderance*** *with the Br addition in product A than in product B.*
Carbocation substitution	*I would expect Product A to be the major product of the reaction because the* ***carbonation [sic] intermediate is tertiary, which is more stable than the secondary carbonation [sic] intermediate in Product B****. The intermediate of Product A has lower energy because it is more stable this will yield the major product.*
Other/None	*You would expect that B would be the major project as H-Br goes in from a backside attack to flip the bond so Br is facing downwards. To get the other reaction would require a different resonance structure that is higher in energy* *The reaction is kinetically controlled so because product B has a lower energy intermediate, that intermediate will lead to the major product.*
Energy Code	**Example**
Intermediate	*No, product A would be the major product because it goes through the faster path of the* ***more stable intermediate****. The positive charge on the intermediate can be delocalized to the ring, so* ***this intermediate is stabilized by resonance and therefore has a lower energy.***
Transition State (with or without intermediate)	*I would not expect product B to be the major product of this reaction because product A carbocation has pi conjugation with the ring which allows for distribution of the positive charge. This also allows product a to have* ***a lower transition state and energy of the intermediate*** *through pi conjugation.*

Table 1.3. (Continued) Structure and energy codes, along with examples, for RQ1.

Energy Code	Example
Intermediate and/or Transition State + Product	*Product A would be the major product of this reaction because the Br being closer to the ring allows for more stabilization of the carbocation intermediate. The ring is able to delocalize the positive change and is therefore a* ***more stable intermediate and product.***
Product	*I would not expect it to be the major product. I would expect product A to be the major product because there is opportunity for pi-conjugation between the carbon the Br is bonded to and the benzene ring.* ***This makes the molecule more stable then product B.***
None/Ambiguous	I *wouldn't expect product B to be the major product since it would have a* ***higher energy pathway.*** *This is because the carbocation cannot delocalize with the pi conjugation from the ring as it would in product A.*

Associations between learning environment enrollment and the distribution of codes describing student explanations were examined using a series of Pearson's chi-square tests. An analogous approach was used to examine associations between the correctness of drawn mechanisms and the distribution of codes describing student explanations. All Pearson's chi-square tests were conducted using SPSS (*SPSS Statistics for Mac*, 2017). The output of each chi-square test included χ^2 and Cramer's V. Cohen's guidelines for effect size were used to interpret the value of Cramer's V (Cohen, 1988). For a contingency table containing three rows, values of 0.071, 0.212, and 0.345 for Cramer's V correspond to small, medium, and large effect sizes, respectively. The threshold of significance used for this study was a $p \leq 0.01$. Post hoc analysis of each chi square test which showed a significant association between variables was conducted in order to support inferences about the driver(s) of that significance. Standardized residuals for each cell were calculated by SPSS. Standardized residuals with positive values indicated more counts than expected by chance while residuals with negative values indicated fewer counts than expected by chance (Cohen, 1988). The magnitude of the standardized residual was compared to a critical value, which was 2.58 for a threshold of significance of $p \leq 0.01$ (Agresti, 2013; MacDonald & Gardner, 2000). Thus, cells with standardized residuals larger than 2.58 or smaller than -2.58 were considered drivers of the significant association.

Findings

RQ1: What knowledge elements did students activate and connect when asked to explain the outcome of a benzylic alkene addition reaction?

When justifying why one alkene addition product would predominate over the other, students commonly activated ideas related to charge/electron delocalization, steric repulsion, and carbocation substitution (Fig. 1.4). Nearly two-thirds of the students attended to differences in charge or electron distribution. Most responses of this sort compared the delocalization of positive charge between the two potential carbocation intermediates. A typical response of this nature is shown here:

> I would expect product A to be the major product because the [sic] proceeds down the lower energy pathway. Product A has a more stabilized intermediate because the positive charge is delocalized around the ring while in product B it cannot be delocalized because the positive charge is separated from the π conjugated ring by an sp3 hybridized carbon therefore raising the energy of the intermediate and favoring product A.

A few students described how the bromine atom would donate or withdraw electron density. Activation of ideas related to substituent donation or withdrawal of negative charge may have occurred due to the utility of these ideas in explaining electrophilic aromatic substitution (EAS) reactions, which was the focus of the chapter prior to the chapter on reactions of benzylic systems. Most explanations of why one EAS product would be favored over other alternatives are grounded in how aryl substituents affect reactivity by altering the electron density of the ring.

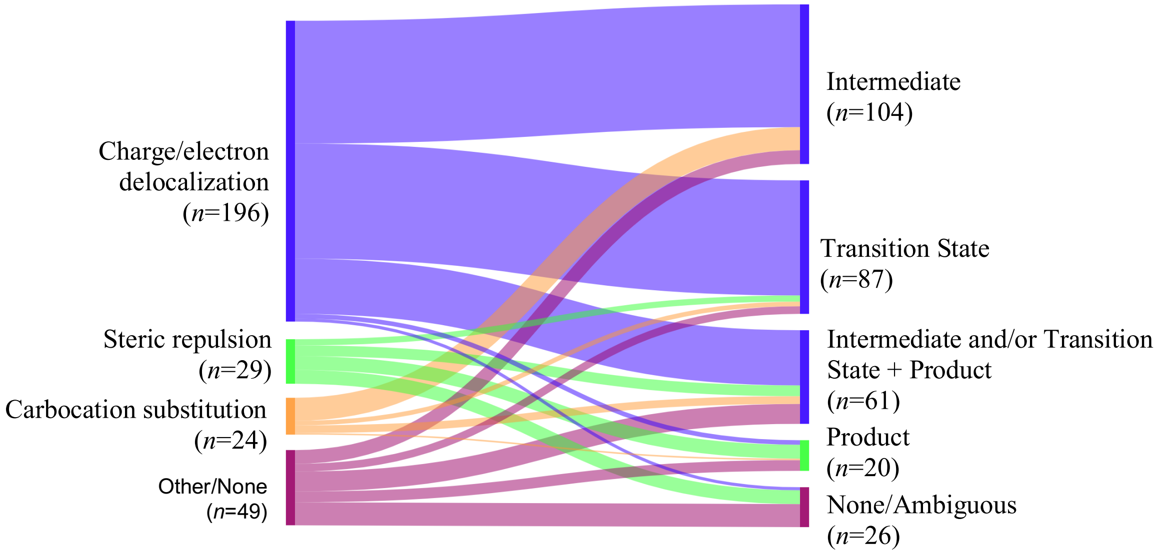


Figure 1.4. Sankey diagram depicting how students connected structural features (left) to the stabilities or relative energies of species (right) involved in the benzylic alkene hydrobromination reaction.

The second most common factor cited in students' explanations was steric repulsion between the large bromine atom and the aryl ring. An example of an explanation utilizing steric repulsion is given here:

> I would expect Product B to be the major product mainly due to sterics. Br is a large atom and being closer to the aromatic ring may cause some steric repulsion, making the molecule more unstable and the intermediate higher energy. Therefore, the reaction make product B would include less sterics than A and thus, would have a lower energy intermediate, form more easily, and have a more stable product.

Steric repulsion is often a productive resource for rationalizing observations in organic chemistry, and at this point in the course, students had encountered it in the context of alkane conformations, other alkene addition reactions, alkyl halide reactions, and electrophilic aromatic substitution reactions. Indeed, the linkage between steric repulsion and the relative energy of intermediates described by the prior response is generally correct and very often useful! However, in reactions that proceed through carbocation intermediates, such as this one, steric repulsion has a minimal effect on the outcome because the nucleophile (i.e., bromide) reacts from above or below the planar reactive carbon.

The other structural feature invoked in several responses was carbocation substitution. In reactions that proceed through carbocation intermediates, the degree of substitution often dictates the regioselectivity of the reaction since increased substitution allows for increased hyperconjugation, which stabilizes the carbocation via charge delocalization. A tag of "carbocation substitution" denotes invocation of intermediate substitution without linking this to electron delocalization (which was a separate tag). Accordingly, it is likely many students whose responses were described by this code used "more substituted carbocations are more stable" as a heuristic (Talanquer, 2014). Since both potential carbocation intermediates in this reaction are secondary, substitution is not a useful means of discriminating between the two reaction pathways referenced in our diagnostic prompt. Some students recognized this and concluded that the products would be formed in roughly equal amounts, as the following response illustrates: "I would expect both of the products to have an equal outcome because they are both secondary carbocations." Other students believed that the carbocation leading to Product A was tertiary or "closer to being tertiary," presumably because it was next to the larger aryl group.

To connect differences in structure to reaction outcome, most students included an energetic component to their explanations (Fig. 1.4). Of the students who described charge/electron delocalization, most connected it to the stability of the intermediate, as shown in the first quote. In addition, a large number of students connected electron delocalization to the energy of the transition states, as follows:

> No, I would not expect it to be the major product. The carbocation intermediate for product B should be less stable because the positive charge cannot be delocalized. The carbocation intermediate for A is resonance stabilized by the benzylic ring adjacent. The ability to delocalize positive charge is very stabilizing. Since the carbocation intermediate for B might be less stable, by Hammond's postulate the transition state for B is higher, so A would form faster.

Only a handful of students used electron delocalization to justify a difference in product energies. More commonly, students connected electron delocalization to intermediate and/or transition state along with product energy:

> I would NOT expect product B to be the major product of this reaction. This is because when the carbocation intermediate is formed for the reaction that leads to product A, it is a more stable carbocation than in B. I determined this because the positive charge is located on the benzylic carbon in reaction A, which is very stabilizing due to it being in pi conjugation with the benzene ring. The positive charge is close enough to the ring that it can be delocalized in the ring which is very stabilizing. Therefore, since the carbocation is more stable in reaction A, this would mean that it would produce the more stable product.

Finally, a few of the students who invoked electron delocalization either described energy in ambiguous terms (*e.g.,* "lower energy pathway") or did not discuss energy or stability at all in their explanation. Attributing the outcome to differences in steric repulsion was mostly associated with descriptions of the relative energy difference between the two possible products. Unsurprisingly, carbocation substitution was mostly connected to the stability of the carbocation intermediate.

Overall, analysis of explanations revealed that students possess useful ideas of how structure impacts the stability of a molecule and how relative energy at various points along competing reaction pathways dictates outcome. From an instructional standpoint, these results suggest that it may be productive to prompt students to consider the relative impacts of various factors on stability and to consider if and how intermediate and/or transition state energy, product energy, and reaction outcome are interrelated. As demonstrated, this type of analysis provides more insight into how instruction can build on students' ideas than an analysis that identifies students' responses as merely "right" or "wrong."

RQ2: How did the electron-pushing mechanisms students drew relate to the explanations they provided?

Most students across all learning environments drew mechanisms with correct intermediates and electron-pushing arrows (Fig. 1.5). Slight differences were observed between the aggregate distributions for Mechanism A and Mechanism B. Students who earned different codes for A and B tended to leave off an arrow in one of the mechanisms. Two students altered the reaction conditions for Mechanism B by adding in light or peroxides to render it a radical reaction, which would favor formation of Product B. Disaggregating by learning environment revealed a significant association between OChem I enrollment and mechanism code ($\chi^2(2) = 24.9$, $p < 0.001$, Cramer's V = 0.29, medium effect). Post hoc analysis of the results of this test revealed that a negative association between enrollment in the high 3D learning environment and drawing an incorrect mechanism was the primary driver of significance. That is, students who were enrolled in the high 3D environment for OChem I were substantially *less* likely to draw an incorrect mechanism than would be expected by chance. A significant association between OChem II enrollment and mechanism code was also observed ($\chi^2(2) = 10.3$, $p < 0.001$, Cramer's V = 0.19, small effect). Post hoc analysis of the results of this test showed no primary driver(s) of the significant association.

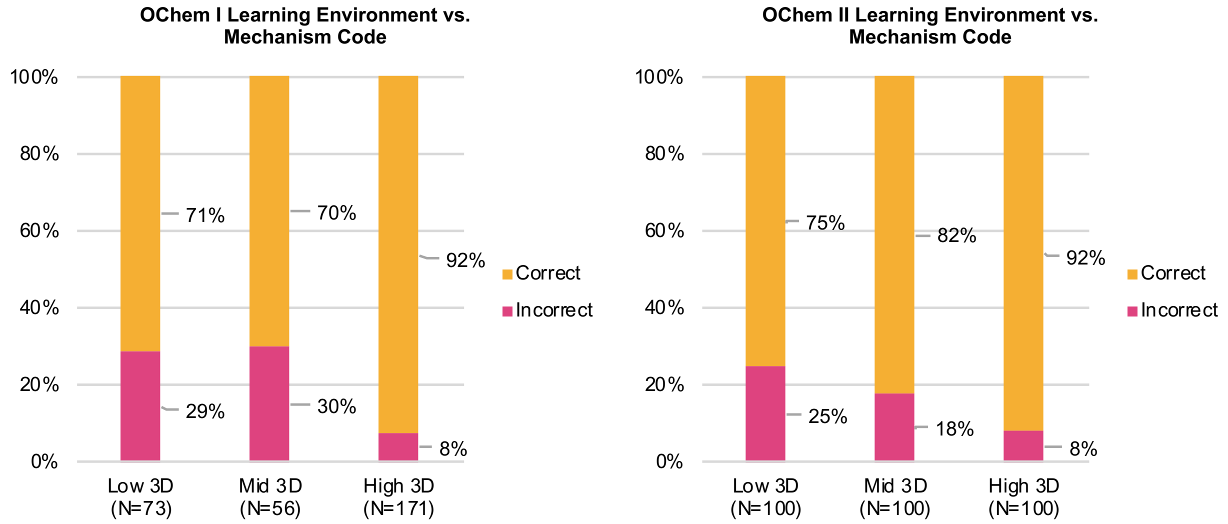
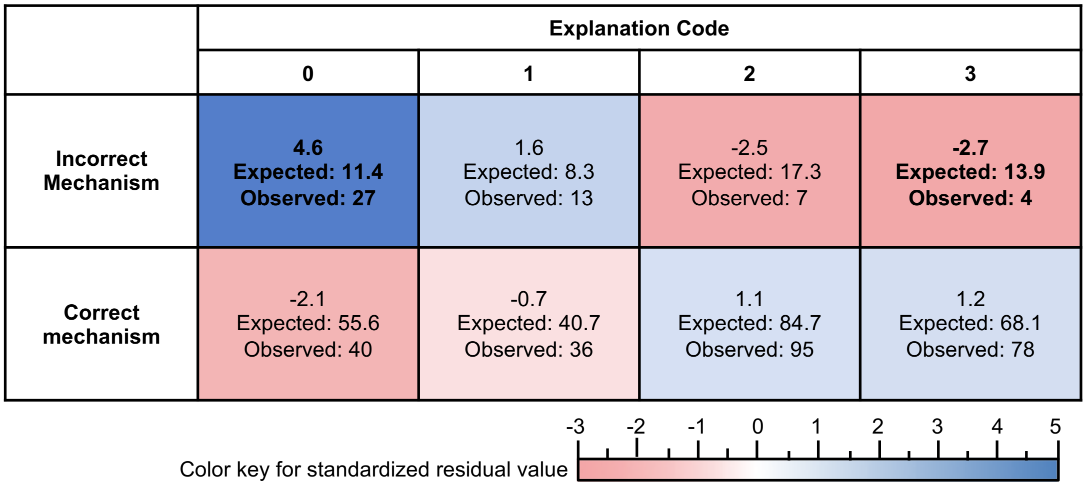


Figure 1.5. Distribution of mechanisms codes according to learning environment for OChem I (left) and OChem II (right). The vast majority of students in all learning environments drew the mechanism correctly.

Next, we examined the relationship between mechanism code and explanation code. A chi-square test revealed a significant association with medium effect size, $\chi^2(3) = 44.9$, $p < 0.001$, Cramer's V = 0.387. Post hoc analysis of this chi-square test showed that the significant association was driven by students who drew incorrect mechanisms (Fig. 1.6). Students who drew incorrect mechanisms were substantially *more* likely to provide explanations that were coded as "0" and substantially *less* likely to provide explanations that were coded as a "3" than would be expected by chance. In fact, of the 51 students who drew an incorrect mechanism, only four offered an explanation coded as a "3." However, for students who drew the correct mechanism, there were no substantive differences between expected and observed counts for each explanation code. This means that a student was unlikely to provide a higher code explanation without the correct mechanism, but drawing a correct mechanism was no guarantee that they could explain the meaning underpinning that mechanism. This is consistent with a large body of prior research (Grove et al., 2012; Houchlei et al., 2021; Bhattacharyya & Bodner, 2005; Bhattacharyya, 2014; Ferguson & Bodner, 2008; Anzovino & Bretz, 2015).

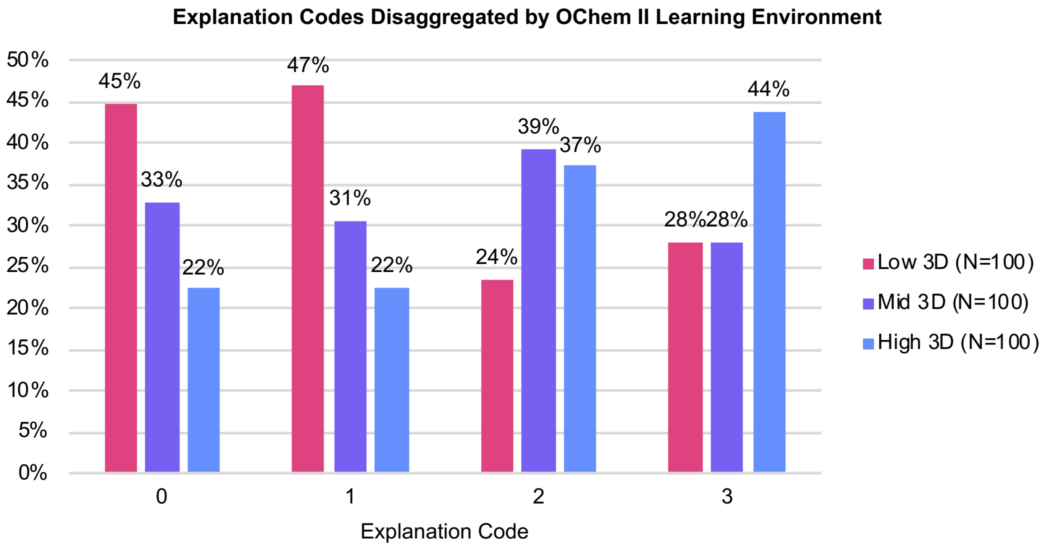


	Explanation Code			
	0	**1**	**2**	**3**
Incorrect Mechanism	**4.6** **Expected: 11.4** **Observed: 27**	1.6 Expected: 8.3 Observed: 13	-2.5 Expected: 17.3 Observed: 7	**-2.7** **Expected: 13.9** **Observed: 4**
Correct mechanism	-2.1 Expected: 55.6 Observed: 40	-0.7 Expected: 40.7 Observed: 36	1.1 Expected: 84.7 Observed: 95	1.2 Expected: 68.1 Observed: 78

Figure 1.6. Contingency table for the χ^2 square test examining association between mechanism code and explanation code. Standardized residuals and expected and observed counts are reported in each cell. Standardized residuals greater than the critical value (±2.58) are in bold. Cells with positive standardized residuals are shaded blue while cells with negative standardized residuals are shaded red.

RQ3: How did the intellectual work emphasized and rewarded on assessments given in each learning environment relate to the structure-energy connections invoked in student explanations?

To examine whether students who were routinely expected to construct explanations on assessments were more likely to productively connect structure- and energy- ideas when responding to our prompt, student responses were disaggregated by learning environment. Initially, we separated student responses according to OChem II learning environment enrollment, which corresponds to the semester in which the prompt was administered (Fig. 1.7). A significant association with a small effect size was found between OChem II learning environment enrollment and the distribution of explanation codes, $\chi^2(6) = 18.2$, $p = 0.006$, Cramer's V = 0.17. A post hoc analysis of the results of this chi-square test found that none of the cells strongly drove the significant association; that is, no cells had a standardized residual greater than 2.58 in magnitude (Fig. 1.7). We then disaggregated responses by OChem I learning environment enrollment and repeated this analysis (Fig. 1.8). A significant association between OChem I learning environment enrollment and explanation code was found, $\chi^2(6) = 36.4$, $p < 0.001$, Cramer's V = 0.25 (medium effect size). A post hoc analysis of the results of the chi-square test showed that students enrolled in the low 3D learning environment were substantially *more* likely than would be expected by chance to construct an explanation coded as "0" while students enrolled in the high 3D learning environment were substantially *less* likely than would be expected by chance to construct an explanation coded as "0" (Fig. 1.8).



	Explanation Code			
	0	**1**	**2**	**3**
Low 3D *N*=100	1.6 Expected: 22.3 Observed: 30	1.6 Expected: 16.3 Observed: 23	-1.7 Expected: 34.0 Observed: 24	-0.8 Expected: 27.3 Observed: 23
Mid 3D *N*=100	-0.1 Expected: 22.3 Observed: 22	-0.3 Expected: 16.3 Observed: 15	1.0 Expected: 34.0 Observed: 40	-0.8 Expected: 27.3 Observed: 23
High 3D *N*=100	-1.6 Expected: 22.3 Observed: 15	-1.3 Expected: 16.3 Observed: 11	0.7 Expected: 34.0 Observed: 38	1.7 Expected: 27.3 Observed: 36

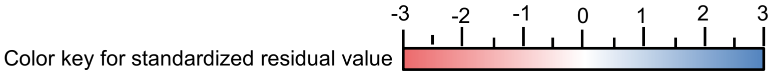
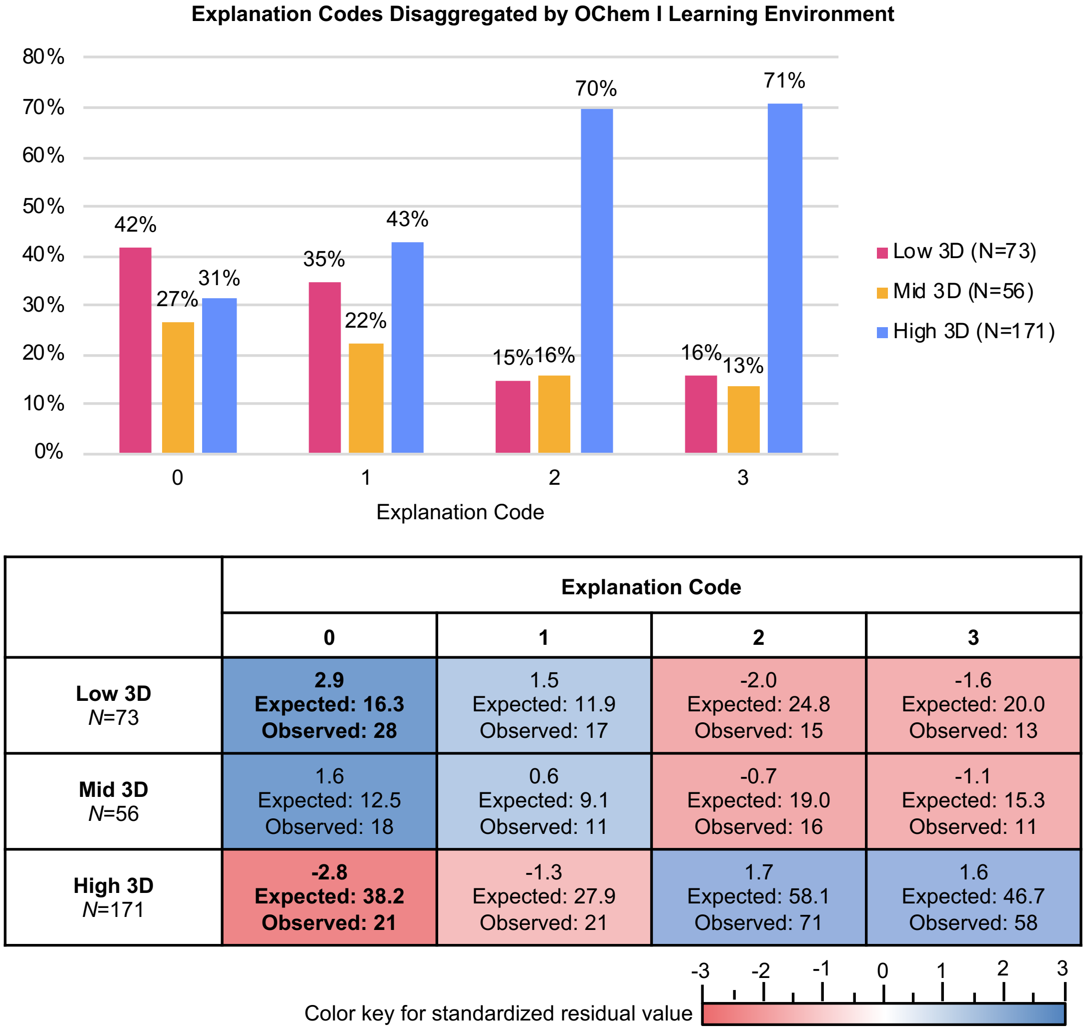


Figure 1.7. Distribution of explanation codes according to OChem II learning environment (top). Contingency table for the χ^2 square test examining association between OChem II learning environment and explanation code (bottom). Standardized residuals and expected and observed counts are reported in each cell. Standardized residuals greater than the critical value (±2.58) are in bold. Cells with positive standardized residuals are shaded blue while cells with negative standardized residuals are shaded red.

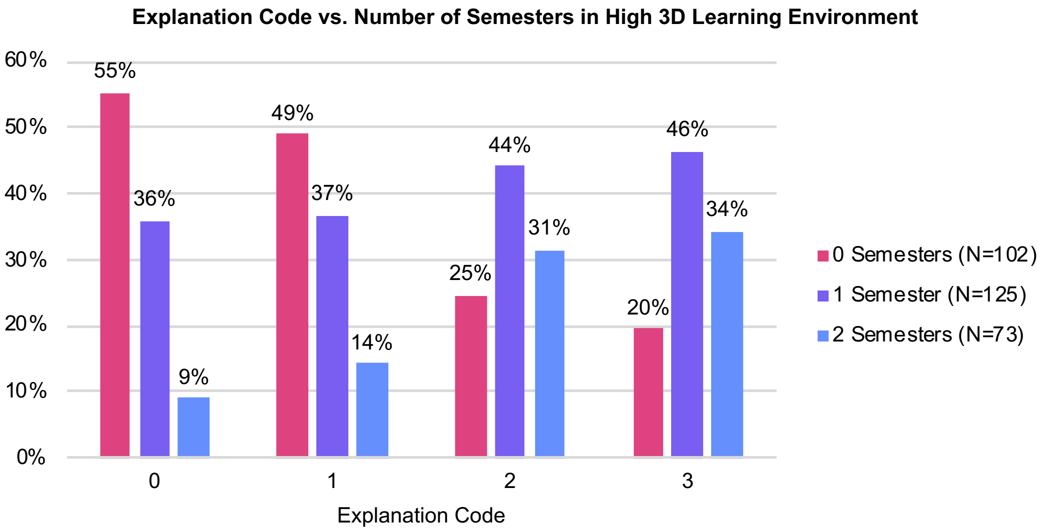


	Explanation Code			
	0	1	2	3
Low 3D *N*=73	**2.9** **Expected: 16.3** **Observed: 28**	1.5 Expected: 11.9 Observed: 17	-2.0 Expected: 24.8 Observed: 15	-1.6 Expected: 20.0 Observed: 13
Mid 3D *N*=56	1.6 Expected: 12.5 Observed: 18	0.6 Expected: 9.1 Observed: 11	-0.7 Expected: 19.0 Observed: 16	-1.1 Expected: 15.3 Observed: 11
High 3D *N*=171	**-2.8** **Expected: 38.2** **Observed: 21**	-1.3 Expected: 27.9 Observed: 21	1.7 Expected: 58.1 Observed: 71	1.6 Expected: 46.7 Observed: 58

Figure 1.8. Distribution of explanation codes according to OChem I learning environment (top). Contingency table for the χ^2 square test examining association between OChem I learning environment and explanation code (bottom). Standardized residuals and expected and observed counts are reported in each cell. Standardized residuals greater than the critical value (±2.58) are in bold. Cells with positive standardized residuals are shaded blue while cells with negative standardized residuals are shaded red.

We hypothesize that the more substantive association between OChem I section enrollment and explanation code distribution is due to the foundational nature of the first semester course. Most broadly useful explanatory ideas are introduced in OChem I, including π conjugation, hyperconjugation, and steric repulsion. In addition, although the reaction used in our diagnostic assessment is revisited in OChem II in the context of benzylic reactivity, electrophilic addition to alkenes is first introduced early in OChem I and could be explained at that time. We should note that students who were enrolled in the low 3D learning environment for OChem I did not have the opportunity to use π conjugation as an explanatory idea until the end of the course. It is conceivable that these students might be less inclined to integrate π conjugation into their explanations on account of having less experience doing so than other study participants.

Finally, we wanted to examine how learning environment enrollment over the two-semester introductory organic sequence related to the distribution of codes describing student explanations. Specifically, we looked at how the number of semesters students spent in the high 3D learning environment related to the distribution of codes describing their explanations (Fig. 1.9). A significant association with medium effect size was observed, $\chi^2(6) = 35.2$, $p < 0.001$, Cramer's V = 0.24. A post hoc analysis of the results of this test revealed that the vast majority of students who were enrolled in the high 3D learning environment for both semesters (82%) connected π conjugation to intermediate or transition state energy in their explanation (Fig. 1.9). By contrast, students who spent two semesters in a course that rarely (or never) emphasized explanations on exams (i.e., low- or mid-3D learning environment) were substantially less likely to connect differences in π conjugation to differences in transition state energies in their explanations.



		Explanation Code			
		0	**1**	**2**	**3**
# of Semesters in High 3D Environment	**0** *N*=102	**3.0** **Expected: 22.8** **Observed: 37**	1.8 Expected: 16.7 Observed: 24	-1.6 Expected: 34.7 Observed: 25	-2.2 Expected: 27.9 Observed: 16
	1 *N*=121	-0.7 Expected: 27.9 Observed: 24	-0.5 Expected: 20.4 Observed: 18	0.4 Expected: 42.5 Observed: 45	0.7 Expected: 34.2 Observed: 38
	2 *N*=73	**-2.6** **Expected: 16.3** **Observed: 6**	-1.4 Expected: 11.9 Observed: 7	1.4 Expected: 24.8 Observed: 32	1.8 Expected: 20.0 Observed: 28

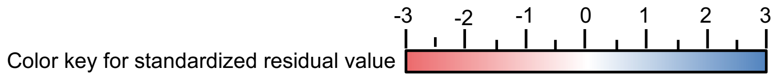


Figure 1.9. Distribution of explanation codes according to number of semesters enrolled in the high 3D learning environment (top). Contingency table for the χ^2 square test examining association between number of semesters enrolled in high 3D learning environment and explanation code (bottom). Standardized residuals and expected and observed counts are reported in each cell. Standardized residuals greater than the critical value (±2.58) are in bold. Cells with positive standardized residuals are shaded blue while cells with negative standardized residuals are shaded red.

The differences in explanation code distributions among learning environments point toward the importance of assessment emphasis in messaging course priorities. While the general format of the learning environments was the same (i.e., recorded lectures, virtual instructor-led problem-solving sessions, weekly discussions, four exams) and all sections made use of the same order of topics and textbook, the performances signaled as important differed markedly. To earn high marks in the high 3D learning environment, students were required to use models to predict and explain reaction outcomes. By contrast, students could succeed perfectly well in the low 3D environment without ever connecting structure- and energy- ideas to why phenomena happen. Given this difference in what was emphasized and rewarded on exams, it makes sense that we found that students who were enrolled in the high 3D learning environment were substantially *less* likely to receive a code of "0" on their explanations. Conversely, students who enrolled in a lecture environment where they were never asked to provide explanations for phenomena on assessments were significantly *more* likely to receive a code of "0" on their explanations. While our study design prevents us from determining if differences in assessments *caused* the differences in explanation code distributions, this work supports the notion that assessments convey strong messages to students regarding the intellectual work central to organic chemistry. If we want to support students in constructing explanations for how and why observable events happen the way they do, we likely need to reward them for doing so throughout the course.

Overall, most students were able to generate an explanation for the outcome of the focal benzylic alkene hydrobromination reaction by using charge/electron delocalization to account for differences in intermediate or transition state energies. Since many phenomena central to organic chemistry can be explained, at least in part, by charge delocalization, it is encouraging that most students activated these ideas when justifying their claim as to the reaction outcome. Whether the stabilizing influence of electron delocalization is viewed as a useful explanatory resource in the context of other reactions remains to be seen. Some of the other factors invoked by students, such as the stabilization from increased substitution on a carbocation or the destabilizing influence of steric repulsion, demonstrate that many of the students who did not discuss π conjugation possess useful resources for understanding other reaction systems.

Limitations

We cannot directly observe students' thinking and must use their drawings and writings to infer what they know and can do in a given context. Due to the inherent restrictions of a virtual semester, we do not know how many of the students worked together to complete the assignment, so not all responses may be indicative of an individual student's thinking. Furthermore, this study focused on student responses to a single prompt at a single point in the semester. If we were to administer this prompt at a different time, under different conditions, or with different wording, students may exhibit new patterns of resource activation. As an example, the potential energy surface question (part 3) may have cued different resources for students in the high 3D learning environment, which emphasized use of potential energy surfaces, compared to students in the other learning environments, which did not. Similarly, if we were to repeat this study using a diagnostic prompt focused on a different phenomenon, we do not know if we would observe the same associations between learning environment enrollment and student explanations. Thus, we are not suggesting that one learning environment is definitively "better" than another. Furthermore, our analysis of the different learning environments only characterized assessment emphasis. While the format of the courses were similar (e.g., video lectures, weekly discussion sections), we did not examine the emphases instructors placed on 3D performances during lecture. It is possible that the instructors differed in how they modeled the practice of constructing explanations, which may have contributed to the differences we observed. Finally, we have no evidence as to how students were framing their engagement with the assessment item. The way students understand the aim of explanation construction, the appropriate sources of knowledge to draw from, and what constitutes a credible justification shape the degree to which they perceive the task as a “school science” exercise versus an opportunity to make sense of a phenomenon (Chinn et al., 2011; Lemke, 1990). Given that one’s framing is largely dependent on past experience, it is likely that students viewed the diagnostic task in a similar manner to other problems given during the course, in which the goal is rapidly producing a “correct answer” (Tannen, 1993; Hammer & Elby, 2002). Future work will focus on methods for characterizing and influencing students’ frames in order to support students in engaging more authentically with organic chemistry.

Conclusions

In this study, we identified the structure-energy connections students made to explain the outcome of an alkene addition reaction, explored how their explanations related to their mechanisms, and examined the relationship between course assessment emphasis and explanations. We found that most students justified their outcome predictions based on the relative energies of intermediates or transition states caused by differences in the extent of electron delocalization. Generally useful ideas relating to steric repulsion and degree of substitution on carbocation intermediates were also observed, illustrating that students who do not arrive at a canonically correct answer may possess productive ideas upon which to build. Unsurprisingly, several of these ideas (e.g., steric repulsion, carbocation substitution, activation energy) matched the concepts Bodé et al. (2019) found embedded in students’ arguments pertaining to an S_N1 reaction, which also proceeds though a carbocation intermediate. Our analysis indicates that we should not assume that students who draw a correct mechanism understand why the reaction proceeds in the manner illustrated. In-line with previously published literature, we found no substantive association between a correctly drawn mechanism and any particular explanation code (Houchlei et al., 2021; Caspari, Weinrich, et al., 2018; Bhattacharyya & Bodner, 2005). We did, however, find a significant association between learning environment enrollment and explanation code with a medium effect size. Students who were routinely expected to construct explanations on assessments tended to connect electron delocalization to intermediate or transition state energy in the context of our prompt.

Since constructing explanations is central to the work of chemists, we should design learning environments to support students in this practice (National Research Council, 2012). Studies have provided evidence demonstrating the positive impact curricular reforms can have towards achieving this goal, but the impact transforming assessments in an otherwise traditional course cannot be determined from these studies (Crandell et al., 2020; Crandell et al., 2019; Webber & Flynn, 2018). Our study begins to address that gap in the literature. Based on our findings, we will continue to probe the relationship between assessment emphasis and students' propensity to construct productive causal accounts for phenomena, as transformation of assessments appears to be a potentially productive avenue for reform. In particular, we are interested in examining students' reasoning using structure-energy connections across multiple prompts over a year of introductory organic chemistry instruction.

Implications for Research

This study lends further support to the oft-repeated claim that assessments signal to students "what counts" in a learning environment (Momsen et al., 2013; Scouller, 1998; Scouller & Prosser, 1994; Snyder, 1973; Entwistle, 1991; Crooks, 1988). Relatedly, the measures researchers use to determine the success of interventions or transformations message the sorts of performances those researchers think are important. We argue that researchers should take the nature of assessments into account when evaluating reform efforts. Noting some sort of improvement in course grade or exam performance means very little if we do not know what "success" meant in the course or on the exam. As the format and content of exams may vary considerably, the nature of the assessment items on those exams should be reported so that the reader knows precisely what sort of performances are influenced by the intervention. Relatedly, we argue that researchers should prioritize use of assessment items that have the potential to elicit detailed evidence of engagement in aspects of “doing science,” such as constructed-response items that ask for the reasoning supporting a claim. Items that require only a claim, including those that instruct students to draw a mechanism, cannot support inferences as to why students claimed what they did.

Implications for Instruction

Assessments are powerful tools that instructors of chemistry can use to shape learning. We urge instructors to reflect on whether the performances they award points to align with intellectual work central to the discipline under study. Do assessment tasks promote construction of causal accounts for phenomena or a reliance on pattern recognition? We also suggest that instructors who want to improve students’ engagement in “doing chemistry” but are unable to implement whole curricular reforms consider how they might modify their formative and summative assessments in order to support a coherent emphasis on constructing explanations for phenomena. Although assessment reform is non-trivial and may increase the grading burden, it is vital that “success” in organic chemistry courses align with productive engagement in intellectual work characteristic of the discipline.

Acknowledgements

Support for this work was provided by the Office of the Vice Chancellor for Research and Graduate Study at the University of Wisconsin–Madison with funding from the Wisconsin Alumni Research Foundation. We would like to thank all instructors who provided their exams for us to analyze.

References

Agresti, A. (2013). Inference for two-way contingency tables. In *Categorical data analysis*. Wiley.

Anzovino, M. E., & Bretz, S. L. (2015). Organic chemistry students' ideas about nucleophiles and electrophiles: The role of charges and mechanisms. *Chemistry Education Research and Practice*, *16*(4), 797–810. https://doi.org/10.1039/C5RP00113G

Arjoon, J. A., Xu, X., & Lewis, J. E. (2013). Understanding the state of the art for measurement in chemistry education research: Examining the psychometric evidence. *Journal of Chemical Education*, *90*(5), 536–545. https://doi.org/10.1021/ed3002013

Bhattacharyya, G. (2014). Trials and tribulations: Student approaches and difficulties with proposing mechanisms using the electron-pushing formalism. *Chemistry Education Research and Practice*, *15*(4), 594–609. https://doi.org/10.1039/C3RP00127J

Bhattacharyya, G., & Bodner, G. M. (2005). "It gets me to the product": How students propose organic mechanisms. *Journal of Chemical Education*, *82*(9), 1402. https://doi.org/10.1021/ed082p1402

Bodé, N. E., Deng, J. M., & Flynn, A. B. (2019). Getting past the rules and to the WHY: Causal mechanistic arguments when judging the plausibility of organic reaction mechanisms. *Journal of Chemical Education*, *96*(6), 1068–1082. https://doi.org/10.1021/acs.jchemed.8b00719

Brown, D. E. (2014). Students' conceptions as dynamically emergent structures. *Science & Education*, *23*(7), 1463–1483. https://doi.org/10.1007/s11191-013-9655-9

Caspari, I., Kranz, D., & Graulich, N. (2018). Resolving the complexity of organic chemistry students' reasoning through the lens of a mechanistic framework. *Chemistry Education Research and Practice*, *19*(4), 1117–1141. https://doi.org/10.1039/C8RP00131F

Caspari, I., Weinrich, M. L., Sevian, H., & Graulich, N. (2018). This mechanistic step is "productive": Organic chemistry students' backward-oriented reasoning. *Chemistry Education Research and Practice*, *19*(1), 42–59. https://doi.org/10.1039/C7RP00124J

Chinn, C. A., Buckland, L. A., & Samarapungavan, A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist*, *46*(3), 141–167. https://doi.org/10.1080/00461520.2011.587722

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Earlbaum Associates.

Cooper, M. M., Corley, L. M., & Underwood, S. M. (2013). An investigation of college chemistry students' understanding of structure–property relationships. *Journal of Research in Science Teaching*, *50*(6), 699–721. https://doi.org/10.1002/tea.21093

Cooper, M. M., & Klymkowsky, M. (2013). Chemistry, Life, the Universe, and Everything: A new approach to general chemistry, and a model for curriculum reform. *Journal of Chemical Education*, *90*(9), 1116–1122. https://doi.org/10.1021/ed300456y

Cooper, M. M., Stowe, R. L., Crandell, O. M., & Klymkowsky, M. W. (2019). Organic Chemistry, Life, the Universe and Everything (OCLUE): A transformed organic chemistry curriculum. *Journal of Chemical Education*, *96*(9), 1858–1872. https://doi.org/10.1021/acs.jchemed.9b00401

Crandell, O. M., Kouyoumdjian, H., Underwood, S. M., & Cooper, M. M. (2019). Reasoning about reactions in organic chemistry: Starting it in general chemistry. *Journal of Chemical Education*, *96*(2), 213–226. https://doi.org/10.1021/acs.jchemed.8b00784

Crandell, O. M., Lockhart, M. A., & Cooper, M. M. (2020). Arrows on the page are not a good gauge: Evidence for the importance of causal mechanistic explanations about nucleophilic substitution in organic chemistry. *Journal of Chemical Education*, *97*(2), 313–327. https://doi.org/10.1021/acs.jchemed.9b00815

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, *58*(4), 438–481.

diSessa, A. A. (1988). Knowledge in pieces. In *Constructivism in the Computer Age* (1st ed., p. 22). Psychology Press.

diSessa, A. A. (2006). A history of conceptual change research: Threads and fault lines. In *The Cambridge handbook of the learning sciences* (pp. 265–281). Cambridge University Press.

Entwistle, N. J. (1991). Approaches to learning and perceptions of the learning environment. *Higher Education*, *22*(3), 201–204. https://doi.org/10.1007/BF00132287

Ferguson, R.; Bodner, G. M. (2008). Making sense of the arrow-pushing formalism among chemistry majors enrolled in organic chemistry. *Chemistry Education Research and Practice*, *9*, 102–113.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*(5), 378–382.

Flynn, A. B., & Ogilvie, W. W. (2015). Mechanisms before reactions: A mechanistic approach to the organic chemistry curriculum based on patterns of electron flow. *Journal of Chemical Education*, *92*(5), 803–810. https://doi.org/10.1021/ed500284d

Goodwin, W. M. (2003). Explanation in organic chemistry. *Annals of the New York Academy of Sciences*, *988*(1), 141–153. https://doi.org/10.1111/j.1749-6632.2003.tb06093.x

Goodwin, W. M. (2008). Structural formulas and explanation in organic chemistry. *Foundations of Chemistry*, *10*(2), 117–127. https://doi.org/10.1007/s10698-007-9033-2

Grove, N. P., Cooper, M. M., & Rush, K. M. (2012). Decorating with arrows: Toward the development of representational competence in organic chemistry. *Journal of Chemical Education*, *89*(7), 844–849. https://doi.org/10.1021/ed2003934

Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 89–119).

Hammer, D., & Elby, A. (2002). On the form of a personal epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 169–190). Lawrence Erlbaum Associates, Inc.

Houchlei, S. K., Bloch, R. R., & Cooper, M. M. (2021). Mechanisms, models, and explanations: Analyzing the mechanistic paths students take to reach a product for familiar and unfamiliar organic reactions. *Journal of Chemical Education*, *98*(9), 2751–2764. https://doi.org/10.1021/acs.jchemed.1c00099

Lamichhane, R., Reck, C., & Maltese, A. V. (2018). Undergraduate chemistry students' misconceptions about reaction coordinate diagrams. *Chemistry Education Research and Practice*, *19*(3), 834–845. https://doi.org/10.1039/C8RP00045J

Laverty, J. T., Underwood, S. M., Matz, R. L., Posey, L. A., Carmel, J. H., Caballero, M. D., Fata-Hartley, C. L., Ebert-May, D., Jardeleza, S. E., & Cooper, M. M. (2016). Characterizing college science assessments: The three-dimensional learning assessment protocol. *PLOS ONE*, *11*(9), e0162333. https://doi.org/10.1371/journal.pone.0162333

Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Ablex Publishing Corporation.

Loudon, M., & Parise, J. (2016). *Organic chemistry* (6th ed.). W. H. Freeman and Company.

MacDonald, P. L., & Gardner, R. C. (2000). Type I error rate comparisons of post hoc procedures for I j chi-square tables. *Educational and Psychological Measurement*, *60*(5), 735–754. https://doi.org/10.1177/00131640021970871

Manz, E., Lehrer, R., & Schauble, L. (2020). Rethinking the classroom science investigation. *Journal of Research in Science Teaching*, *57*(7), 1148–1174. https://doi.org/10.1002/tea.21625

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, *25*(4), 6–20. https://doi.org/10.1111/j.1745-3992.2006.00075.x

Momsen, J., Offerdahl, E., Kryjevskaia, M., Montplaisir, L., Anderson, E., & Grosz, N. (2013). Using assessments to investigate and compare the nature of learning in undergraduate science courses. *CBE—Life Sciences Education*, *12*(2), 239–249. https://doi.org/10.1187/cbe.12-08-0130

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. National Academies Press. https://doi.org/10.17226/10019

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press. https://doi.org/10.17226/13165

Popova, M., & Bretz, S. L. (2018a). Organic chemistry students' challenges with coherence formation between reactions and reaction coordinate diagrams. *Chemistry Education Research and Practice*, *19*(3), 732–745. https://doi.org/10.1039/C8RP00064F

Popova, M., & Bretz, S. L. (2018b). Organic chemistry students' interpretations of the surface features of reaction coordinate diagrams. *Chemistry Education Research and Practice*, *19*(3), 919–931. https://doi.org/10.1039/C8RP00063H

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*(2), 211–227. https://doi.org/10.1002/sce.3730660207

Robertson, A. D., Atkins, L. J., Levin, D. M., & Richards, J. L. (2015). What is responsive teaching? In *Responsive teaching in science and mathematics* (pp. 19–53). Routledge.

Rosenberg, S., Hammer, D., & Phelan, J. (2006). Multiple epistemological coherences in an eighth-grade discussion of the rock cycle. *The Journal of the Learning Sciences*, *15*(2), 261–292.

Schwarz, C. V., Passmore, C., & Reiser, B. J. (2017). *Helping students make sense of the world using next generation science and engineering practices*. NSTA Press.

Scouller, K. M. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, *35*(4), 453–472. https://doi.org/10.1023/A:1003196224280

Scouller, K. M., & Prosser, M. (1994). Students' experiences in studying for multiple choice question examinations. *Studies in Higher Education*, *19*(3), 267–279. https://doi.org/10.1080/03075079412331381870

Snyder, B. (1973). *The hidden curriculum*. The MIT Press: Cambridge, MA.

SPSS statistics for Mac. (2017). [Computer software]. IBM Corp.

Stowe, R. L., & Cooper, M. M. (2017). Practicing what we preach: Assessing “critical thinking” in organic chemistry. *Journal of Chemical Education*, *94*(12), 1852–1859. https://doi.org/10.1021/acs.jchemed.7b00335

Stowe, R. L., Esselman, B. J., Ralph, V. R., Ellison, A. J., Martell, J. D., DeGlopper, K. S., & Schwarz, C. E. (2020). Impact of maintaining assessment emphasis on three-dimensional learning as organic chemistry moved online. *Journal of Chemical Education*, *97*(9), 2408–2420. https://doi.org/10.1021/acs.jchemed.0c00757

Stowe, R. L., Scharlott, L. J., Ralph, V. R., Becker, N. M., & Cooper, M. M. (2021). You are what you assess: The case for emphasizing chemistry on chemistry assessments. *Journal of Chemical Education*, *98*(8), 2490–2495. https://doi.org/10.1021/acs.jchemed.1c00532

Talanquer, V. (2014). Chemistry education: Ten heuristics to tame. *Journal of Chemical Education*, *91*(8), 1091–1097. https://doi.org/10.1021/ed4008765

Talanquer, V., & Pollard, J. (2010). Let’s teach how we think instead of what we know. *Chemistry Education Research and Practice*, *11*(2), 74–83. https://doi.org/10.1039/C005349J

Tannen, D. (1993). *Framing in discourse*. Oxford University Press.

The Nobel Committee for Chemistry. (2021). *Enamine and iminium ion-mediated organocatalysis*. The Royal Swedish Society of Chemistry. https://www.nobelprize.org/uploads/2021/10/advanced-chemistryprize2021-3.pdf

Webber, D. M., & Flynn, A. B. (2018). How are students solving familiar and unfamiliar organic chemistry mechanism questions in a new curriculum? *Journal of Chemical Education*, *95*(9), 1451–1467. https://doi.org/10.1021/acs.jchemed.8b00158

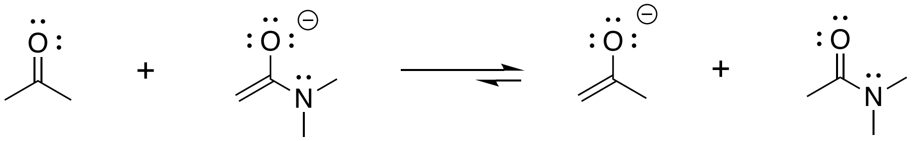
Supporting Information

The Supporting Information includes the following components: examples of 3D and non-3D assessment items, a canonical answer to the assessment item used in this study, and a detailed description of the explanation coding process.

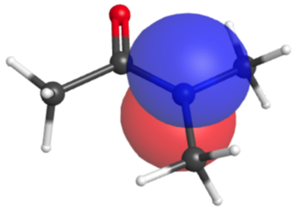
Examples of 3D and Non-3D Assessment Items

An example of a 3D assessment item is shown below (Fig. 1.S1). This item meets the criteria for “Developing and Using Models” (Fig. 1.S2). In this item, the following phenomenon is presented: the relative amount of reactants (amide enolate and ketone) and products (ketone enolate and amide) at equilibrium. A representation is provided in the form of line-angle drawings of structures and WebMO-generated images of the orbitals containing the nitrogen lone pairs. Students are asked to use the representations to explain the phenomenon, i.e., engage in scientific modeling. Core ideas of Electrostatic and Bonding Interactions, Atomic/Molecular Structure and Properties, Energy, and Change and Stability in Chemical Systems are needed to address the questions in this prompt (Fig. 1.S2). Finally, the crosscutting concepts of Cause and Effect, Structure and Function, and Stability and Change provide a broad frame for how students could productively approach the problem (Fig. 1.S2). Because this item has the potential to engage students in the scientific practice of modeling using core ideas related to chemistry underpinned by crosscutting concepts, this item is characterized as 3D.

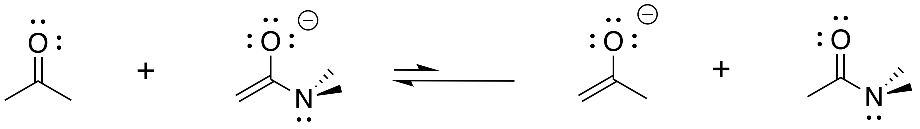
Simple model systems such as the system illustrated below enable inferences about the relative acidity of alpha positions on different carbonyl containing compounds. In this case, the product system is lower in energy than the reactant system and thus the equilibrium is product-favored.



Explain why the product system in the example above is lower in energy than the reactant system. Make explicit reference to how structural features of each system contribute to the relative energy of that system. An image of the N-atom lone pair is provided.



If the amide nitrogen in the system depicted in part A is rotated 90°, the energy difference between reactant and products systems changes (shown below). ***For the purposes of this problem, assume the amide nitrogen is locked in place and cannot rotate.*** Explain why the product system shown below is now higher in energy than the reactant system. An image of the N-atom lone pair is provided.



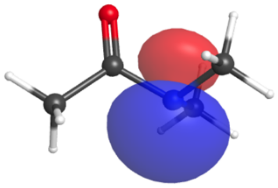
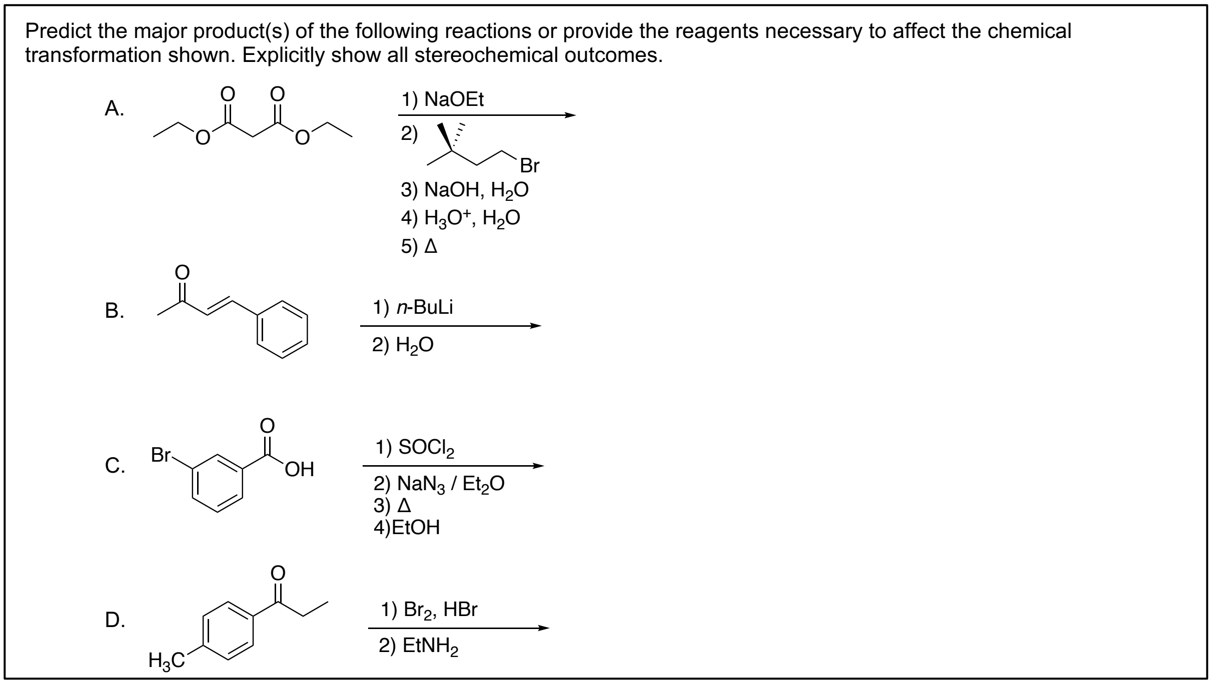


Figure 1.S1. An example of a 3D prompt taken from one of the exams administered during the Spring 2022 semester. This item has the potential to engage students in the scientific practice of Developing and Using Models using core ideas of Electrostatic and Bonding Interactions, Atomic/Molecular Structure and Properties, Energy, and Change and Stability in Chemical Systems as framed by the crosscutting concepts Cause and Effect, Structure and Function, and Stability and Change.

Scientific Practice	Core Ideas	Crosscutting Concepts
Developing and Using Models - Question gives an event, observation, or phenomenon for the student to explain or make a prediction about. - Question gives a representation or asks student to construct a representation. - Question asks student to explain or make a prediction about the event, observation, or phenomenon. - Question asks student to provide the reasoning that links the representation to their explanation or prediction.	**Electrostatic and Bonding Interactions** Attractive and repulsive electrostatic forces govern noncovalent and bonding (covalent and ionic) interactions between atoms and molecules. The strength of these forces depends on the magnitude of the charges involved and the distances between them. **Atomic/Molecular Structure and Properties** The macroscopic physical and chemical properties of a substance are determined by the three-dimensional structure, the distribution of electron density, and the nature and extent of the noncovalent interactions between particles. **Energy** Energy changes are either the cause or consequence of change in chemical systems, which can be considered on different scales and can be accounted for by conservation of the total energy of the system of interest and the surroundings. **Change and Stability in Chemical Systems** Energy and entropy changes, the rates of competing processes, and the balance between opposing forces govern the fate of chemical systems.	**Cause & Effect** Question provides at most two of the following: 1) a a cause, 2) an effect, and 3) the mechanism that links the cause and effect, and the student is asked to provide the other(s). **Structure & Function** Question asks the student to predict or explain a function or property based on a structure, or to describe what structure could lead to a given function or property. **Stability & Change** Question asks the student to determine 1) if a system is stable and provide the evidence for this, or 2) what forces, rates, or processes make a system stable (static, dynamic, or steady state, or 3) under what conditions a system remains stable, or 4) under what conditions a system is destabilized and the resulting state.

Figure 1.S2. Descriptions of the specific scientific practice, core ideas, and crosscutting concepts that the assessment item in Figure 1.S1 has the potential to elicit. These descriptions are taken from the 3D-LAP Protocol[1].

An example of a non-3D assessment item is shown in Figure 1.S3. Note that students are only asked to provide the products of the reactions. They do not need to explain how the products form or why they are formed in greater amounts compared to other potential products. Therefore, this assessment item does not meet the criteria for any of the scientific practices found in the 3D-LAP, so it cannot be considered 3D.



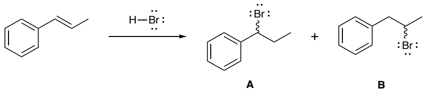
Predict the major product(s) of the following reactions or provide the reagents necessary to affect the chemical transformation shown. Explicitly show all stereochemical outcomes.

Figure 1.S3. An example of a non-3D prompt taken from one of the exams administered during the Spring 2022 semester. This item does not meet the criteria for potentially engaging students in any scientific practice, so it is not 3D.

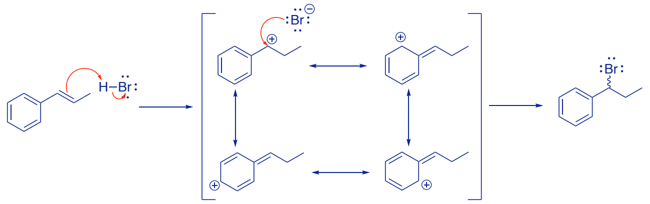
Canonical Answer to Assessment Item

The following figure provides a canonically correct answer to the assessment item administered in this study (Fig. 1.S4). The prompt is in black, and the answers are in blue.

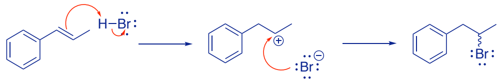
Consider the reaction between hydrobromic acid (HBr) and the alkene shown below. You read a claim that **Product B** is the major product of this reaction.



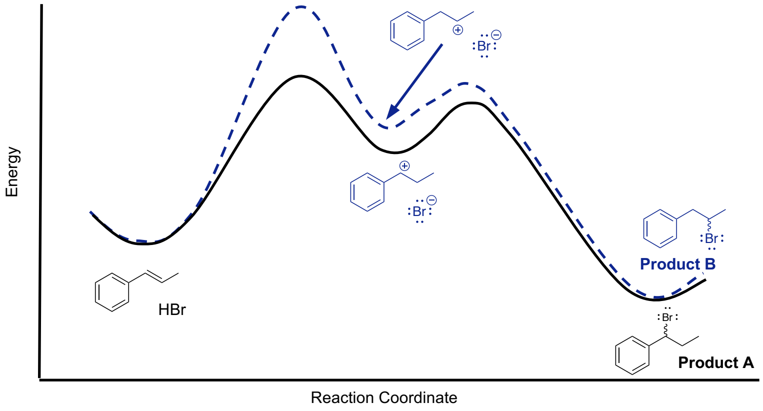
1. Draw a mechanism for the reaction above that leads to **Product A**.



2. Draw a mechanism for the reaction above that leads to **Product B**.



3. The potential energy surface for the pathway leading from reactants to **Product A** is drawn below. Using a dashed line, draw the potential energy surface for the pathway leading from reactants to **Product B** on the same axes. Label all intermediates and products.



4. Would you expect **Product B** to be the major product of this reaction? Explain your answer using the mechanisms and potential energy surface you drew in parts 1–3.

I would not expect Product B to be the major product of this reaction. This is a kinetically controlled reaction (i.e. irreversible), so the major product will be the one that has the lowest transition state energy for the rate-determining step. For this reaction, formation of the high energy carbocation intermediate is the rate-determining step. Using Hammond's Postulate, we can estimate the relative energy of the transition states using the relative energies of the closest energy species, which are the carbocation intermediates. The carbocation leading to Product A is lower in energy than the carbocation leading to Product B because its positive charge is delocalized via pi conjugation with the phenyl group whereas the positive charge on the carbocation leading to Product B is not delocalized via pi conjugation. Since the carbocation leading to Product A is lower in energy, its transition state must be lower in energy, so Product A must be the major product.

Figure 1.S4. The assessment item administered in this study with correct answers provided in blue.

Development of Explanation Coding Scheme

Coding of the student-generated explanations proceeded through five rounds, with Fleiss's Kappa calculated for each round (see Table 1.S1 for a summary). In the first round, KSD, CES, and NJE independently coded approximately 25% of the student assessment responses and obtained a value of 0.56 for Fleiss's Kappa. Discussion of disagreements led to several refinements to the coding scheme. For example, we decided that if a student referred to the activation energy or the energy barrier of the reaction, we would assume they were referring to the first elementary step of the reaction (*i.e.,* carbocation formation) unless they specified otherwise. Therefore, the response that follows was coded as a "3."

> I would not expect product B to be the major product because it has a higher energy of activation than product A and its carbocation intermediate is not delocalized via a pi conjugated system. Product A's intermediate is stabilized via pi conjugation

We also decided to base our decisions on the substance of students' responses rather than their use of specific terms. For example, in the following response, the student uses the term "transition state" but is clearly describing the structures of the two possible intermediates. Thus, we agreed to code this response as a "2" rather than a "3."

> Yes, Product B would be the major product of the reaction. The main difference between the formation of the two products is the position of the positive ion on the propyl carbon in the transition state. The transition state of product A has the positive ion closer to the benzene ring which can donate elections via pi bonding making the carbon less active. On the other hand, the transition state on product B has the positive charge further away from the electron rich benzene ring. This makes the molecule more active/electrophilic which would increase the rate of the nucleophile (bromine) attaching.

A few responses described the relative rates of the two competing reactions. While the rate of a reaction is related to the activation energy of its rate-determining step, we decided that a response needed to specifically describe the activation energy, energy barrier, or energy of the transition state to earn a code of "3." Therefore, the following response, which identified formation of Product A as the faster process but did not describe the transition state in any way, was coded as a "2."

> I would expect A to be the major product of this reaction because it has a lower energy intermediate. This will make the reaction go faster (kinetic) and end up with more product

> A than B. Product A's intermediate is lower in energy due to resonance as the positive charge on the carbocation can be donated around the ring for stabilizing effects that lower the energy.

Table 1.S1. Inter-rater agreement for coding of student explanations.

Round	Number of Responses	Fleiss's Kappa[a]
1	69	0.56
2	64	0.64
3	68	0.75
4	64	0.78
5	35	0.81

[a] A value of 0.7 or greater is considered good agreement.

In the second round of coding, we analyzed another quarter of the data set, and Fleiss's Kappa was increased to 0.64. Further clarifications and refinements were made to the coding scheme following discussion of discrepancies. For example, it was somewhat common for students to compare the energy pathways leading to each potential product without referring to any specific species (e.g., intermediate, transition state) along the pathways. In these cases, we decided to classify them as "2's".

> I wouldn't expect product B to be the major product since it would have a higher energy pathway. This is because the carbocation cannot delocalize with the pi conjugation from the ring as it would in product A.

Two more rounds of coding were performed, each consisting of approximately one quarter of the responses. Fleiss's Kappa for these rounds was calculated to be 0.75 and 0.78, respectively. Since a value of 0.7 or greater for Fleiss's Kappa is considered acceptable agreement, we made no further alterations to our coding scheme.[3] However, we still met to resolve any discrepancies. Thus, we ultimately achieved consensus codes for the entire data set.

Belatedly, we decided to only include the responses of students who had taken OChem I the previous semester (Fall 2020) in our data set so that we could also test for associations between OChem I enrollment and explanation code. We replaced the responses of students who had taken

OChem I during a different semester or at a different institution with those who had taken it during Fall 2020, which required us to code an additional 35 responses. Fleiss's Kappa for this portion of the data set was 0.81, indicating good agreement. As with the original data set, we discussed any discrepancies and arrived at consensus codes for all the replacement responses.

Chapter 2

Students’ Conceptual and Epistemic Resources for Predicting Reaction Mechanisms in a Big-Idea Centered Introductory Organic Chemistry Course

This work was conducted in collaboration with Ryan L. Stowe.

Abstract

Recognition that chemistry courses are often perceived by students as bundles of disconnected facts and algorithms has led to many initiatives to define “big ideas” central to the study of atomic/molecular behavior. Historically, these lists of ideas have consisted of concepts expert chemists thought all students should know. There is growing realization in the higher education community that science learning environments should emphasize making sense of phenomena rather than simply “knowing stuff”. Accordingly, “big ideas” should be those conceptual tools that allow students to predict and explain a wide variety of chemical phenomena. Here, we seek to describe “big ideas” in terms of small-grain intellectual resources that students commonly rely on to construct explanations for the atomic/molecular cause for perplexing phenomena. Using data from think aloud interviews focused on construction of predictive models, we explore the small-grain knowledge elements that were activated in moments where students were uncertain as to the “right answer”. Our analysis entailed 1) operationalizing “moments of uncertainty” in terms of activation of epistemological resources inferred from interview audio and transcripts and 2) characterizing the conceptual resources activated during these moments of uncertainty. Students commonly invoked conceptual tools related to electrostatics and energy when figuring out the likely outcome for a complex phenomenon. To support inferences as to why students expected these conceptual tools to be productive for sensemaking, we examined the intellectual work emphasized and rewarded on the assessments given in their organic chemistry course using the 3D-Learning Assessment Protocol. We found that course assessments placed substantial emphasis

on use of ideas related to electrostatics and/or energy to explain and/or model phenomena. Implications from this research suggest that, if students are to perceive certain concepts as broadly useful for explaining chemical phenomena, they should be consistently engaged in and rewarded for using those concepts to figure out how and why increasingly complex phenomena occur.

Introduction

Modern chemistry coursework has all too often become bloated with disconnected topics—a mile wide and an inch deep (Cooper, 2010; Lloyd & Spencer, 1994; Nameroff & Busch, 2004). Accordingly, "the typical course often appears to the novice as a disjointed, brisk trot through a host of unrelated topics" (Cooper, 2010, p. 231). In an effort to combat the sprawling, disjointed nature of most chemistry curricula, there have been a number of calls to structure high school and college chemistry instruction around large-grain ideas (Table 2.1). Among the first efforts to streamline the content of college chemistry courses were Gillespie (Gillespie, 1997) and Atkins' (Atkins, 1999) definition of "great ideas" that, "form the basis of modern chemistry" and that "every high school and college course should include" (Gillespie, 1997). There is a fair degree of overlap between both lists of "great ideas" with both including "bonding" in some form as well as substantial emphasis on energy. No literature from the learning sciences is cited in either Atkins' or Gillespie's manuscripts detailing "great ideas"—it is thus reasonable to infer that there was no theoretical basis for either list.

Table 2.1. Big idea lists.

Gillespie (1997)

1. Atoms, Molecules, and Ions
2. The Chemical Bond
3. Molecular Shape and Geometry
4. Kinetic Theory
5. Chemical Reaction
6. Energy and Entropy

Atkins (1999)

1. Matter consists of about 100 elements
2. Elements are composed of atoms
3. The orbital structure of atoms accounts for their periodicity
4. Chemical bonds form when electrons pair
5. Shape is central to function
6. Molecules attract and repel each other
7. Energy is blind to its mode of storage
8. Reactions fall into a small number of types
9. Reaction rates are summarized by rate laws

ACS Exams Institute (2012)

1. Atoms
2. Bonding
3. Structure/Function
4. Intermolecular Forces
5. Chemical Reactions
6. Energy and Thermodynamics
7. Kinetics
8. Equilibrium
9. Measurement and Data
10. Visualization and Scale
11. Systems Thinking

CLUE (2013)

1. Electrostatic and Bonding Interactions
2. Atomic/Molecular Structure and Properties
3. Energy
4. Change and Stability in Chemical Systems

Recognition that expert knowledge appears to be organized around fundamental disciplinary ideas (National Research Council, 2000) prompted several 21st century reform efforts to attempt their own definition of "big ideas" in chemistry. Prominent initiatives include the Advanced Placement reinvention project (The College Board, 2014), the American Chemistry Society's Examinations Institute's Anchoring Concept Content Maps (Holme & Murphy, 2012; Murphy et al., 2012), and the general chemistry curriculum Chemistry, Life, the Universe, and Everything (Cooper, Posey, & Underwood, 2017). As shown in Table 2.1, there are many commonalities between the big ideas arrived at by each of these initiatives—energy features prominently in all lists, as does structure and bonding. However, lists differ markedly on the character of what they categorize as a "big idea". For example, "bonding" and "energy" are required to explain many phenomena, while “measurement and data” and “visualization and scale” have little to no explanatory power. The different kinds of things that populate lists of “big ideas” stand as evidence that the chemistry education community does not have consensus on the essential character of such ideas.

Demarcation of concepts fundamental to the study of chemistry has historically been focused on ideas experts deem “extremely important”. There has not, until relatively recently (Cooper & Klymkowsky, 2013; Cooper et al., 2017; The National Research Council, 2012), been a concomitant focus on explicitly defining how students should use their knowledge to engage in practices characteristic of work in science. There is growing consensus in the science education and discipline-based education research communities that science learning environments should engage students in explaining and modeling aspects of natural and designed worlds (Cooper et al., 2015; Schwarz et al., 2016; The National Research Council, 2012). Such engagement has the potential to position students as knowers and doers of science and illumine the broad utility of scientific ways of knowing. Chemistry education research scholars are beginning to operationalize “sensemaking” in college contexts. For example, pioneering work by Cooper and colleagues suggests that courses structured around the use of fundamental concepts to make sense of successively more complex systems supports students in making sense of atomic emission spectra (Minter, 2019), acid-base reactions (Cooper et al., 2016; Crandell et al., 2019), phase changes (Noyes & Cooper, 2018; Stowe et al., 2019), and dissolution (Judd, 2018). Focus on making sense of phenomena requires refashioning “big ideas” from concepts experts think are important to ideas that students find broadly useful in explaining why things happen the way they do. The National Academies’ consensus study *A Framework for K-12 Science Education* makes this shift explicit by stating that disciplinary core ideas should, “Provide a key tool for understanding or investigating more complex ideas and solving problems" (The National Research Council, 2012, p.31). However, stating that "big ideas should be powerful sensemaking tools" does not clarify what it means for a student to weave such ideas together as they construct and critique explanations and models for phenomena they do not yet understand. How can we know when students are constructing an explanation *de novo* instead of reciting a memorized passage? What are the qualitative signatures of students drawing on large-grain ideas as they make sense of observable events? How do the ideas students expect to be productive sensemaking tools relate to those ideas emphasized and rewarded on assessments? How can we support students in internalizing the broad utility of disciplinary "big ideas"? If a central goal of chemistry learning environments is to support molecular level sensemaking, we must have consensus and clarity on what it means for students to leverage fundamental ideas in crafting explanations and models across contexts.

Research Questions

Here, we seek to describe "big ideas" in terms of clusters of small-grain intellectual resources students can use to predict and explain perplexing phenomena in terms of atomic/molecular behavior. Our perspective foregrounds the utility of certain conceptual tools for figuring out how and why things happen rather than the opinions of domain experts as to which ideas are "extremely important". This shift is needed if we are to position chemistry-enrolled students as knowers and doers of science instead of receivers of facts. We demonstrate how "big ideas" may be operationalized as clusters of small-grain knowledge elements by analyzing the resources organic-enrolled students call on and connect when asked to explain complex chemical phenomena. Our analysis focuses on moments when students' verbal and paraverbal cues signal that they are uncertain of the "correct answer". Subsequently, we examine association between the intellectual "heavy lifting" emphasized in assessments and conceptual resources commonly cued in an interview setting. We argue that course assessments help structure student expectations as to which conceptual tools will be useful in explaining chemical phenomena described by a written prompt. The findings of this study can inform future work on supporting and assessing atomic/molecular sensemaking grounded in "big ideas".

Our work was guided by the research questions:

1) How can one infer that students are constructing, rather than recalling, an explanation for a chemical phenomenon?
2) What conceptual tools do students call to mind and connect to make sense of unfamiliar phenomena?
3) How does the intellectual work emphasized and rewarded on course assessments relate to the conceptual tools students activate when figuring out the cause for perplexing phenomena?

Theoretical Framework

In order to elicit evidence that students consider particular ideas as powerful sensemaking tools, we must consider what is known about the character of student knowledge of chemistry. For many years, studies in chemistry education research operated off an implicit theory of cognition in which students were thought to possess coherent and stable "wrong theories" that could be detected by "wrong" answers on concept inventories and overridden by instruction (Cooper & Stowe, 2018). This perspective was an outgrowth of Strike and Posner's work on conceptual change published in the 1980s (Posner et al., 1982). More modern studies have found evidence that students' knowledge of atomic/molecular behavior is often not coherent or stable and is better modeled as a dynamic conceptual ecology in which small-grain knowledge elements are strung together *in situ*. The instability of students' chemistry knowledge was persuasively demonstrated by Cooper and colleagues via a qualitative study in which students were asked to pick which substance in a pair would have the higher boiling point and to explain their choice (Cooper, Corley, & Underwood, 2013). Students employed inconsistent and idiosyncratic reasoning that varied according to the structure of the substances being compared. Taber and Garcia-Franco also observed such inconsistent reasoning when high school students were asked to explain observable phenomena in terms of atomic/molecular behavior (Taber & García-Franco, 2010).

The "pieces of knowledge" activated by students in Cooper and Taber's studies may be considered "conceptual resources" drawn from formal instruction as well as personal experience (Hammer et al., 2005; Redish, 2004; Scherr & Hammer, 2009). Individual resources are not always inherently "right" or "wrong" but rather may be activated in more or less appropriate ways within a given context (Sayre & Wittmann, 2008). For example, the notion that "more effort begets more result" (called Ohm's p-prim by diSessa; diSessa, 1993; Hammer, 2000) may be invoked appropriately to explain chemical phenomena (*e.g.*, stronger bonds require more energy to disrupt), or inappropriately (*e.g.*, the molecule with more oxygens always has the higher boiling point). Resource Theory grew out of diSessa's "Knowledge in Pieces" perspective on conceptual change (diSessa, 1988). Due to this lineage, many "resources" described in the literature can be considered "primitives" – that is, thoughts that cannot be consciously reduced down to constituent parts by an individual (diSessa, 1993; Sherin, 2001). However, more modern literature has broadened the meaning of "resource" to include constructs of a variety of grain sizes (Sayre & Wittmann, 2008). This work defines "resources" as "small reusable pieces of thought" (Hammer, 2000) that may be active or inactive in a given moment and may link to other resources deemed complimentary for addressing a particular task (Sayre & Wittmann, 2008). Brown proposes a model in which students' conceptions emerge dynamically from interactions of conceptual resources. According to this dynamic systems perspective, it may not be possible to demarcate between a "conceptual system" and a "conceptual fragment" (Brown, 2014). Here, we do not attempt to make claims as to whether an utterance represents activation of a primitive or activation of some larger-grain emergent structure. Our perspective on what makes a "resource" encompasses both. Our focus is characterizing the conceptual resources students consider productive in explaining perplexing chemical phenomena and how those align with "big ideas" rewarded by learning environment designers.

The conceptual resources students consider "productive" are influenced by their sense of "what is going on" in a particular scenario – that is, their *frame* (MacLachlan & Reid, 1994; Scherr & Hammer, 2009; Tannen, 1993). The process of framing is shaped by past experiences and expectations of what will be required in the future. A significant amount of research in physics education has been dedicated to *epistemological framing*, or a students' answer to "How should I approach knowledge?" (Scherr & Hammer, 2009). For example, qualitatively different stances on knowledge and knowledge construction have emerged from studies of physics students' problem solving (Chari et al., 2017). Crucially, framing is a dynamic and context sensitive process (Tannen & Wallat, 1987). Students' sense of appropriate knowledge construction work may change moment-to-moment and be either stabilized or destabilized by various factors in their learning environment (Berland & Hammer, 2012). Thus, students' sense of the nature and appropriate use of knowledge is best considered as a localized coherence of *epistemological resources*. As with "resources" generally, *epistemological resources* are re-usable, fine-grained knowledge elements that may be activated and connected in-the-moment. Epistemological resources proposed in the literature include ideas that relate to the nature of knowledge (Hammer & Elby, 2002), forms of knowledge product (Collins & Ferguson, 1993)), processes of knowledge construction (Russ &

Luna, 2013) and the goals of knowledge construction activities (Berland & Crucet, 2016). To summarize, those conceptual tools students activate and link in response to a scientific phenomenon are affected by their view of the nature and appropriate use of knowledge in that moment, which is a function of their past experience.

"Big ideas", as defined by *The Framework*, may be thought of in terms of collections of conceptual resources that experts have found to be broadly useful in predicting, explaining, and/or modeling phenomena. Stated differently, conceptual resources related to "big ideas" have a high cuing priority for experts due to their utility in making sense of a wide variety of scenarios (diSessa, 1993). For example, various small-grain knowledge-elements related to Coulomb's law are useful in explaining the energetics of formation and breakage of intermolecular forces or the relative stability of charged species. Examples of these resources might be "opposite charges attract" or "more concentrated charge = more reactive", the latter of which is likely a restatement of Ohm's p-prim (diSessa et al., 2004).

Chemistry learning environments with sensemaking as a central goal should help students develop and activate core-idea related resources to make sense of a variety of phenomena in instructional and assessments settings. The purpose of such environments should not be to "know about big ideas", but to figure out how and why things happen the way they do. As a student's sense of "what is going on" in a given moment is structured by past experiences they deem similar (MacLachlan & Reid, 1994; Scherr & Hammer, 2009; Tannen, 1993), we hypothesize that there will be substantial overlap between the conceptual resources students consider useful and those ideas rewarded on assessments. In an environment focused on weaving together big ideas to make sense of phenomena, students should receive a substantial amount of course credit for constructing and critiquing explanations and explanatory models for observable events. The learning environment context for this work may be considered a prototype of what a sensemaking-focused organic chemistry learning environment might look like.

Methods

Participants and Setting

Participants for this study were students enrolled in a second-semester organic chemistry course at a research-intensive university in the Midwestern United States. At this university, approximately 1,000 students enroll in organic chemistry per semester. The course study participants were enrolled in is intended for non-chemistry STEM majors. Each course section meets as a large (~350 student) group for 150 minutes/week. Enrolled students also have the opportunity to engage in weekly discussion meetings which emphasize construction of models, arguments and explanations under the guidance of a teaching assistant, and complete similar open-ended homework assignments several times a week. Interviewees were volunteers who were solicited by email near the end of the second semester of the course (N = 12). As our goal was characterizing the conceptual tools successful students use when making sense of complex unfamiliar phenomena, interviewees were selected from among the highest achieving volunteers, as measured by their grades on the first two mid-term exams. Informed consent was sought and obtained from all study participants prior to commencement of the interview in accordance with the Institutional Review Board protocol approved for this study. Of the 12 participants, 5 were male and 7 were female. All students interviewed were very successful in the course and went on to earn an A or a B – their responses thus represent the “best-case scenario” in terms of resources activated to figure out the cause for perplexing phenomena. Additionally, all participants intended to take biochemistry following completion of organic chemistry. Table 2.2 lists each participant by pseudonym, sex, major, grade in both semesters of organic chemistry, and the prompt sequence used to structure their interview.

Table 2.2. Demographic and achievement characteristics of interviewed students listed alphabetically by pseudonym.

Pseudonym	Gender	Major[a]	O-Chem 1 GPA[b]	O-Chem 2 GPA[b]	Prompt[c]
Adam	Male	Physiology	4.0	4.0	M-2
Alex	Male	Neuroscience	4.0	4.0	L
Aurelia	Female	Microbiology	4.0	3.5	M-2
Austin	Male	Biochemistry & Molecular Biology	4.0	3.5	L
Christy	Female	Human Biology	4.0	4.0	M-2
David	Male	Human Biology	4.0	4.0	M-1
Kim	Female	Neuroscience	4.0	4.0	L
Matt	Male	Physiology	4.0	3.5	L
Megan	Female	Human Biology	4.0	3.5	L
Melissa	Female	Microbiology	4.0	4.0	L
Rachel	Female	Zoology	4.0	4.0	L
Sadie	Female	Human Biology	4.0	3.0	M-1

[a] Students' declared major at the end of the second semester of organic chemistry.

[b] GPA is reported on a 4-point scale.

[c] M-1 = variant 1 of more scaffold prompt, M-2 = variant 2 of more scaffolded prompt, L = less scaffolded prompt (see Figure 2.1).

[d] Student received prompts marked with an asterisk (*) depicted a trans-esterification under neutral conditions with alcohol nucleophiles. These prompts are labeled Variant 1 in Figure 2.1.

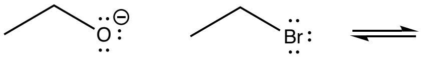
All who were interviewed for this work enrolled in two semesters of a transformed organic chemistry curriculum known as Curriculum A (Cooper et al., 2019). This course is designed around scaffolded progressions of "big ideas" identified as part of a general chemistry transformation effort at University A (Cooper et al., 2017). These include: "electrostatic and bonding interactions", "atomic/molecular structure and properties", "energy", and "change and stability in chemical systems". As with the core ideas defined in *The Framework*, Curriculum A "big ideas" were selected due to their broad predictive and explanatory power. Students are to use these ideas throughout both semesters of Curriculum A to make sense of increasingly complex systems. We will examine how "making sense of increasingly complex systems" was operationalized in Curriculum A via analysis of course formative and summative assessments. We anticipate that student expectations as to the conceptual tools useful for making sense of unfamiliar phenomena will be informed by those resources useful in making sense of contexts assessed in Curriculum A. Importantly, while the course was designed to emphasize use of ideas experts find useful for sensemaking, there is as yet no evidence that Curriculum A-enrolled students tend to use "big idea"-related conceptual resources to predict and explain likely outcomes in novel contexts.

Interview Protocol

Semi-structured interview protocols were used to elicit evidence of students' epistemological framing (RQ 1) as well as the conceptual resources students used when making sense of chemical phenomena (RQ 2). These protocols were centered around construction and justification of models to predict the outcome of chemical processes. The prompts meant to initiate model construction are given in Figure 2.1. In all prompts, students were to leverage a curved arrow mechanism to predict and explain the outcome of a set of reactions. Mechanisms of this type are understood by expert chemists to convey donor-acceptor interactions in which electrons are donated from filled to empty orbitals. It is well-established in the literature that proper arrow depiction does not necessarily indicate an understanding of the phenomenon being depicted (Grove et al., 2012). For this reason, students were prompted by the interviewer to both predict the major product that would result from a reaction system and to explain why.

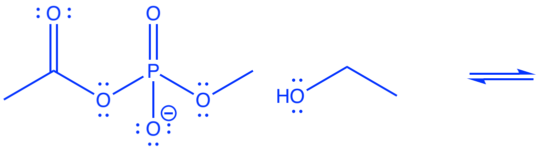
Two sequences of prompts were used in this study, a "more scaffolded" sequence consisting of prompts A-C given in Figure 2.1, and a "less scaffolded" sequence consisting only of prompts A and C given in Figure 2.1. By "scaffolding", we refer to prompt features that focus learners on "critical features"(Reiser & Tabak, 2014; Wood et al., 1976) salient to productively addressing the task. Both prompt sequences began with a straightforward bimolecular substitution (S_N2) reaction analogous to many systems explored in both semesters of Curriculum A. This prompt served chiefly to familiarize students with verbalizing their thoughts – a notoriously unnatural experience (Herrington & Daubenmire, 2014). The prompts shown in Figure 2.1B and C were meant to engage students in making sense of complex, biologically relevant phenomena they had not seen previously – transesterification of phosphate esters. Biological relevance was chosen as a focus due to the intention of all study participants to enroll in biochemistry and the intention of many to pursue careers in the healthcare sector. The system students were asked to model in Figure 2.1C was meant to model attachment of an amino acid to tRNA. In biological contexts, this process is catalyzed by aminoacyl-tRNA synthetase (Agarwal & Nair, 2012), though the principles of the reaction may be understood in the absence of the enzyme. A canonical explanation for the formation of the major product observed upon treating a phosphate ester with an alkoxide is presented in the supplemental information for the interested reader. In this contribution, we seek to characterize the small-grain knowledge elements students activate and connect when attempting to predict and explain a phenomenon they do not yet understand. Accordingly, the impact of prompt scaffolding is not the focus of the present study.

A. Using a curved arrow mechanism, predict the major product that would result from adding sodium ethoxide to a solution of ethyl bromide in DMF (ethoxide and ethyl bromide are shown below).

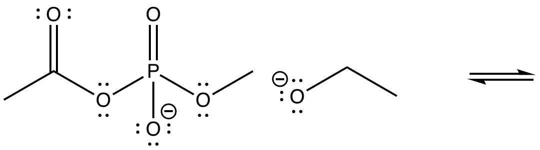


B. Using a curved arrow mechanism, predict the major product that would result from addition of ethoxide (or ethanol) to a solution containing the phosphate ester shown below.

Variant 1:

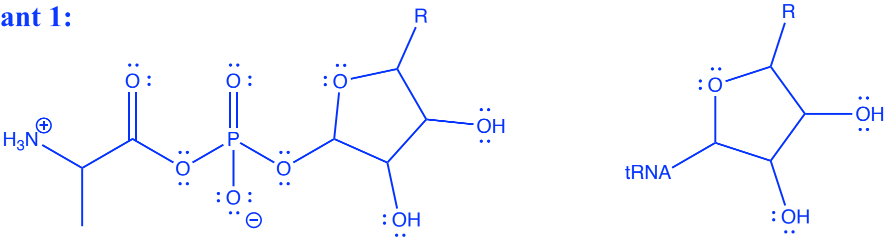


Variant 2:



C. Many physiologically important molecules are synthesized in part through the reaction of species similar to those shown below. Using a curved arrow mechanism, predict what might happen when these two molecules collide.

Variant 1:



Variant 2:

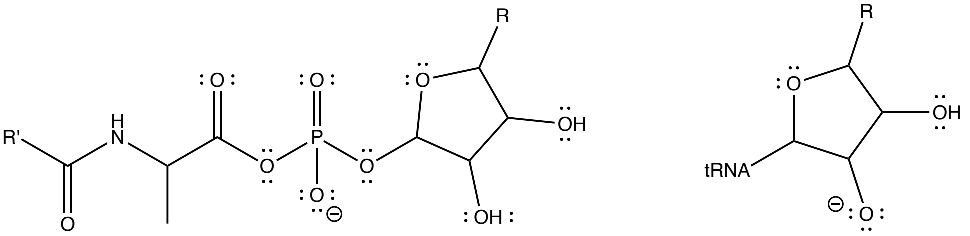


Figure 2.1. Interview prompts administered to students. The more scaffolded version consisted of parts A–C while the less scaffolded version consisted of only parts A and C. A few students received Variant 1 of parts B and C (shown in blue), which involved a neutral alcohol, while the rest received Variant 2 of parts B and C (shown in black), which involved a negatively-charged alkoxide.

As students may frame an interview in a different manner than the interviewer, which may affect those resources activated and linked in the interview context (Russ et al., 2012), participants were advised from the outset that our principle aim was to elicit evidence of how they think through problems, not evaluate the correctness of responses. Statements such as, “you needn’t approach this as an exam”, “we’re just interested in your process”, and “try not to feel like this is a high-pressure situation because it really isn’t” were said by the interviewer in an attempt to stabilize a frame in which reflective reasoning was seen as productive rather than maximally efficient problem solving. Relatedly, in situations where students expressed that a particular answer was “good enough” for an exam, the interviewer would respond that the interview was not a high-stakes test and that students should aim for the best possible model-based prediction, without worrying about time pressure or correctness.

Data Collection

Interviews

The second author conducted all twelve interviews analyzed for this study. The length of the interviews varied from 15 to 30 minutes depending on the amount of information the students provided. Interview audio was redundantly recorded using both a digital voice recorder and a *LiveScribe* pen, which can replay both audio and student drawings in real-time (Linenberger & Bretz, 2012). Prompts were printed on dot-matrix paper and annotations were made using a *LiveScribe* pen. This enabled simultaneous capture of student writing and dialogue. Student-constructed representations were collected at the end of each interview.

Assessment Forms

To get a sense as to the emphasis Curriculum A places on use of “big ideas” in atomic/molecular sensemaking, all high stakes assessments and low stakes discussion activities taken by the interviewees over both semesters of the course were collected and analyzed. Our high stakes assessment dataset was comprised of: three mid-term exams and the final exam given during the first semester course together with three mid-term exams given during the second semester course. The final exam given during the second semester of Curriculum A was a test authored by the ACS Examinations Institute, which has little potential to elicit evidence of student ability to explain phenomena (Stowe & Cooper, 2017). Our low stakes assessment dataset consisted of 28 discussion activities, which were completed by small groups of Curriculum A-enrolled students during each week of instruction.

Data Selection

As seven students received the "less scaffolded" prompt variant and five received the "more scaffolded" variant, interviews included responses to 29 individual prompts. Intelligible recordings were obtained from each interview and thus all prompt responses were included in our dataset. As the emphasis of the present study is the conceptual tools students activate when predicting and explaining perplexing phenomena, dialogue that did not pertain to relating big ideas to phenomena was omitted from our analysis (*e.g.*, conversation concerning the logistics of the interview, narration of actions, reflections on the course). As an example of the sort of dialogue omitted, consider the common occurrence of students narrating what they are drawing without subsequently justifying why that drawing is reasonable (*e.g.*, "an arrow from a lone pair on oxygen should go to carbon"). Absent additional elaboration, describing the mechanics of drawing a representation does not support inferences as to what, if any, conceptual resources underpin that representation. Narration of actions followed by justification as to why those actions are reasonable was included in our analysis (*e.g.*, "I can represent donation of electrons from the negative oxygen to the partial positive carbon by drawing an arrow from a lone pair on oxygen to carbon").

Dialogue pertaining to construction of explanations and/or explanatory models was sub-divided into segments representing complete thoughts. These segments were typically one to three sentences long. The following excerpt from Kim illustrates what we mean by a "segment":

> Segment 1: Um, I guess just looking at what it's attached to. So, this is attach-, like a carbon bonded to an oxygen.
> Segment 2: Oh, well, I guess there's a carbonyl right there too actually.
> Segment 3: So I didn't look hard enough.
> Segment 4: But I would look at what it's attached to, so like this carbon, how it's attached to an R group, like another C, like carbon, it wouldn't be that, like there'd be no really charge here because they're the same amount of electronegativity. But, with these carbonyls, since this has got two bonds to oxygen, that makes this have even more partial positive.

The first two sentences concern the attachment of atoms; this passage begins with a general statement, followed by a statement specific to this situation. Thus, these two sentences were treated as one segment (Segment 1). In the following sentence, Kim paused her explanation to note a feature of the prompt, so this was considered a new segment (Segment 2). The next sentence (Segment 3) is a reflective statement that did not pertain to reasoning about the reaction, so it was not coded. The remaining two sentences focus on the same topic (bond polarity) and were treated as one segment since the latter sentence provides a counterexample to the example given in the former sentence.

Data Analysis

Each segment of dialogue was characterized via two layers of coding. First, epistemological resources that could be inferred from dialogue segments were coded via adapting the code definitions put forth by Hammer and Elby (Hammer & Elby, 2002). Second, we inductively coded for the small-grain reusable knowledge elements that represented conceptual resources students activated when they made sense of phenomena presented by prompts. Transcript coding was facilitated by NVivo 12 for Mac (QSR International Pty Ltd., 2018).

Coding Epistemological Resources

To characterize aspects of students' sense of what was going on with respect to knowledge, each segment of dialogue was examined for evidence of the knowledge type and epistemic activity students appeared to engage in (Table 2.3). Especial attention was given to paraverbal cues (pauses, tone changes, transitions etc.), as these often served to demarcate between epistemic activities. For example, engaging in the epistemic activity *causal storytelling* was marked by smoothly relaying a causal explanation that was likely previously learned. By contrast, if a student was *forming* an explanation from prior knowledge, they were dredging up and connecting ideas in the moment to make something new – this was characterized by halting speech and tone changes that signify uncertainty. It was crucial to listen to interview audio recordings while coding the transcripts to detect changes in speech patterns.

To build a case for the reliability of our coding of epistemological resources, the first and second authors independently coded two of the transcripts and subsequently met to discuss coding inconsistencies. The codebook was refined in response to this discussion. For example, it was decided that a statement did not have to explicitly name the source of information to be coded as *propagated* knowledge. Independent coding of transcripts and subsequent discussion occurred twice until acceptable agreement was reached (Cohen's kappa of 0.70). The first author then coded the remaining transcripts. In total, consensus codes were reached for five of the twelve interviews while seven interviews were coded solely by the first author.

Coding Conceptual Resources

As the goal of RQ 2 was identification of conceptual resources cued when explaining a phenomenon, our analytical approach began with detailed consideration of individual cases using open coding (Strauss & Corbin, 1990) to identify phrases and/or drawings that appeared to reflect conceptual tools activated and connected to make sense of phenomena. Our analysis was refined via iteratively revisiting the interview transcripts and student drawings using a process of constant comparison between the two authors' codes (Glaser, 1978; Taber, 2000). Our identification of "resources" is based on the definition of Sayre and Wittman (Sayre & Wittmann, 2008) that they are both "individual" – that is not necessarily coupled to a particular context – and "reusable" – that is they might be productively used in many contexts. We do not assume a particular grain size of "resource" in our analysis – those items identified thus range from "primitives" to fairly large-grain ideas likely composed of smaller nested resources. These larger-grain knowledge elements may be considered structures emergent from dynamic interactions between smaller-grain elements (Brown, 2014). Interview transcripts were examined segment-by-segment for words or phrases indicating use of discrete, reusable conceptual tools – these were tagged with a descriptive code. These tools consisted both of primitives (*e.g*., "more means more") and ideas related to more sophisticated concepts (*e.g*., "opposite charges attract"). In instances where resources related to several "big ideas" were invoked, multiple codes were used (*e.g*., "energy" and "opposite charges attract" would describe the phrase, "attraction between oppositely charged regions of two molecules lowers system potential energy"). Use of multiple codes allows us to infer how students connected multiple conceptual resources to explain aspects of the phenomena they were examining. Transcripts were sub-divided by phenomenon (*e.g*., "bimolecular substitution", "simple trans-esterification", "complex trans-esterification") to aid in inferring which resources students used across contexts.

To build a case for the reliability of our conceptual resource coding scheme, both authors coded two interviews (one structured around the more scaffolded prompt series and one structured around the less scaffolded prompts) and examined coding consistency paragraph by paragraph using the interrater-reliability tools within NVivo. Discussion of inconsistencies led to numerous refinements of our codebook. For example, it became apparent that many codes could be expressed as a statement as well as its inverse (*e.g.*, “concentrated charge = more reactive” can also be expressed as “delocalized charge = less reactive”). We settled on the five codes, described in Table 2.4, to describe the conceptual resources leveraged by students to reason across phenomena. Note that each resource represented by a code is reusable in the sense that it does not require linkage to a particular phenomenon. The two authors were able to independently code two interviews (~16% of the total dataset), using the final elaborated codebook, with a high level of consistency (Kappa = 0.8). Six of the twelve transcripts were jointly coded as part of establishing coding scheme reliability. For these six, a consensus was reached as to which transcript segments corresponded to activation of which resources. The first author coded the remaining six transcripts using our established codebook. In order that others, should they choose, can evaluate the validity of the claims we make from interview data, we have appended full, coded transcripts of all analyzed interviews in the supplemental information. We recognize that codes are claims, not data, and that it is important for the community to have the ability to double-check consistency between all claims that we make here and the dataset informing those claims (Hammer & Berland, 2014).

Assessment Analysis

We hypothesized that the conceptual resources students default to when asked to make sense of an unfamiliar phenomenon would relate to the intellectual “heavy lifting” emphasized on formative and summative assessments. Assessments send strong implicit messages to students as to the *true* focus of a learning environment (Crooks, 1988; Entwistle, 1991; Momsen et al., 2013; Scouller, 1998; Scouller & Prosser, 1994; Snyder, 1973). In order for a course to truly emphasize predicting, explaining, and modeling phenomena, atomic/molecular sensemaking must be allotted substantial points on course quizzes and exams. Stated differently, students' expectations regarding knowledge construction (that is, their frame) will be built from past experiences they deem similar to the present context (Tannen, 1979, 1993). We would therefore expect students to cue knowledge elements that have been useful for constructing explanations on assessments so long as they see the interview prompts as similar to assessment prompts. Analysis of each assessment leveraged the three-dimensional learning assessment protocol (or 3D-LAP; Laverty et al., 2016). This protocol provides criteria an assessment item must meet to have the potential of eliciting evidence that students are using what they know to predict, explain, and/or model phenomena. Importantly, fulfilling these criteria does not guarantee that a prompt will in fact elicit the evidence desired. The potential of each assessment item to elicit evidence of 3-dimensional learning was coded independently by both authors. Following this initial coding, the authors met and reached consensus on assignment of all codes. As the focus of this piece is the use of core ideas to explain and model phenomena, especial attention was given to which core idea(s) might be involved in addressing questions requiring an explanation or model. As will become obvious, core ideas almost always co-occur in atomic/molecular explanations and models. All assessments coded for this analysis are appended to as part of the supplemental information together with our consensus codes for each assessment item.

Results and Discussion

Constructing an Explanation In-the-moment is Qualitatively Distinct from Recall

Discerning the conceptual resources students find useful when explaining perplexing phenomena requires that we identify instances when students are constructing, rather than recalling, explanations for events. Literature related to students' epistemological framing suggests that students' perspective on "what is going on here" with respect to knowledge should differ depending on whether they regard a scenario as an opportunity to recall a “correct answer” or to construct an explanation for something not yet fully understood (Odden & Russ, 2019; Russ et al., 2012). Russ and colleagues found that students framing an interview as an oral examination saw “their task as producing a correct answer to a prompt or question in a clear or concise fashion” (Odden & Russ, 2019, p. 1058). The qualitative signatures of an oral examination frame included clearly and confidently delivering a (very likely recalled) answer to an interview question that often incorporates scientific vocabulary (Russ et al., 2012). Students constructing an explanation in the moment, from prior knowledge deemed useful, often speak in a halting fashion with many stops and starts (Russ et al., 2012). Students framing an interview activity as an opportunity to engage in inquiry also tend to use hedging language (*e.g*., I think, maybe). Recent work has sought to characterize the epistemological resources that are activated and connected to create a frame in-the-moment (Shar et al., 2020). These include knowledge elements pertaining to the type of knowledge (*e.g*., propagated, fabricated) and epistemic activity (*i.e*., how the knowledge was obtained). An oral examination frame initiated in response to a prompt to “explain why” may represent a local coalescence of resources including “knowledge as propagated stuff” and “causal storytelling”, as the student relays a previously learnt explanation for a phenomenon. An inquiry frame initiated in response to a similar prompt may come about, in part, via activation of “knowledge as fabricated stuff” and “forming” as a student draws from prior knowledge to stitch together an explanation on-the-fly.

Here, we demarcate between instances when students recalled a memorized response and instances when they constructed a response in-the-moment by characterizing aspects of students’ framing throughout our think-aloud interviews. Drawing from prior work published by Hammer and Elby (Hammer & Elby, 2002), we characterized activation of knowledge elements related to knowledge type and epistemic activity (Table 2.3). We argue that viewing an activity as an opportunity to recall facts will result in activation of knowledge type and epistemic activity resources in a manner distinct from the epistemological resources activated when students see a scenario as a time to engage in *in situ* explanation construction. We focused our analysis on resources related to knowledge type and epistemic activity because we found these knowledge elements straightforward to characterize and the activation of these resources sufficient to support claims as to whether students were recalling or constructing explanations.

Table 2.3. Themes that emerged from the interviews were summarized using the codes shown below. An example response described by each code is provided.

Code	Definition	Example
Knowledge Type		
fabricated	Stated knowledge is figured out by student using their prior knowledge	*"Because the oxygen will be able to hold onto those electrons easier than the nitrogen would because the nitrogen would just want to pick up a hydrogen or something because it's not as stable with extra electrons."*
propagated	Stated knowledge comes from outside source, such as a professor or textbook. Note that the student does not have to explicitly identify the outside source	*"We learned they have the same oxidation states. I think that was what we talked about in class."*
directly perceived	Stated knowledge is readily apparent or obvious because it is directly given on the page	*"I guess there's a negative over there too."*
intuited	Stated knowledge is determined from a gut-feeling or sense that it's true. It is usually indicated by the phrase "I feel"	*"I feel like this is the best way to be able to stabilize the oxygens and make the whole entire thing overall more stable than this."*
Epistemic Activity		
forming	Students is using their prior knowledge to reason about how to solve an unfamiliar problem.	*"So, I'm thinking I don't know which one I would pick for where there would most likely an attack, to be honest... Maybe this one? Er, maybe that one because it's less cluttered, it has less hindrance?"*
causal storytelling	Students explains why something happens by connecting a cause and an effect.	*"It's more electronegative, so it's pulling electrons away from that carbon, so with less negative electron charge, then this becomes more positive."*
accumulating	Student notices a feature of the prompt or recalls factual information.	*"A weak base is a good leaving group."*

Knowledge type refers to students' understanding of the nature of their knowledge; resources of this type that were relevant to our analysis include *knowledge as fabricated stuff* (built from students' previous knowledge), *knowledge as propagated stuff* (received from another person or authority), *knowledge as intuited stuff* (obtained from an inarticulate "sense"), and *knowledge as directly perceived stuff* (gained through use of the senses). Epistemic activity describes how students use their knowledge; of the many examples given, *accumulating*, *causal storytelling*, and *forming* were most applicable to our data. *Accumulating* describes instances where students are "collecting" information, whether by using one's senses, talking to another person, reading a book, or through some other process. *Causal storytelling* describes instances where students know the answer and are communicating their reasoning by relating causes and effects while *forming* describes instances where students are constructing an answer to an unknown problem in the moment. These are not exhaustive lists of epistemological resources but rather those that were most useful in addressing our research questions. It is important to note that framing is a dynamic process; a student may shift fluidly among several frames during a single interview. In our data set, we observed episodes of epistemological resource activation ranging in duration from a few seconds to several minutes. The following paragraphs will illustrate how we applied these codes to our data.

After receiving the written prompts (Fig. 2.1), most students began by verbalizing prompt features that caught their attention. These instances represented activation of *knowledge as directly perceived stuff* and *accumulating*. For example, Megan's first response to the complex trans-esterification prompt (Fig. 2.1C) was to note, "There's a negative charge on this oxygen." These instances were often followed up activation of one of two epistemological resource clusters*:* *knowledge as fabricated stuff* and *causal storytelling*, or *knowledge as fabricated* (or *intuited*) *stuff* and *forming*. If the student knew how a noticed feature related to a canonical answer they could recall, they drew on their prior knowledge (*fabricated*) to describe the cause for the phenomenon represented by the prompt (*causal storytelling*). For example, Rachel explained, "Because oxygen is negatively charged, and I know that the Br [bromine] has a delta minus and the carbon has a delta plus, that the negative is going to be attracted to the partial positive." Rachel drew on her prior knowledge of how charge is distributed between bromide and carbon and how charges interact to justify why oxygen would react with carbon. Rachel's explanation was delivered confidently and without hesitation, likely signifying that much of it was recalled rather than constructed in that moment.

All interviewed students experienced one or more moments where they were unable to recall and smoothly deliver a canonical answer to a prompt or interviewer question. During these instances, study participants iteratively proposed and connected up ideas they thought would be useful in addressing the task at hand. After identifying a negatively charged oxygen atom, Christy said, "I think it's going to attack the carbonyl 'cause this is partial negative and this is partial positive. I don't know where else it would go. Yeah, I think it's going to go like this..." Her use of the phrases "I think" and "I don't know where it else it could go" suggests that she did not know the outcome of this reaction and instead had to draw on her knowledge of how negative and partial charges interact (*fabricated)* to predict what would happen (*forming*). Note that, even though Christy included a cause-and-effect relationship in her reasoning, this transcript passage is better described as an instance of on-the-fly construction (that is, *forming*) rather than recall of a causal mechanism (*causal storytelling*). Paraverbal signatures captured by audio recordings, such as uncertain tones and halting speech, were very useful in distinguishing between the epistemic activities *forming* and *casual storytelling*. We also observed students attempting to figure out what would happen (*forming*) based on a gut instinct (*intuited*) rather than prior knowledge (*fabricated*). For example, Alex said, "So I see that there is an oxygen in both molecules with a negative charge on it, which I feel like would react with something." Alex sensed that negatively charged oxygens were involved in the reaction, although he did not articulate what prior knowledge gave him this sense.

A final cluster of epistemological resources that was frequently observed was *knowledge as propagated stuff* and *accumulating*. These codes describe instances where students were recalling information that they had been told by someone or had read in a textbook. For example, when Aurelia was asked why she thought the negatively charged oxygen made ethoxide a good nucleophile, she responded, “Umm... It's a strong-ish base. It's a strong base,” a statement recalled from class. As observed with the other clusters containing *accumulating*, these factual statements were often followed up by students explaining in their own words why the recalled fact made sense or how it related to the specific situation given in the prompt (*fabricated* + *causal storytelling*) or an attempt to determine what it meant or how to apply it (*fabricated* or *intuited* + *forming*).

When predicting the outcome of the first phenomenon given during the interview (Fig. 2.1A), most students activated the knowledge type/epistemic activity clusters *propagated* + *accumulating* and *fabricated* + *causal storytelling*. That is, students frequently recalled information they thought relevant to the prompt (*propagated* + *accumulating)* and used this information to deliver a recalled causal explanation (*fabricated* + *causal storytelling*). This was expected, as the prompt was meant to serve as a warmup to thinking aloud and foregrounded a phenomenon students had explained many times previously. However, follow-up questions by the interviewer occasionally affected students’ epistemological resource activation. For example, David readily identified the role of bromine as a leaving group in the bimolecular substitution reaction because it could “support the negative charge” (*casual storytelling*). However, when asked to explain why bromide could support a negative charge, his pattern of speech changed noticeably, from relatively smooth and fast-paced to halting and slower-paced, suggesting that he did not readily know how to answer the question and was instead constructing an explanation on the spot from prior knowledge (*forming*).

In contrast, most students were engaged in *forming* when predicting and explaining phenomena given by the other two prompts. Student dialogue related to making sense of the transesterification reactions (Fig. 2.1B and Fig. 2.1C) was often characterized by frequent long pauses, halting speech patterns, and most telling, consideration of multiple possible solutions. These verbal and paraverbal cues indicate an iterative process of calling to mind and connecting knowledge elements to build an explanation on-the-fly. As an example, consider how Matt began contemplating what might happen when a phosphate ester collides with a nucleophile:

> I'm thinking it will take off an alpha hydrogen, because those hydrogens are acidic. And then the bond from the alpha carbon and the alpha proton are going to go onto the alpha carbon, giving it a full negative charge. And then I was thinking that it was going to do something intramolecular, like this carbon was going to be attracted to something that was partial positive.

He voiced his reasoning for considering two possible first steps, which he then had to weigh as he constructed his prediction. These types of *forming* instances provide rich insight into the conceptual resources students utilized when asked to predict and explain the outcome of an unknown reaction. In the next section, we turn to a discussion of students' conceptual resources and conclude with an extended example from one student in which we compare the conceptual resources activated in *causal storytelling* instances with those activated in *forming* instances.

Students called to mind and connected knowledge elements related to electrostatics and/or energy when constructing explanations for perplexing phenomena

Before we may consider the conceptual resources students activated when constructing an explanation on-the-fly, it is necessary to define the small-grain knowledge elements that emerged from our interview data. Five conceptual resources were observed in most or all of the interviews: *charge=reactive*, *bond polarity*, *opposite charges attract*, *more charge=more reactive*, and *energy* (Table 2.4). The activation of each of these knowledge elements across contexts supports the argument that our codes represent "individual reusable thoughts," or resources. In the following paragraphs, we will describe each code in detail and offer examples of how the resource that code represents was used by students. We will finish by examining one student's reasoning in detail to illustrate a common sequence of conceptual resource activation and discuss how their framing of an interview scenario affected the ideas the student called to mind. Before we delve into these examples, it is important to note that we are not interested in students' abilities to use proper scientific vocabulary but rather the ideas underpinning their reasoning; thus, we accepted colloquial or imprecise verbiage in lieu of more precise language when coding. Past work by Crandell et. al. has shown that proper use of jargon is not associated with appropriate explanations for atomic/molecular phenomena (Crandell et al., 2020). Similarly, because we are characterizing the resources that *students* perceive as useful, we do not focus on the correctness of their reasoning, although we recognize that developing explanations in line with scientific canon is a major goal of higher education.

Table 2.4. Codes for conceptual resources that emerged from the interviews are summarized below. An example response described by each code is provided.

Code	Definition	Example Response
charge = reactive	Students use charge to identify reactive part(s) of a molecule.	*"So I guess the first thing that I notice is the negatives on the oxygens, they're not protonated. So I feel like those are going to be the starting point."*
bond polarity	Students discuss bond polarity to rationalize partial charges.	*"Because this bromine has a higher Zeff or effective nuclear charge. It's more electronegative, so it's pulling electrons away from that carbon, so with less negative electron charge, then this becomes more positive."*
opposite charges attract	Students describe a negative charge being attracted to a positive charge (or vice versa) or an atom with a negative charge reacting with an atom with a positive charge.	*"causing this negatively-charged pair of electrons on the oxygen with the ethyl to be attracted to the partial positive carbon"*
more charge = more reactive	More charge can mean either a greater magnitude of charge or a greater concentration of charge. Student connects a greater magnitude/concentration of charge with being more reactive/less stable or a smaller magnitude/more delocalized charge with being less reactive/more stable.	*"if you have a lot of negative charge in one place, it tends to be unstable." and "if you could delocalize the electrons between different atoms in the molecule, then it makes it more stable."*
energy	Student describes the stability or reactivity of a molecule or charge. Note that they do not have to talk about energy specifically (and most do not).	*"Square structures that are a little more unstable than, I mean obviously say like a cyclohexane, but it all matches up."*

Charge=Reactive

Identification of electron deficient and/or electron rich regions of molecules in a reacting system was very often the first step in students' sensemaking, which led to the development of the code *charge=reactive.* Every student used the presence of a charge to identify a reactive region on a molecule at least once, and it was commonly the first verbalized "individual reusable thought" expressed. David illustrated how charge can serve as a powerful cue to students as to where a reaction is likely to occur. In the bimolecular substitution problem (Fig. 2.1A), David immediately identified the reactive nucleophile as "the oxygen with a negative charge." In the simple transesterification reaction (Fig. 2.1B, Variant 1), he consciously applied the same strategy:

"Similar to the last one, I'm going to use the oxygen with a negative charge as the nucleophile in this case," referring to the oxygen bonded to phosphorus, the only atom with a formal negative charge present. David was then prompted by the interviewer to consider what would happen if instead the oxygen in ethanol was deprotonated, giving it a negative formal charge as well. He said, "In my opinion, we're going to have competing reactions." His use of *charge=reactive* is clearly evident; only oxygen atoms with formal negative charges cued him into the possibility of a reaction occurring. In the complex transesterification reaction, David demonstrated a slightly different use of *charge=reactive*. After proposing a reaction step and drawing the resulting molecule, the interviewer asked if anything else would happen. David responded, "Yes, definitely. Because we have the two negative charges there." To him, the presence of the negative charges indicated that the molecule would react further. As Caspari has observed, students may be heavily influenced by surface features such as formal charges (Anzovino & Bretz, 2015; Caspari et al., 2018; Galloway et al., 2017; Graulich & Bhattacharyya, 2017). Our findings are consistent with this work in that the presence of written charges prompted students to think about reactivity.

Explicit formal charges readily caught students' attentions, but they also used implicit partial charges to predict reactivity. For example, in both the simple bimolecular substitution reaction and simple transesterification reaction, Christy drew partial positive charges on the starting materials and provided her rationale for doing so: "When I first looked at [the problem], I didn't know what was going to attack what, but then [I] draw the charge, like the negative I knew would be attracted to this partial positive. It just kind of solidified my thoughts, I guess." It seems that Christy considered identifying implicit charges as a necessary first step for predicting and/or explaining reactivity. Christy was not the only student who looked for implicit partial charges; seven of the twelve students interviewed also added them to the starting materials in at least one of the reactions. Thus, *charge=reactive* encompasses not only explicit formal charges but implicit partial charges as well, and students seek out these implicit charges to help them predict and explain reactivity.

As partial charges were not explicitly depicted in the prompts, students used the resource *bond polarity* to predict and justify their presence. *Bond polarity* describes the distribution of electron density within a bond and derives from the interaction of negatively-charged electrons and positively-charged atomic nuclei. For example, Aurelia described the polarity of a carbon-oxygen bond to explain why the carbon atom had a partial (delta) positive charge: "Oxygen is really electronegative, so that means it's pulling electrons more towards it than the carbon in this bond, so it's going to put a delta positive on the carbon." Interestingly, most students seemed to use this resource unconsciously; they would describe or draw partial charges without verbalizing their reasoning. However, when prompted for an explanation by the interviewer, most students readily provided one. For example, after Kim stated that a carbon atom would have a partial positive charge, the interviewer asked her to explain why. She replied, "Because this bromine has a higher Z_{eff} or effective nuclear charge. It's more electronegative, so it's pulling electrons away from that carbon, so with less negative electron charge, then this becomes more positive." Thus, although Kim did not initially describe how she thought about *bond polarity* when she first used it, she was able to do so when the interviewer signaled to her that this depth of reasoning was desired. The example from Kim and several other students has implications for how assessments are written, which we will return to later.

Opposite Charges Attract

After using *charge=reactive* and *bond polarity* to identify *where* a reaction would occur, most students then employed the resource *opposite charges attract* to rationalize *how* the charged molecules would react. Eleven of the twelve students interviewed utilized this resource at least once throughout their interview. Matt relied extensively on *opposite charges attract* as he reasoned through the given problems and reflected aloud on the value of this resource as a general tool in organic chemistry. In the context of the simple bimolecular substitution problem, Matt used *opposite charges attract* to justify why a bond would form between an oxygen atom and a carbon atom, explaining that “this negatively charged oxygen with the ethyl [is] attracted to the partial positive carbon.” Later, when asked to predict the outcome of a complex transesterification reaction, he invoked *opposite charges attract* to identify several possible reaction pathways. He suggested that the alpha carbon could be deprotonated, “giving it a full negative charge,” which would then be “attracted to something positive.” Alternatively, he reasoned that “the lone pair of electrons [on the oxygen] could be attracted to the carbonyl carbons because there’s a partial positive on the carbonyl carbon.” Furthermore, Matt perceived activating *opposite charges attract* as broadly useful in organic chemistry. When asked how he determined what could react with the negatively charged oxygen (which he had previously identified as reactive), he said, “I always think it’s attracted to something. The lone pair of electrons are going to be attracted to something like an acidic hydrogen or something that has a partial positive charge on it.” His use of the word “always” suggests that he approached many, if not all, problems with this resource in mind. Matt clearly believed that knowing *opposite charges attract* was central to understanding chemistry in general and, like most of the students interviewed, he leveraged this resource to make predictions about molecular behavior in a complex and unfamiliar reaction context.

More Charge = More Reactive

The *more charge=more reactive* resource summarizes the relationship between concentration of charge and reactivity. This resource is typically expressed with comparative language and was often stated in opposite terms, such as “less concentrated,” “more stable,” or “more spread out.” Note that *more charge=more reactive* may be a specific case of Ohm’s p-prim *more means more* described by diSessa (diSessa et al., 2004). Ten of the twelve students interviewed used *more charge=more reactive*, most frequently in the context of identifying the leaving group, a feature that is common to many reactions in organic chemistry. For example, when Adam was considering the outcome of a simple transesterification reaction, he made the following statement about the phosphate group, which he had not seen before: “So it seems like a good leaving group to me because it can stabilize that negative charge” through the “smearing of negative charge.” He explained his reasoning further by discussing the converse: “if you have a lot of negative charge in one place, it tends to be unstable.” The idea that *more charge=more reactive* allowed Adam to predict that the phosphate group would be a good leaving group because the delocalization of the added electrons across the molecule would make it more stable. The ease with which Adam predicted the reactivity of the phosphate group demonstrates the broad utility a resource like *more charge=more reactive* may have for making sense of new chemistry phenomena.

Energy

All but one of the students invoked *energy* at some point in their explanations. Usually, it co-occurred with the previously discussed resources *charge=reactive* and *more charge=more reactive* since *energy* is implicit in these, but it was occasionally observed independently from charge in students’ reasoning. Three students considered an intramolecular reaction that would form a four-membered ring, which led them to comment on ring stability. Sadie explained, “Usually five to six membered rings are preferred over lesser-membered rings because... lesser ones have a lot of strain in them, so they’re not as stable.” Similarly, David said, “I think this one will be more likely to attack this because this is going to form the more stable product because you have the square cycle there and that's a lot less stable than if we were just to attack here." Neither of these students were asked to explain why four-membered rings are less stable than five- or six-membered rings, so we cannot determine whether they were invoking this as a rule or whether they could explain the source of ring strain. Another example of *energy* activation occurred when Matt considered which product would be formed in the trans-esterification reaction. He asked himself, "What product can you get that'll be the most stable, that'll be at the lowest energy, the least reactive?" The general way in which Matt described his strategy of looking for the lowest energy product that can form implies that he viewed consideration of *energy* as a broadly useful tool for predicting the outcome of unknown reactions. Although other students did not articulate this strategy as explicitly as Matt did, most of them made at least one prediction based on the relative stabilities (*i.e*. energies) of the molecules involved.

Case Study: Melissa's Activation of Conceptual and Epistemological Resources

To illustrate how conceptual and epistemological resources were activated as students made sense of multiple reactions, we will unpack the dialogue of one student: Melissa. Melissa was chosen for the following reasons: 1) She was very explicitly prompted to unpack her decisions and statements as she predicted the outcome of a bimolecular substitution reaction, giving us considerable evidence of the conceptual resources she activated in the context of a familiar reaction. 2) Melissa vocalized much of her reasoning while working through the complex transesterification reaction. 3) Melissa's use of conceptual and epistemological resources was typical of several students interviewed for this study.

Melissa clearly viewed the bimolecular substitution prompt as an opportunity to recall a previously learned, canonical explanation. She predicted and explained the reaction outcome without hesitation or frequent restarts, indicating activation of *knowledge as fabricated stuff* and *causal storytelling*. When elaborating on her prediction as to the outcome of the bimolecular substation reaction, Melissa occasionally shifted from recalling a previously learned causal mechanism to remembering a fact from a source of authority. This frame shift represented activation of epistemological resources including *knowledge as propagated stuff* and *accumulating.* For example, she stated that good leaving groups were "typically things at the end of the periodic table" that were "large and electronegative"; two facts about leaving groups she had been previously taught.

When delivering a causal explanation for the outcome of a bimolecular substation reaction, Melissa activated and connected nearly all of the conceptual resources we identified. As this prompt presented a familiar context, invocation of conceptual resources may indicate prior success rather than a view that resource clusters were useful in making sense of unfamiliar contexts. She started by reasoning, "So this one, the oxygen has a negative charge on it. So I would then attack this carbon, and move these electrons onto the Br because this carbon has a partial positive on it, because the bromine is electronegative." As part of this segment of dialogue, Melissa identified both the negative charge and the partial positive charge (*charge=reactive)* and utilized *opposite charges attract* to predict how they would react. She further elaborated on this idea, saying, "So the electrons are negative obviously. So then those are going to be more likely to attack something that's partially positive compared to attacking something that's already electron-rich." She hinted at *bond polarity* when she stated, "So this [carbon] has the partial positive on it because it's next to the electronegative bromine," but a deeper explanation was not prompted for. Finally, when asked to justify why bromide was a good leaving group, Melissa invoked *energy*: "[Good leaving groups] can handle the extra negative charge from the new electrons coming in." As a reminder, we are not concerned with the precise terminology used, so we interpret the phrase "handle the extra negative charge" as a stability argument. She further characterized good leaving groups like bromine as "large and electronegative. So they can balance the negative charge more when they're added on." The inclusion of a size criterion is indicative of *more charge=more reactive*; a larger atom has a larger area to spread negative charge over, which makes it more stable, or less reactive.

After predicting and justifying the outcome of a bimolecular substitution reaction, Melissa was given the complex transesterification reaction to consider (Fig. 2.1C). She relied on the same conceptual resources she used to explain the previous scenario but exhibited a qualitatively distinct sense of "what was going on" with respect to knowledge. Most notably, activation of *knowledge as fabricated stuff* and *forming*, which were not observed in the context of a bimolecular substitution reaction, were common. Activation of this cluster of epistemological resources is associated with students framing an interview scenario as an opportunity to construct (rather than recall) an explanation in-the-moment from prior knowledge. This was consistent with our expectation that the first reaction would be very familiar to students whereas the transesterification reactions would be unfamiliar and require students to figure out reasonable paths forward.

Melissa began trying to make sense of the complex transesterification reaction by looking for negative and positive charges (*charge=reactive*) that could react with each other (*opposite charges attract*). However, it soon became apparent to her that this strategy was limited in the context of a complex system. She said,

> So, with the last one it was easier because it's a very, very simple system where, "Okay, here's what's electronegative, it's going to attack something that's partially positive." Whereas this one, there's so many options for what could be partially positive or partially negative. Because there's a negatively charged oxygen on both of them. So you don't which one's going to attack this one.

She ended up relying in part on an intuitive sense of what was reasonable grounded in prior examples seen recently in-class. "I'm thinking this one will attack somewhere over here because this has carbonyl groups, which we've been reacting a lot with, where they'll like go and react at the carbonyl center and kick up the electrons to the oxygen.” After determining that the negatively charged oxygen on one reactant would likely donate electron density to a carbonyl carbon, she drew on her knowledge (*fabricated*) of *bond polarity* and *more charge=more reactive* to determine which carbonyl carbon had a greater partial positive charge and therefore which would be more reactive (*forming*). “Then out of all the carbonyls, there's only two of them. I'm thinking it's going to attack this one because it has the adjacent ... nitrogen group, which would make it more partially positive, because that would also be withdrawing electrons.” After drawing out this proposed step, she debated whether the oxygen would be protonated or the carbonyl would reform. “I know we've used like ester-y looking things as leaving groups before, so I could also... maybe picture that happening? But for me, I feel like the easiest thing to do would just be to like protonate this oxygen.” In this situation, Melissa identified two possible outcomes and made a decision based on a gut instinct about which would be “easiest” (*intuited* and *forming*).

After choosing to protonate the oxygen in a drawn intermediate, Melissa was prompted to consider an alternative pathway: reformation of the carbonyl functional group. In reflecting upon how a carbonyl could be reasonably reformed, Melissa considered which group attached to her drawn intermediate would be most stable (that is, lowest in energy) were it to depart. As with her prior explanation of bromide’s leaving group ability, she invoked an *energy* argument, noting that a good leaving group “can stabilize the more negative charge more easily.” The topic of leaving group ability seemed to return Melissa to familiar ground. She confidently concluded, “Yeah, so then the oxygen is more electronegative, then that's more likely to leave because that can like handle... The electrons will be more easily pulled to that compared to the nitrogen or the alkyl group” (*fabricated* and *causal storytelling*). However, in following this reasoning, she ended up reforming the starting material, which she described as “not ideal.” This illustrates an interesting epistemological shift from using her knowledge of chemistry to reason about the most likely outcome of a reaction to using her knowledge to arrive at a “right” answer or “ideal” outcome.

After reforming the starting material, Melissa recognized that her initial proposed step was inconsistent with her knowledge of electronegativity, and she proposed a different reaction instead. This once again led her to a consideration of leaving group ability in which she utilized her knowledge of *energy* and *more charge=more reactive*.

> [The phosphate group] should be a better leaving group, because if the negative charge is in all these different places and then you add more negative to it, it can kind of ... I feel like the new negative can kind of move through the same system. Whereas if you have something where the negative like has to go on the oxygen, I feel like that's going make the system less stable because there's no potential for the negative to be shared in other places. But if you have something with a lot of resonance forms possible, then that would be a good leaving group because the charge could be more delocalized ... I think? Yeah.”

The phosphate group was unfamiliar to Melissa, but she was able to use her prior knowledge of charge delocalization and stability to figure out how it was likely to react (*fabricated* and *forming*).

In these two contexts, one familiar and one unfamiliar, Melissa employed the same suite of resources in roughly the same sequence. She started by identifying charges that signal reactivity (*charge=reactive*), using *bond polarity* to determine where the partial charges were, and predicted how they would interact using *opposite charges attract*. In both situations, she used ideas related to *energy* and *more charge=more reactive* to predict and/or justify her choice of leaving group. Melissa’s example suggests that she viewed the resources she activated to explain the bimolecular substitution reaction as valuable for reasoning about unfamiliar reactions.

Students' Organic Chemistry Learning Environment Rewarded Use of Ideas Related to Electrostatics and/or Energy on Assessments

Students' sense of the conceptual resources useful in constructing predictions and explanations for phenomena presented during our interview was likely influenced by prior experiences they deemed similar. Given that Curriculum A was designed to support students' construction of causal mechanistic explanations for phenomena, we suspected Curriculum A assessments would present prompts similar to those which structured our interview. As assessments send strong implicit messages to students about what is of value in a course (Crooks, 1988; Entwistle, 1991; Momsen et al., 2013; Scouller, 1998; Scouller & Prosser, 1994; Snyder, 1973), it is reasonable to suppose that students would perceive patterns of resource activation rewarded on assessments as broadly useful for sensemaking. Here, we characterize the intellectual "heavy lifting" assessed in Curriculum A and relate the conceptual resources activated during our interview to the ideas apportioned credit on homework and exams.

In order to describe the intellectual work rewarded on Curriculum A assessments, we analyzed the multiple choice and short answers assessment items given on exams throughout both semesters of the course using the 3D-LAP (Laverty et al., 2016). Since we focused on students' ability to explain phenomena, not their ability to recall facts or algorithms (*e.g*. nomenclature), we limited our analysis to 3D questions– that is, questions which have the potential to engage students in using core ideas in practices characteristic of science to make sense of phenomena. In total, 33% of the points on examined assessments were allotted to 3D questions. This compares favorably to organic chemistry assessments given at "elite" institutions, which often allocate greater than 90% of assessment points to questions that cannot elicit evidence of student engagement in a scientific practice (Stowe & Cooper, 2017). The majority (57%) of Curriculum A assessment items that met the 3D-LAP criteria for having the potential to elicit evidence of 3D learning required students to use more than one core idea. This is unsurprising, given that most chemical phenomena can only be understood through the use of conceptual tools linked to multiple core ideas. Indeed, linkage between "big ideas" in the course is explicit in some of the codes we used to describe resource activation observed during interviews (*e.g.*, *more charge=more reactive* involves both electrostatics and energy). A quantitative Venn-diagram depicting the percentage of 3D assessment points allocated to items coded with one or more chemistry core idea is shown in Figure 2.2. This Venn Diagram was generated using the nVennR package in R (Pérez-Silva et al., 2018). Of the four core ideas Curriculum A was built around, "energy" was the most frequently emphasized by 3-Dimensional questions – 94% of the points Curriculum A allocated to 3D exam questions asked students to weave energy ideas into explanations, predictions and/or models. A requirement that students invoke "energy" almost always co-occurred with an expectation that students leverage some combination of the other course "big ideas" (*e.g.*, electrostatics and/or atomic/molecular structure and properties). This makes sense as one cannot explain energy changes at the particulate-level without saying something about forces, donor-acceptor interactions and/or molecular structure. "Electrostatics" was the second most commonly coded core idea among 3D Curriculum A assessments – 57% of points allocated to 3D items received this code. Interestingly, all items coded as requiring students to use ideas related to "atomic/molecular structure and properties" also required activation of electrostatic ideas. This is not surprising, as many of the characteristics of molecular structure germane to reactivity can be thought of in electrostatic terms (*e.g.*, polarized bonds, partial charges on particular functional groups).

An important takeaway from our assessment analysis is that Curriculum A allocated substantial points, in both semesters, to students applying a relatively small set of conceptual tools to explain why things happen. As assessments convey strong implicit messages to students as to the focus of a course, if we wish for construction of models and explanations grounded in “big ideas”, we need to allot points to these performances. As Curriculum A allots points for use of ideas clustered under the headings of “energy”, “electrostatics”, and “atomic/molecular structure” to craft causal explanations, we should not be too surprised that successful students find those ideas useful in understanding new contexts that have similar features to in-class assessments. All of our assessment prompts bore some similarity to 3D tasks the students have experience with in-class. Importantly, emphasis on figuring out how and why chemical phenomena occur was found on both high and low-stakes assessments. 61% of recitation activities, which were completed weekly by groups of Curriculum A students working together, included at least one question that had the potential to elicit evidence of 3-Dimensional learning. Grades on recitation activities were assigned on the basis of effort rather than correctness and problems were not assigned a point value. Accordingly, we cannot report the percentage of recitation points dedicated to assessing 3D learning.

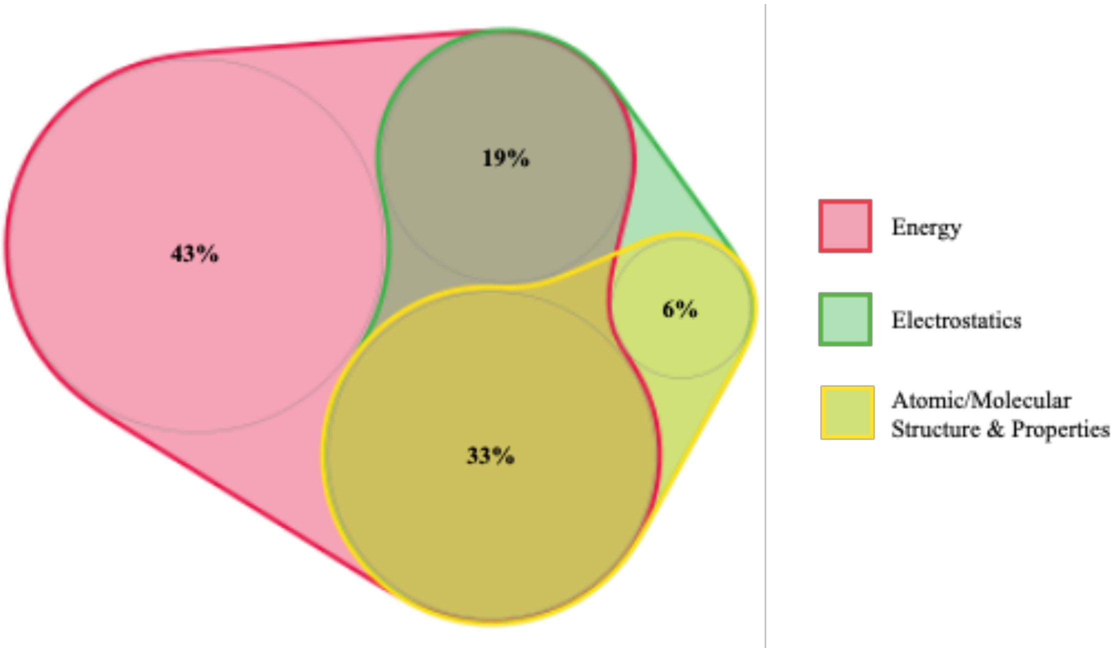


Figure 2.2. Venn diagram showing the percentage of 3D points associated with each core idea on Curriculum A assessments. Energy is shown in red, electrostatics is shown in green, and atomic/molecular structure & properties is shown in yellow.

Limitations

One limitation of our study is the focus we placed on eliciting evidence of the epistemological and conceptual resource activation of select, successful students. Given the high achieving character of our study subjects, it is possible that the patterns of coordinating resources *in situ* that we characterize here are not representative of the sensemaking strategies used by the bulk of students enrolled in organic chemistry learning environments. However, Crandell and colleagues have shown that a significant portion (>50%) of Curriculum A-enrolled undergraduates invoke electrostatics appropriately when describing the cause of a bimolecular substitution, so it would seem that conceptual tools related to Coulomb's law are somewhat commonly used in core idea-centered organic chemistry learning environments (Crandell et al., 2020).

Given the dynamic and context-sensitive nature of cognition, it is also likely that students activate conceptual resources in ways that were not surfaced by our interview. Thus, one should not read the themes we identified as the sum total of the sensemaking strategies students might employ to explain a phenomenon. It is quite likely that tweaking the prompt or verbal scaffolding would result in activation of different resources. Relatedly, we have no evidence that students would default to the same sensemaking strategies we observed when asked to think through the cause for a phenomenon on a high-stakes, time-limited exam or quiz. It is possible the constraints of such high-stakes assessment contexts foreground the activation of other resources associated with quickly producing an answer.

Conclusions

In this study, we characterized the conceptual tools that high-achieving students activated when predicting and explaining perplexing phenomena. We demarcated between moments when students were recalling an explanation and moments when they were constructing an explanation *in situ* by characterizing the epistemological resources related to knowledge source and epistemic activity which were activated throughout the interview. Recall of a memorized passage was signified by smooth, confident delivery of a causal story, while on-the-fly formation of an explanation was often indicated by slow, halting speech and many stops and starts. When explaining both familiar and unfamiliar phenomena, students very commonly invoked small-grain knowledge elements related to electrostatics (*e.g.*, "opposite charges attract") and energy (*e.g.*, "charge=reactive"). Discussion of electrostatic ideas, such as "charge", very commonly co-occurred with discussion of ideas related to energy, such as "stability" or its converse "reactivity". This co-occurrence may indicate that students see linkage of conceptual resources related to these two core ideas as a productive sensemaking strategy. Students also were capable of providing a nuanced analysis of competing pathways when prompted to do so.

It is worth mentioning that all students interviewed were capable of constructing more sophisticated explanations than their first utterance or drawing would suggest. For example, while initial explanations of the bimolecular substitution (Fig. 2.1A) tended to be fairly simple, many students were able to unpack the electrophilicity of the carbon attached to bromine via invocation of electronegativity and bond polarization when prompted. Articulation of competing reaction pathways almost always required explicit prompting from the interviewer, especially in the more complex trans-esterification phenomenon students were asked to consider. All of this indicates 1) students possessed a range of powerful conceptual tools and, when asked, were able to use them to predict and explain the outcome of a complex chemical system, and 2) explicit prompting (either written or verbalized) is often required to help students construct the most elaborated model and/or explanation they are capable of constructing. An underdeveloped answer to a problem requesting a model or model-based explanation should thus not be read to imply that students necessarily lack an understanding of the phenomenon in question – it is possible that the way the question is structured is not cuing students into which resources would be best employed. This perspective is consistent with work by Cooper and colleagues which found that the way students were asked to explain an acid-base reaction powerfully affected the sophistication of the responses constructed (Cooper et al., 2016).

The learning environment our students were enrolled in placed substantial emphasis on students explaining how and why processes occurred. Course exams dedicated 33% of assessment points to items that had the potential to engage students in predicting, explaining, and/or modeling phenomena. Additionally, 61% of recitation activities included at least one item with the potential to elicit evidence of 3D learning. Those exam prompts which required construction of explanations or models overwhelmingly emphasized energy and electrostatics, which were well represented among the conceptual tools interviewed students used for sensemaking. This finding highlights the importance of high and low stakes assessments in telegraphing what matters in any given course. Students who succeeded in Curriculum A were rewarded for their appropriate use of tools linked to large-grain ideas and so seemed to internalize those tools as broadly useful in engaging with prompts similar to exam tasks.

Implications

Implications from our study rest upon the presumption that a central goal of chemistry learning environments is to support students in predicting and explaining phenomena in terms of atomic/molecular behavior (*i.e.*, sensemaking). We argue that, without sensemaking as a central theme, chemistry coursework is often reduced to a collection of unimportant factoids and algorithms. In a sensemaking-focused organic chemistry learning environment, central disciplinary concepts should be regarded as those ideas that are useful for predicting and explaining a broad swathe of phenomena. Students should have ample opportunities, during class and on assessments, to figure out how and why events happen by leveraging fundamental ideas in the construction of predictive and explanatory models. This study indicates that, in such a phenomena-focused learning environment, successful students may internalize conceptual resources that cluster under “big ideas” as commonly helpful and thus make use of those tools when confronted with a new scenario to figure out. Notably, focus on explaining phenomena in terms of electrostatics, energy, and atomic/molecular structure and properties started early in Curriculum A and continued for both semesters of the course. This suggests that explaining and modeling phenomena can be integrated throughout a course and that chemistry learning environments need not focus on “skill building” exclusively prior to asking “why” questions.

There is growing evidence that students may be supported in weaving ideas together into particle-level explanations and models by making sense of increasingly complex systems (Cooper & Klymkowsky, 2013; Cooper et al., 2019; Stowe et al., 2019). This study suggests that coherent emphasis on figuring out why events happen on high and low stakes assessments plays an important role in students seeing some conceptual resources as broadly applicable tools useful for crafting explanations and models. If all aspects of a course (*e.g.*, homework, in-class interactions, exams) focus on engaging students in weaving "big ideas" together as they explain and model phenomena, it is likely that students will internalize the broad utility of these ideas and call upon them when attempting to figure out novel scenarios. Finally, we would like to emphasize that building from simple systems is likely quite important to helping students make atomic/molecular sense of observable events. There is no chance at-all that a student fresh from general chemistry would have the tools needed to figure out the enzyme-catalyzed transesterification process we gave in Figure 2.1C. It should be noted that we have no evidence the conceptual sequencing in Curriculum A is the "best" way to build system complexity – indeed, we would hypothesize that there is no "best sequence" but instead a variety of reasonable ways to engage students in increasingly sophisticated sensemaking.

References

Agarwal, V., & Nair, S. K. (2012). Aminoacyl tRNA synthetases as targets for antibiotic development. *MedChemComm*, *3*(8), 887–898. https://doi.org/10.1039/C2MD20032E

Anzovino, M. E., & Bretz, S. L. (2015). *Organic chemistry students' ideas about nucleophiles and electrophiles: The role of charges and mechanisms*. *16*(4), 797–810. https://doi.org/10.1039/C5RP00113G

Atkins, P. (1999). Chemistry: The great ideas. *Pure and Applied Chemistry*, *71*(6), 927–929. https://doi.org/10.1351/pac199971060927

Berland, L., & Crucet, K. (2016). Epistemological Trade-Offs: Accounting for Context When Evaluating Epistemological Sophistication of Student Engagement in Scientific Practices. *Science Education*, *100*(1), 5–29. https://doi.org/10.1002/sce.21196

Berland, L. K., & Hammer, D. (2012). Framing for scientific argumentation. *Journal of Research in Science Teaching*, *49*(1), 68–94. https://doi.org/10.1002/tea.20446

Brown, D. E. (2014). Students' Conceptions as Dynamically Emergent Structures. *Science & Education*, *23*(7), 1463–1483. https://doi.org/10.1007/s11191-013-9655-9

Caspari, I., Kranz, D., & Graulich, N. (2018). Resolving the complexity of organic chemistry students' reasoning through the lens of a mechanistic framework. *Chemistry Education Research and Practice*, *19*(4), 1117–1141. https://doi.org/10.1039/C8RP00131F

Chari, D. N., Nguyen, H. D., Zollman, D. A., & Sayre, E. C. (2017). Student and instructor framing in upper-division physics. *ArXiv.Org*. https://arxiv.org/abs/1704.05103

Collins, A., & Ferguson, W. (1993). Epistemic forms and epistemic games: Structures and strategies to guide inquiry. *Educational Psychologist*, *28*(1), 25–42. https://doi.org/10.1207/s15326985ep2801_3

Cooper, M. (2010). The Case for Reform of the Undergraduate General Chemistry Curriculum. *Journal of Chemical Education*, *87*(3), 231–232. https://doi.org/10.1021/ed800096m

Cooper, M., & Klymkowsky, M. (2013). Chemistry, Life, the Universe, and Everything: A New Approach to General Chemistry, and a Model for Curriculum Reform. *Journal of Chemical Education*, *90*(9), 1116–1122. https://doi.org/10.1021/ed300456y

Cooper, M. M., Caballero, M. D., Ebert-May, D., Fata-Hartley, C. L., Jardeleza, S. E., Krajcik, J. S., Laverty, J. T., Matz, R. L., Posey, L. A., & Underwood, S. M. (2015). Challenge faculty to transform STEM learning. *Science*, *350*(6258), 281–282. https://doi.org/10.1126/science.aab0933

Cooper, M. M., Corley, L. M., & Underwood, S. M. (2013). An investigation of college chemistry students' understanding of structure–property relationships. *Journal of Research in Science Teaching*, *50*(6), 699–721. https://doi.org/10.1002/tea.21093

Cooper, M. M., Kouyoumdjian, H., & Underwood, S. M. (2016). Investigating Students' Reasoning about Acid–Base Reactions. *Journal of Chemical Education*, *93*(10), 1703–1712. https://doi.org/10.1021/acs.jchemed.6b00417

Cooper, M. M., Posey, L. A., & Underwood, S. M. (2017). Core Ideas and Topics: Building Up or Drilling Down? *Journal of Chemical Education*, *94*(5), 541–548. https://doi.org/10.1021/acs.jchemed.6b00900

Cooper, M. M., & Stowe, R. L. (2018). Chemistry Education Research—From Personal Empiricism to Evidence, Theory, and Informed Practice. *Chemical Reviews*, *118*(12), 6053–6087. https://doi.org/10.1021/acs.chemrev.8b00020

Cooper, M. M., Stowe, R. L., Crandell, O. M., & Klymkowsky, M. W. (2019). Organic Chemistry, Life, the Universe and Everything (OCLUE): A Transformed Organic Chemistry Curriculum. *J. Chem. Educ.*, *96*(9), 1858–1872. https://doi.org/10.1021/acs.jchemed.9b00401

Crandell, O. M., Kouyoumdjian, H., Underwood, S. M., & Cooper, M. M. (2019). Reasoning about Reactions in Organic Chemistry: Starting It in General Chemistry. *Journal of Chemical Education*, *96*, 213–226. https://doi.org/10.1021/acs.jchemed.8b00784

Crandell, O. M., Lockhart, M. A., & Cooper, M. M. (2020). Arrows on the Page Are Not a Good Gauge: Evidence for the Importance of Causal Mechanistic Explanations about Nucleophilic

Substitution in Organic Chemistry. *Journal of Chemical Education*, *97*(2), 313–327. https://doi.org/10.1021/acs.jchemed.9b00815

Crooks, T. J. (1988). The Impact of Classroom Evaluation Practices on Students. *Review of Educational Research*, *58*(4), 438–481. https://doi.org/10.2307/1170281

diSessa, A. A. (1988). Knowledge in pieces. In G. E. Forman & P. B. Pufall (Eds.), *Constructivism in the computer age* (pp. 49–70). Lawrence Erlbaum Associates.

diSessa, A. A. (1993). Toward an Epistemology of Physics. *Cognition and Instruction*, *10*(2/3), 105–225.

diSessa, A. A., Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science*, *28*(6), 843–900. https://doi.org/10.1207/s15516709cog2806_1

Entwistle, N. J. (1991). Approaches to learning and perceptions of the learning environment. *Higher Education*, *22*(3), 201–204. https://doi.org/10.1007/BF00132287

Galloway, K. R., Stoyanovich, C., & Flynn, A. B. (2017). Students' interpretations of mechanistic language in organic chemistry before learning reactions. *Chemistry Education Research and Practice*, *18*(2), 353–374. https://doi.org/10.1039/C6RP00231E

Gillespie, R. J. (1997). The Great Ideas of Chemistry. *Journal of Chemical Education*, *74*(7), 862–864. https://doi.org/10.1021/ed074p862

Glaser, B. G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. Sociology Press.

Graulich, N., & Bhattacharyya, G. (2017). Investigating students' similarity judgments in organic chemistry. *Chemistry Education Research and Practice*, *18*(4), 774–784. https://doi.org/10.1039/C7RP00055C

Grove, N. P., Cooper, M. M., & Rush, K. M. (2012). Decorating with Arrows: Toward the Development of Representational Competence in Organic Chemistry. *Journal of Chemical Education*, *89*(7), 844–849. https://doi.org/10.1021/ed2003934

Hammer, D. (2000). Student resources for learning introductory physics. *American Journal of Physics*, *68*(S1), S52–S59. https://doi.org/10.1119/1.19520

Hammer, D., & Berland, L. K. (2014). Confusing Claims for Data: A Critique of Common Practices for Presenting Qualitative Research on Learning. *Journal of the Learning Sciences*, *23*(1), 37–46. https://doi.org/10.1080/10508406.2013.802652

Hammer, D., & Elby, A. (2002). On the form of a personal epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing* (pp. 169–190). Lawrence Erlbaum Associates.

Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. In J. P. Mestre (Ed.), *Transfer of Learning from a Modern Multidisciplinary Perspective* (pp. 89–119). IAP. https://dhammer.phy.tufts.edu/home/publications_files/rft.pdf

Herrington, D. G., & Daubenmire, P. L. (2014). Using Interviews in CER Projects: Options, Considerations, and Limitations. In *Tools of Chemistry Education Research* (Vol. 1166, pp. 31–59). American Chemical Society. https://doi.org/10.1021/bk-2014-1166.ch003

Holme, T., & Murphy, K. (2012). The ACS Exams Institute Undergraduate Chemistry Anchoring Concepts Content Map I: General Chemistry. *Journal of Chemical Education*, *89*(6), 721–723. https://doi.org/10.1021/ed300050q

Judd, O. (2018). *Characterizing Student Thinking About Solutions and the Solvation Process: The Search for Mechanistic Understanding*. Michigan State University.

Laverty, J. T., Underwood, S. M., Matz, R. L., Posey, L. A., Carmel, J. H., Caballero, M. D., Fata-Hartley, C. L., Ebert-May, D., Jardeleza, S. E., & Cooper, M. M. (2016). Characterizing College Science Assessments: The Three-Dimensional Learning Assessment Protocol. *PLOS ONE*, *11*(9), e0162333. https://doi.org/10.1371/journal.pone.0162333

Linenberger, K. J., & Bretz, S. L. (2012). A novel technology to investigate students' understandings of enzyme representations. *Journal of College Science Teaching*, *42*(1), 45–49.

Lloyd, B. W., & Spencer, J. N. (1994). The Forum: New Directions for General Chemistry: Recommendations of the Task Force on the General Chemistry Curriculum. *Journal of Chemical Education*, *71*(3), 206. https://doi.org/10.1021/ed071p206

MacLachlan, G. L., & Reid, I. (1994). *Framing and Interpretation*. Melbourne University Press.

Minter, C. (2019). *Characterization of Students' Reasoning about Atomic Emission Spectra—A Design-Based Research Study to Improve Students' Understanding of Light-Matter Interactions*. Michigan State University.

Momsen, J., Offerdahl, E., Kryjevskaia, M., Montplaisir, L., Anderson, E., & Grosz, N. (2013). Using Assessments to Investigate and Compare the Nature of Learning in Undergraduate Science Courses. *CBE-Life Sciences Education*, *12*(2), 239–249. https://doi.org/10.1187/cbe.12-08-0130

Murphy, K., Holme, T., Zenisky, A., Caruthers, H., & Knaus, K. (2012). Building the ACS Exams Anchoring Concept Content Map for Undergraduate Chemistry. *Journal of Chemical Education*, *89*(6), 715–720. https://doi.org/10.1021/ed300049w

Nameroff, T. J., & Busch, D. H. (2004). Exploring the Molecular Vision: Report from a SOCED Invitational Conference. *Journal of Chemical Education*, *81*(2), 177–179. https://doi.org/10.1021/ed081p177

National Research Council. (2000). *How People Learn: Brain, Mind, Experience, and School: Expanded Edition* (2 edition). National Academies Press.

Noyes, K., & Cooper, M. M. (2018). Understanding the Causes of London Dispersion Forces: A Longitudinal Study. *Journal of Chemical Education*, *in preparation*.

Odden, T. O. B., & Russ, R. S. (2019). Vexing questions that sustain sensemaking. *International Journal of Science Education*, *41*(8), 1052–1070. https://doi.org/10.1080/09500693.2019.1589655

Pérez-Silva, J. G., Araujo-Voces, M., & Quesada, V. (2018). nVenn: Generalized, quasi-proportional Venn and Euler diagrams. *Bioinformatics*, *34*(13), 2322–2324. https://doi.org/10.1093/bioinformatics/bty109

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*(2), 211–227. https://doi.org/10.1002/sce.3730660207

QSR International Pty Ltd. (2018). *NVivo 12 for Mac*. https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

Redish, E. F. (2004). A Theoretical Framework for Physics Education Research: Modeling Student Thinking. *ArXiv Preprint Physics/0411149*. https://arxiv.org/abs/physics/0411149

Reiser, B. J., & Tabak, I. (2014). Scaffolding. In R. K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (2nd ed., pp. 44–62). Cambridge University Press.

Russ, R. S., Lee, V. R., & Sherin, B. L. (2012). Framing in cognitive clinical interviews about intuitive science knowledge: Dynamic student understandings of the discourse interaction. *Science Education*, *96*(4), 573–599. https://doi.org/10.1002/sce.21014

Russ, R. S., & Luna, M. J. (2013). Inferring teacher epistemological framing from local patterns in teacher noticing. *Journal of Research in Science Teaching*, *50*(3), 284–314. https://doi.org/10.1002/tea.21063

Sayre, E. C., & Wittmann, M. C. (2008). Plasticity of intermediate mechanics students’ coordinate system choice. *Physical Review Special Topics - Physics Education Research*, *4*(2), 020105. https://doi.org/10.1103/PhysRevSTPER.4.020105

Scherr, R. E., & Hammer, D. (2009). Student Behavior and Epistemological Framing: Examples from Collaborative Active-Learning Activities in Physics. *Cognition and Instruction*, *27*(2), 147–174. https://doi.org/10.1080/07370000902797379

Schwarz, C. V., Passmore, C., & Reiser, B. J. (2016). *Helping Students Make Sense of the World Using Next Generation Science and Engineering Practices—PB351X*. National Science Teachers Association - NSTA Press.

Scouller, K. (1998). The influence of assessment method on students’ learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, *35*(4), 453–472. https://doi.org/10.1023/A:1003196224280

Scouller, K. M., & Prosser, M. (1994). Students’ experiences in studying for multiple choice question examinations. *Studies in Higher Education*, *19*(3), 267–279. https://doi.org/10.1080/03075079412331381870

Shar, K., Russ, R. S., & Laverty, J. T. (2020). Student epistemological framing on paper-based assessments. *Physical Review Physics Education Research*, *16*(2), 020101. https://doi.org/10.1103/PhysRevPhysEducRes.16.020101

Sherin, B. L. (2001). How Students Understand Physics Equations. *Cognition and Instruction*, *19*(4), 479–541.

Snyder, B. (1973). *The Hidden Curriculum*. The MIT Press.

Suidan, L., Badenhoop, J. K., Glendening, E. D., & Weinhold, F. (1995). Common Textbook and Teaching Misrepresentations of Lewis Structures. *Journal of Chemical Education*, *72*(7), 583. https://doi.org/10.1021/ed072p583

Stowe, R. L., & Cooper, M. M. (2017). Practice What We Preach: Assessing "Critical Thinking" in Organic Chemistry. *The Journal of Chemical Education*, *94*(12), 1852–1859. https://doi.org/10.1021/acs.jchemed.7b00335

Stowe, R. L., Herrington, D. G., McKay, R. L., & Cooper, M. M. (2019). The Impact of Core-Idea Centered Instruction on High School Students' Understanding of Structure–Property Relationships. *Journal of Chemical Education*, *96*(7), 1327–1340. https://doi.org/10.1021/acs.jchemed.9b00111

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques* (2nd edition). SAGE Publications, Inc.

Taber, K. S. (2000). Case studies and generalizability: Grounded theory and research in science education. *International Journal of Science Education*, *22*(5), 469–487. https://doi.org/10.1080/095006900289732

Taber, K. S., & García-Franco, A. (2010). Learning Processes in Chemistry: Drawing Upon Cognitive Resources to Learn About the Particulate Structure of Matter. *Journal of the Learning Sciences*, *19*(1), 99–142. https://doi.org/10.1080/10508400903452868

Tannen, D. (1979). What's in a frame? Surface evidence for underlying expectations. In R. Freedle (Ed.), *New Directions in Discourse Procesing* (pp. 137–181). Ablex. https://static1.squarespace.com/static/5523ffe4e4b012b2c4ebd8fc/t/56ddb1df20c647a2ce5d005d/1457369568149/what%27s+in+a+frame.pdf

Tannen, D. (1993). *Framing in Discourse*. Oxford University Press.

Tannen, D., & Wallat, C. (1987). Interactive frames and knowledge schemas in interaction: Examples from a medical examination/interview. *Social Psychology Quarterly*, *50*(2), 205–216. https://doi.org/10.2307/2786752

The College Board. (2014). *AP Chemistry: Course and Exam Description*. The College Board. http://media.collegeboard.com/digitalServices/pdf/ap/ap-chemistry-course-and-exam-description.pdf

The National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. National Academies Press. http://nap.edu/catalog/13165

Wood, D., Bruner, J. S., & Ross, G. (1976). The Role of Tutoring in Problem Solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Supporting Information

The contents of this supporting information include elaborated codebooks for characterizing conceptual and epistemological resources and a description of the trans-esterification reaction.

Elaborated Conceptual Resources Codebook

The codebook refined by both authors and ultimately used to characterize the conceptual resources students activated during interviews is described here.

General Coding Guidelines

- Stability and reactivity are inversely related, so codes that are expressed in terms of stability can also be understood in terms of reactivity. For example, *more charge = more reactive* was used whenever a student discussed something with more charge as being less stable.
- Attract and repel are also inversely related, so if a student said that like charges repel, this was coded as *opposites attract.*

Opposites Attract

The code *opposites attract* was applied when students described a negative charge being attracted to a positive charge (or vice versa) or stated that an atom with a negative charge reacts with an atom with a positive charge. An example is provided in the following quote in which the student stated that the oxygen with a partial negative charge would react with the partially positive carbon.

> I know that this carbon is partially positive because of the bromine, so then one of the lone pairs from oxygen, because oxygen has a partial negative charge, is going to want to attack this carbon.

In the second example (see below), the student stated generally that the negative charge would be attracted to positive charges and then specifically identified where the positive and negative charges were on the molecules given.

> So the negative charge would be attracted to something that's positive. And there's two carbonyl carbons, and I know that those have a partial positive charge. So the negative charge from the O, I'm guessing, would be attracted to the charges on the carbonyl carbon.

Charge = reactive

The code *charge = reactive* was applied wherever students used charge to identify a reactive part of a molecule. In the following example, the student used the presence of the negative charge to predict which part of the molecule would react: “So I see that there is an oxygen in both molecules with a negative charge on it, which I feel like would react with something.” Students also received this code wherever they described charges as unstable or a lack of charge as less reactive or more stable, as shown in the following quote: “I feel like that's more stable. I don't see any full partial negatives or partial positive.”

More charge = more reactive

More charge was defined as either a greater magnitude of charge or a greater concentration of charge. Thus, the code *more charge = more reactive* was applied to segments in which students connected a greater magnitude or concentration of charge with increased reactivity. The following quote provides a good example: “When they're super concentrated, it would be more reactive just because there's a concentration of charge on one thing.” The impact of greater charge could also be expressed in terms of decreased stability rather than increased reactivity, as shown the in the following quote: “If you have a lot of negative charge in one place, it tends to be unstable.” For an example that invokes magnitude of charge, consider the following quote: “Yeah. But then the phosphate would have two O minus if it was kicked off... So it would increase the potential energy a lot in that molecule, which is really unstable, and un-stability is unfavorable.” Here the student is arguing that because the phosphate has a greater magnitude of charge (two negative charges versus one), it would be less stable. Finally, students could also express the concept *more charge*

= *more reactive* in terms of its converse—a smaller magnitude of charge or a more delocalized charge results in decreased reactivity or greater stability—as shown in the following quote: “If you could delocalize the [negative] electrons between different atoms in the molecule, then it makes it more stable.”

Bond polarity

The code bond polarity was used to describe segments in which students rationalized partial charges using their understanding of bond polarity. The following quote provides an example of a segment that was given this code.

> Because this bromine has a higher Z_{eff} or effective nuclear charge, it's more electronegative, so it's pulling electrons away from that carbon, so with less negative electron charge, then this becomes more positive.

Here, the student explained that the electronegativity difference between bromine and carbon results in more electron density on the bromine, making the carbon more positive. Similarly, in the next example, the student argued that a carbon bonded to more electronegative atoms, which are “pulling electrons away,” is more positive.

> In my mind the nitrogen was more electronegative, which is actually wrong. So I was basing it off of that fact and that then the carbon would be more partially positive because there'd be more things pulling electrons away from it.

Energy

The energy code was applied to any segment in which a student described the stability or reactivity of a molecule or charge. They did not have to use the term “energy” specifically for their words to receive this code (and most did not). An example invoking stability is shown here: “Square structures that are a little more unstable than, I mean obviously say like a cyclohexane, but it all matches up.” Here the student asserted that four-membered rings (*i.e.* “square structures”) are more unstable, or higher in energy, than six-membered rings (*i.e.* cyclohexanes). In the next example, the student realized that the negatively charged CH_2Br molecule would be quite reactive, or high in energy. They said, “this would make this really reactive cause it would be CH_2Br minus.”

Elaborated Epistemological Resources Codebook

Fabricated

Stated knowledge is figured out by student using their prior knowledge. In the following example, the student predicted a reaction based on their knowledge of where partial negative and partial positive charges were in the molecule.

> I think that this is going to go and attack something. I think, let me see. I think it's going to attack the carbonyl ‘cause this is partial negative and this is partial positive. I don't know where else it would go.

In the next example, the student drew on their knowledge of charge delocalization to determine that the molecule would be a good leaving group: “But if you have something with a lot of resonance forms possible, then that would be a good leaving group because the charge could be more delocalized ... I think?”

Propagated

Stated knowledge comes from outside source, such as a professor or textbook. Note that the student does not have to explicitly identify the outside source. For example, in the following quote, a student recalls a fact about oxidation states that they learned in class: “We learned they have the same oxidation states. I think that was what we talked about in class.” In the second example that follows, the student makes a factual statement about electronegativity that must have been learned from an outside source at some point: “But I just know it's electronegative. Its affinity for electrons is higher due to the trends from periodic table.”

Directly Perceived

Stated knowledge is readily apparent or obvious because it is directly given on the page. For example, a student identifies a negative charge that they see on a molecule, as demonstrated by the following quote: “I guess there’s a negative over there too.” Similarly, in the next example the student notices the carbonyl function group present in the molecule: *“*So, we have another carbonyl on the larger molecule that can be attacked.”

Intuited

Stated knowledge is determined from a gut-feeling or sense that it’s true. For example, in the following segment the student hesitates to kick off a group of atoms that were added earlier based on the feeling that undoing a previous step would be “silly.” The students says, “You'd have to kick something else off, but I feel like it wouldn't be this because we just put that on there, so that would just kind of be silly.” Intuited knowledge is often indicated by the phrase “I feel.” For example, consider the following quote: “I feel like this is the best way to be able to stabilize the oxygens and make the whole entire thing overall more stable than this.” The student believes that what they have suggested results in the greatest stability, but they do not articulate why they believe so.

Forming

Students is using their prior knowledge to reason about how to solve an unfamiliar problem. In the following example: the student is thinking through how either of the oxygen atoms could react and what product would result from the reaction: “I feel like either one of these oxygens, because they have a bunch of electrons on them too, could attack this carbonyl and maybe make it like a ring and close it up.” In the next example, the student is attempting to figure out why a molecule may be stable: “There's a whole lot of oxygens, and um, maybe it might be able to stabilize itself because it can have a lot of bonds.”

Causal Storytelling

Students explains why something happens by connecting a cause and an effect. In the following example, the student is describing how a difference electronegativity affects electron distribution in a bond: “It's more electronegative, so it's pulling electrons away from that carbon, so with less negative electron charge, then this becomes more positive.” In the second example, the student is explaining that having bromine on the molecule makes it electrophilic: “This looks like an electrophile because it is attached to bromine, which is a good leaving group.”

Accumulating

Student notices a feature of the prompt or recalls factual information. An example of factual recall is shown here (interviewer dialogue shown in brackets): [So do we have a name for these sorts of things?] “Oh, a tetrahedral intermediate. [And what do those often do?] They collapse back down.” The student recalls what type of intermediate they have drawn and what those intermediates do. The next example demonstrates a student accumulating information from the prompt: “I see there is a carbonyl there.” The student identifies a carbonyl group in the molecule by looking at the worksheet.

Fabricated + Causal Storytelling

Students knows the answer and is confidently and smoothly explaining it to the interviewer. In the following example, the student uses their knowledge of what makes a molecule stable to explain why nitrogen would react with a hydrogen more readily than oxygen:

> Because the oxygen will be able to hold onto those electrons easier than the nitrogen would. Because the nitrogen would just want to pick up a hydrogen or something because it's not as stable with extra electrons.

In the second example that follows, the student uses their knowledge of nucleophiles to explain that deprotonated oxygen makes it a good nucleophile:

> Okay, so I would probably start from oxygen and draw my arrow to this carbon over here, and that's because oxygen is deprotonated in this situation and it'd be probably a pretty good nucleophile to attack here and Br is a good leaving group.

Fabricated + Forming

Student does not know the answer and is trying to figure it out based on what they do know. In the first example shown below, the student draws on knowledge they have regarding acid-base reactions as they attempt to figure out how the starting materials will interact.

> I guess one of these hydrogens could get pulled off? If there was a really strong base in the solution? Like, er, not. One of the ones attached to the alpha carbon, which, would be, this is an alpha carbon, well, I guess it's the only alpha carbon left here. Well, maybe not actually.

The next example shows the student utilizes their knowledge of leaving groups to predict what would happen, although they are unsure as indicated by “potentially?”.

> And since this is a pretty big group here with a lot of area to stabilize a charge, I think it would be a good leaving group in this situation because I know the tetrahedral intermediate isn't the most stable so it would most likely collapse back. So I think that I would then to kick off that leaving group, potentially move this electron pair down there and then have the electrons from this bond leave and go onto the oxygen there... potentially?

Intuited + Forming

Student is putting together a solution based on an intuitive sense or pattern recognition and does not articulate any prior knowledge or reasoning as they do so but rather a “feeling” or “sense.” In the following example, the student is trying to determine where the reaction will occur and is basing it off a feeling of what the products should look like: “I feel like this is like one separate phosphate molecule and this is ... this carbonyl is like bonded to two oxygens, and I feel like it should be a place for ... a reaction.” In the second example, the student predicts an intramolecular reaction but does not explain why they feel this is reasonable: “But in my opinion, I think this would be more likely to attack this because I feel like the molecule is going to favor an intramolecular nucleophilic attack.”

Propagated + Accumulating

Student is making factual statements, often in response to direct questions from the interviewer. Consider the following dialogue:

> Interviewer: So what does it mean to be a good leaving group?
> Student: It can handle the extra negative charge from the new electrons coming in. So typically the end of the periodic table.
> Interviewer: What sort of thing characterizes those?
> Student: Just being large and electronegative.

The student recalls general features of good leaving groups in response to specific, direct questions. In the following example, the student recalls a trend in leaving groups based on the periodic table: “When we look at leaving groups, they get better as you go down the periodic table as the molecule get bigger they become better leaving groups.”

Directly Perceived + Accumulating

Student is noticing features present on the molecules. These features must be explicitly written/drawn, such as formal negative charges, not implicit, such as partial positive charges. In the following example, the student perceives the negative charge in the prompt and “gathers” that observation by vocalizing it: “I guess there's a negative over there too.” Similarly, the student in the next example sees an oxygen in a phosphate group with a negative charge: “Okay, so this is an oxygen that's part of a phosphate group, which has a negative charge on it.”

Description of Trans-Esterification of a Phosphate Ester

Formation of the major product that arises from treating the phosphate ester shown in Figure 2.SA1 with an alkoxide comes about via donation of electrons from the negatively charged alkoxide oxygen to the partial positively charged carbonyl carbon. As a single bond forms between the alkoxide oxygen and carbonyl carbon, the carbonyl double-bond breaks, giving rise to the tetrahedral intermediate pictured above. Formation of this tetrahedral intermediate is reversible; the starting material may re-form. However, phosphate is a much lower energy species than an alkoxide anion due to substantial charge delocalization. Collapse of the tetrahedral intermediate and simultaneous departure of phosphate thus results in a lower energy system than existed at the start of the process. For this reason, phosphate departure is functionally irreversible, and an ester is the major product observed from this system. Charge delocalization on a phosphate anion is often depicted via three resonance structures, shown in Figure 2.S1B. This is a useful model for predicting species stability in this case, but the structure of phosphate is best represented as the “major resonance contributor” listed (Suidan et al., 1995).

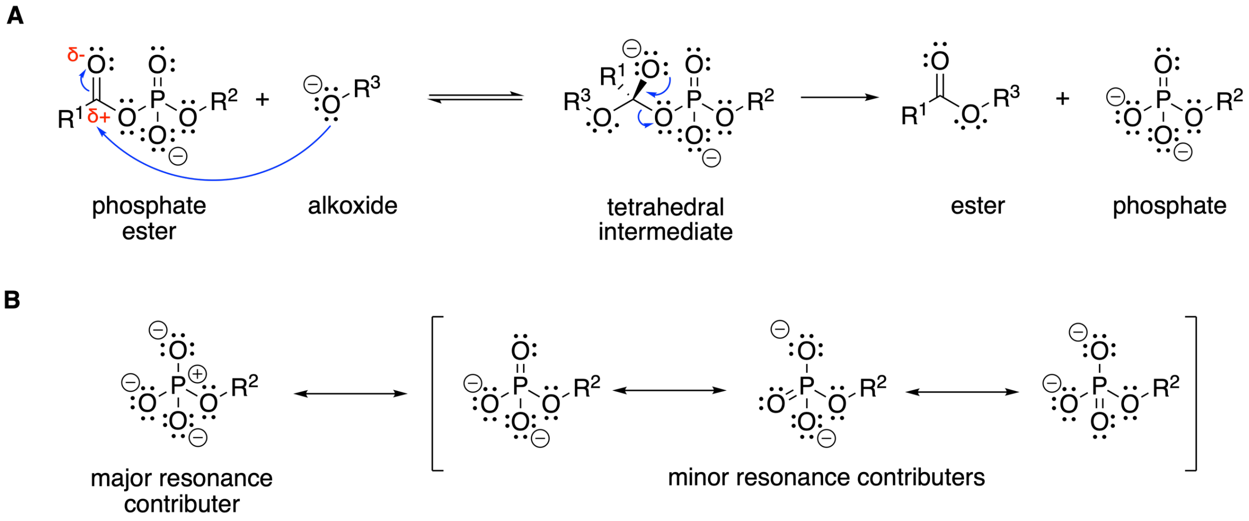


Figure 2.S1. A) Mechanism of a simple trans-esterification reaction under basic conditions. The electron-pushing arrows are shown in blue, and the partial charges on the phosphate ester are shown in red. “R-groups” (*i.e.*, R^1, R^2) represent carbon chains with variable lengths, branches and functionality. B) The major resonance contributor to the structure of the phosphate anion (Suidan et al., 1995) and minor resonance contributors depicting the delocalization of charge among three oxygen atoms.

Chapter 3

Modeling Students' Epistemic Cognition in Undergraduate Chemistry Courses: A Review

This work was conducted in collaboration with Ryan L. Stowe.

Abstract

Thinking about knowledge and knowing (*i.e.*, epistemic cognition) is an important part of student learning and has implications for how they apply their knowledge in future courses, careers, and other aspects of their lives. Three classes of models have emerged from research on epistemic cognition: developmental models, dimensional models, and resources models. These models can be distinguished by how value is assigned to particular epistemic ideas (hierarchy), how consistent epistemic ideas are across time and/or context (stability), and the degree to which people are consciously aware of their own epistemic ideas (explicitness). To determine the extent to which these models inform research on epistemic cognition in chemistry education specifically, we reviewed 54 articles on undergraduate chemistry students' epistemologies. First, we sought to describe the articles in terms of the courses and unit of study sampled, the methods and study designs implemented, and the means of data collection utilized. We found that most studies focused on the epistemic cognition of individual students enrolled in introductory chemistry courses. The majority were qualitative and employed exploratory or quasi-experimental designs, but a variety of data collection methods were represented. We then coded each article for how it treated epistemic cognition in terms of hierarchy, stability, and explicitness. The overwhelming majority of articles performed a hierarchical analysis of students' epistemic ideas. An equal number of articles treated epistemic cognition as stable versus unstable across time and/or context. Likewise, about half of the studies asked students directly about their epistemic cognition while approximately half of the studies inferred it from students' responses, course observations, or

written artifacts. These codes were then used to infer the models of epistemic cognition underlying these studies. Eighteen studies were mostly consistent with a developmental or dimensional model, ten were mostly aligned with a resources model, and twenty-six did not provide enough information to reasonably infer a model. We advocate for considering how models of epistemic cognition—and their assumptions about hierarchy, stability, and explicitness—influence the design of studies on students' epistemic cognition and the conclusions that can be reasonably drawn from them.

Introduction

Ideas about knowledge and knowing underlie many of the behaviors people exhibit and the decisions they make. Whenever people ask a question, they are pursuing some type of knowledge product (*e.g.*, factual information). When they engage in a particular behavior to answer that question, (*e.g.*, typing their question into a search engine), it is likely because they view it as a reliable process for obtaining the desired knowledge product. During the process, they are also relying on ideas regarding appropriate justifications for knowledge (*e.g.*, obtained from a trustworthy source). Importantly, given how often people need to reason with or about knowledge on a daily basis, these considerations tend to be made subconsciously (Hammer & Elby, 2002; Chinn et al., 2011).

Thinking about knowledge and knowing (*i.e.*, epistemic cognition) is also part of student learning in the classroom. Some ways of thinking are independent of subject matter, as exemplified by the common course goal of improving students' critical thinking skills (Tiruneh et al., 2014). Each discipline also has its own norms regarding what constitutes good knowledge, how it is obtained, what form it should take, etc. (Buehl & Alexander, 2001; Talanquer et al., 2015). These norms may be communicated to students along with specific disciplinary practices and content

(Louca et al., 2004; Rosenberg et al., 2006; Russ, 2018). For example, a teacher might respond to a student's answer with a comment like, "That is correct, but how did you arrive at that answer?" Such a comment conveys that reasoning is equally or perhaps more important than a knowledge claim. Like epistemic cognition in daily life, some of what constitutes appropriate use of knowledge in class is communicated tacitly, for example, based on which questions are asked on assessments and which responses to those questions earn points (Entwistle, 1991; Momsen et al., 2013; Scouller, 1998; Scouller & Prosser, 1994; Snyder, 1973).

The epistemic learning that occurs in the classroom has implications for how students understand and apply what they are learning in the course, in future courses, and in other aspects of their lives. Hammer et al. (2005) argue that the conceptual knowledge a person activates in a given context is influenced by the epistemic knowledge they are drawing upon. For example, the role of epistemic knowledge as a "control structure" (Bing & Redish, 2009) was used to understand why a student in an introductory physics class did not see her intuition and everyday experiences as allowed sources of knowledge (Lising & Elby, 2005) and to explain the teacher-initiated shifts in reasoning observed in a group of eighth graders discussing the rock cycle (Rosenberg et al., 2006). These studies suggest that in order to support student learning and transfer of knowledge, which are stated goals of virtually all STEM education reform efforts (National Research Council, 2012; American Association for the Advancement of Science, 2011; Cooper & Klymkowsky, 2013; Talanquer & Pollard, 2017), instructors and researchers need to attend to students' epistemic cognition.

To support further research on students' epistemic cognition, we conducted a review of the work that has been done so far in the context of undergraduate chemistry courses. We chose to focus most of our attention on the model of epistemic cognition used, as this informs how data is collected and interpreted. We found, however, that most studies did not include an explicit theory of epistemic cognition. Therefore, we compared three prominent models of epistemic cognition in the literature and identified major differences between them that would likely be evident in a study. We developed a coding scheme based on these differences and used it to infer models of epistemic cognition in the articles we analyzed. We present the results of our analysis and, drawing on scholarship the broader field of epistemology research, offer recommendations for future research on epistemic cognition in chemistry education.

Literature Background

Models of Individual Epistemic Cognition

Personal epistemology research concerns how individuals think about, construct, and evaluate knowledge. It draws on scholarship across several fields including philosophy, the learning sciences, psychology, and discipline-based education. Several different terms have been used to describe an individual's ideas about knowledge, including "epistemic beliefs" (Hofer & Pintrich, 1997), "epistemic cognition" (Kitchener, 1983; Greene et al., 2008; Sandoval, 2016), "epistemic resources" (Hammer & Elby, 2002), and "epistemic games" (Collins & Ferguson, 1993).

Given the variety in terminology, it is perhaps unsurprising that multiple ways of modeling epistemic cognition have been developed. Below we describe the three classes of models that have emerged, along with prominent examples for each (Table 3.1). For more detailed descriptions and examples, see the reviews published by Hofer and Pintrich (1997) and Sandoval et al. (2016).

Table 3.1. Models of Epistemic Cognition.

Type of Model	Description	Literature Examples
Developmental Models	Development of a person's epistemology proceeds through stages of increasing sophistication	Perry (1970) Kuhn (1999) King & Kitchener (1994)
Dimensional Models	A person's epistemology consists of multiple aspects (*e.g.*, simplicity of knowledge, certainty of knowledge, sources of knowledge, justifications for knowledge), each of which can vary in sophistication independently of the others.	Schommer-Aikins (2004) Hofer & Pintrich (1997)
Resources Models	A person's epistemology is constructed in the moment and the sophistication or utility varies according to the situation. These models typically organize epistemic ideas into categories (*e.g.*, epistemic aim, epistemic form).	Hammer & Elby (2002) Chinn, Buckland, & Samarapungavan (2011)

Developmental Models. Early models of epistemic cognition reported in the literature can be classified as developmental. A frequently cited example is Perry's scheme of intellectual and ethical development (1970). Through interviews with male college students, he proposed a developmental progression in which students initially view knowledge as objective and unchanging and over time develop an understanding of knowledge as contextual and tentative. Later scholars built upon Perry's work and proposed similar developmental models. Based on her work on reasoning in everyday life, Kuhn (1999) detailed three stages of epistemological development: absolutist, multiplist, and evaluatist. The reflective judgement model developed by King and Kitchener (2004) describes epistemic cognition for thinking about ill-structured problems and contains seven stages grouped into three broad categories: pre-reflective, quasi-reflective, and reflective. Underlying these developmental models is the assumption that different aspects of epistemology (such as the nature of knowledge and sources of knowledge) are correlated and progress in tandem.[1] Furthermore, as a person's epistemology develops, it becomes more sophisticated or expert-like.

Dimensional Models. Later researchers questioned the assumption of correlation and proposed models in which epistemology is characterized along separate, independent dimensions rather than a single developmental sequence. For example, Schommer-Aikins proposed five dimensions: simple knowledge, certain knowledge, source of knowledge, ability to learn, and quick learning (2004).[2] In their 1997 review on personal epistemology research, Hofer and Pintrich synthesized ideas from developmental and dimensional models and arrived at a model consisting of four dimensions: simplicity of knowledge, certainty of knowledge, source of knowledge, and justification of knowledge. Although the dimensional models detangle various aspects of epistemology, they still ascribe to the assumption that epistemological beliefs lie on continuums of increasing sophistication.

Resources Model. More recently, Hammer and Elby rejected the idea that some epistemic ideas are inherently more sophisticated than others (2002). In their proposed resources model, a person's epistemology is constructed in the moment from smaller-grained epistemological resources (Hammer & Elby, 2002). Activation of these resources is dynamic and may change in a matter of seconds in response to cues from the environment. While these epistemological resources can be grouped into categories, such as form of knowledge or stance toward knowledge, the resources within a category do not exist on a sophistication continuum. Instead, Hammer and colleagues argue that some epistemological resources may be productive (*i.e.*, useful in progressing toward a goal) in one circumstance while a different set of epistemological resources may be productive in another. Thus, epistemological sophistication, according to Elby and Hammer (2001), "consists of having resources to sort out the complexity of knowledge in different contexts" rather than having "a global, decontextualized opinion about [an] issue." They point out that while most developmental and dimensional models consider constructing one's own knowledge more sophisticated than receiving knowledge from authority, there are times where this is not necessarily true. For example, they argue that it is probably more worthwhile to accept the biologists' claim than cows have multiple stomachs rather than go out and dissect one yourself. Likewise, Muis et al. (2006), Chinn et al. (2011), and Sandoval et al. (2016) have also argued for a context-dependent view of epistemic cognition.

Key Assumptions in Models of Epistemic Cognition

Developmental and dimensional models can be distinguished from the resources model of epistemic cognition by attending to differences in a few key assumptions (Table 3.2). These assumptions manifest themselves in the way data is collected and analyzed, as described in the following sections.

Table 3.2. Comparison of models of epistemic cognition with regards to hierarchy, stability, and explicitness

Model of Epistemology	Hierarchy	Stability	Explicitness
Developmental	hierarchical	stable	explicit[a]
Dimensional	hierarchical	stable	explicit[a]
Resources	variable utility	unstable	implicit

[a] Not inherent to model but consistent with how most studies using this model have been conducted.

Hierarchy. Approaches to evaluating students' epistemic ideas can vary significantly depending on the model of epistemic cognition. Developmental and dimensional models organize epistemic beliefs in a hierarchical manner. Research conducted according to these models typically seeks to assign students' epistemic beliefs to levels or stages and evaluate interventions designed to advance students toward more sophisticated or expert-like epistemic beliefs. The resources model, on the other hand, contends that no epistemic idea is universally better than another. The research focus is therefore on understanding the interplay between contextual factors and students' epistemic cognition.

Stability. A second point of difference between models concerns the stability of students' epistemic cognition. Researchers employing a developmental model typically treat a person's epistemic beliefs as relatively stable over long periods of time. Dimensional models also treat epistemic beliefs as having trait-like or theory-like characteristics and thus assume they are relatively stable over time. This assumption is evident in the frequent collection of pre- and post-test data, often at the beginning and end of a course. In contrast, resources models assume epistemic cognition is often unstable and can shift rapidly in response to comments from a teacher or peer, for example. Importantly, resources models do not preclude epistemic stability—if one finds a set of resources is frequently useful in a given context, they may consistently activate these resources in contexts that (implicitly) seem similar. Researchers employing a resources model thus tend to collect data over short periods of time, such as one class period.

Assumptions of stability also manifest themselves in the implied generalizability or specificity of epistemic ideas. Some researchers utilizing developmental or dimensional models expect epistemic beliefs to be stable across contexts, as indicated by the domain-general descriptions employed (*e.g.*, absolutists believe in one knowable truth). They would expect students to answer the same way whether a survey is given in a science class, a math class, or an English class. Others have limited their claims to a particular area of knowledge. In fact, Nature of Science (NOS) research has emerged as a somewhat separate field of study (Lederman, 1992). Resources models go even further, contending that researchers should not assume the same epistemological resources are activated in all chemistry classes. Thus, studies that use a resources model typically focus on generating or expanding upon theories rather than obtaining statistically generalizable results.

Explicitness. One additional assumption worth highlighting regarding the nature of epistemic cognition concerns how it can be studied. A resources model of epistemic cognition regards activation of epistemological resources as a largely subconscious process (Hammer & Elby, 2002). As such, evidence for students' epistemic cognition is best obtained by observing students' behavior in the situation of interest (*e.g.*, classroom interactions). Developmental and dimensional models do not discuss whether epistemic cognition is tacit or not, but historically scholars ascribing to these models have probed epistemic beliefs through surveys and interviews in which participants are asked about their beliefs directly. It is assumed, often without strong or clear evidence, that the correlation between self-reported epistemic beliefs and epistemic beliefs inferred from observed behavior is strong.

Social Epistemology

Historically, researchers sought to characterize an individual's epistemic beliefs. Like the emergence of social constructivism from constructivism, epistemology researchers began to emphasize in published research the role others, such as the classroom community or society more broadly, play in shaping an individual's epistemic cognition. This has given rise to social epistemology, the study of how people collectively determine how knowledge is created and evaluated (Schmitt, 2017). In the context of education, researchers draw upon social epistemology to understand how classroom communities negotiate epistemic norms. Their emphasis is on the interactions between individuals and between individuals and the wider cultural context in which their education takes place, rather than on the individual's thoughts and behaviors (Sandoval et al., 2016).

Research Questions

Our first goal in this review, intended primarily for researchers, is to describe how studies on undergraduate students' epistemic cognition have been conducted. We were guided by the following questions:

1) What student populations are the samples drawn from?
2) Does the study focus on the epistemic ideas of individual students or groups of students?
3) What methodologies and study designs have been employed?
4) What means of data collection have been used?

Our second goal is to discuss the models of epistemic cognition that explicitly or implicitly informed studies on undergraduate chemistry students' epistemic cognition. Since most articles did not report a model of epistemic cognition, we attempted to infer the model by addressing the following questions:

1) Do chemistry education researchers characterize epistemic ideas as hierarchical in nature or as varying in utility depending on context?
2) Is epistemic cognition assumed to be stable over time and/or across contexts?
3) Did researchers infer students' epistemic ideas from explicit statements about knowledge and knowing or from observations of behavior or interpretation of students' statements?

We intentionally do not summarize the findings of the studies, largely, because there was so much variation in what it meant to study epistemic cognition. It is difficult to compare, for example, a study that characterized the class consensus on appropriate justifications for arguments with a study that documented changes in individual students' Likert-scale responses to items describing the simplicity of knowledge. Without more consensus regarding the nature of epistemic cognition and how it should be studied and measured, it seems unproductive, and in some cases impossible, to synthesize results across studies.

Methods

Selection of Articles

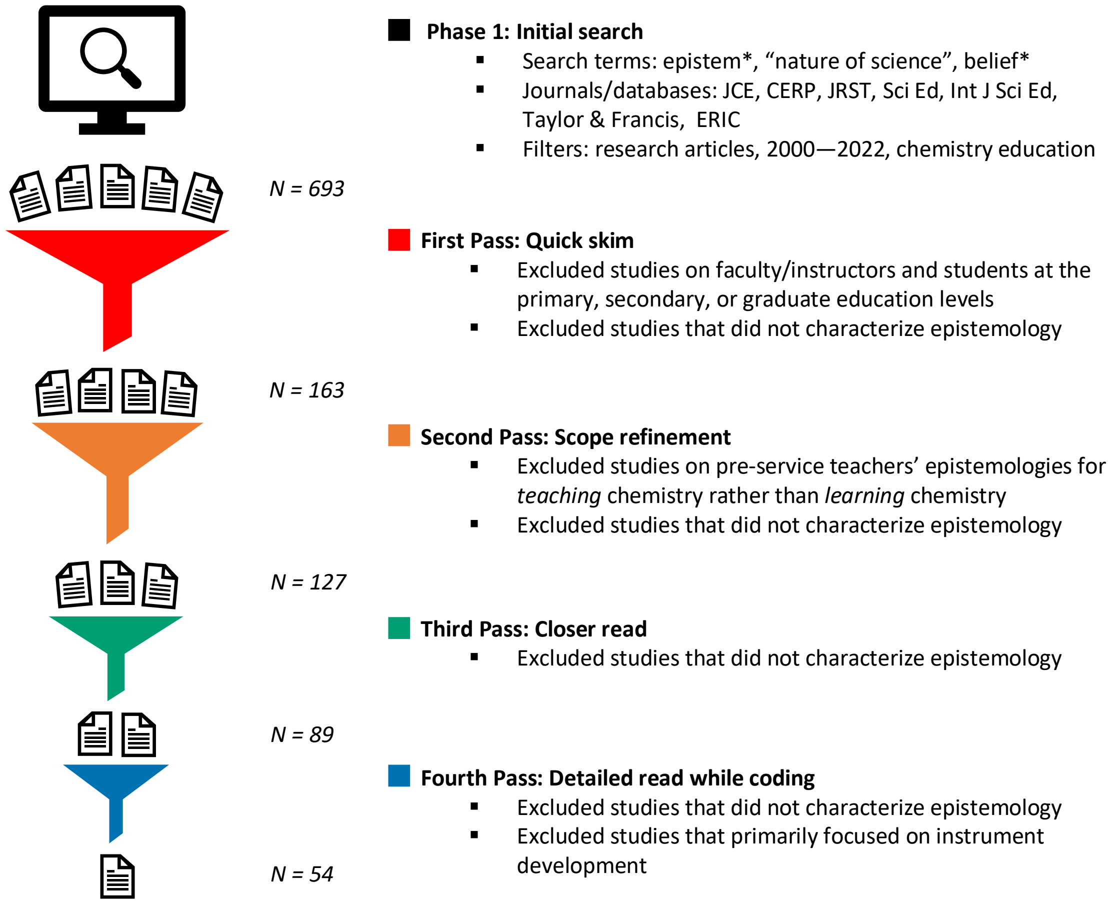
For this review, we chose to focus on undergraduate chemistry students' epistemic cognition. Therefore, we started by collecting articles published in the chemistry education journals *Journal of Chemical Education* and *Chemistry Education Research and Practice*. We then expanded our search to more general science education journals: *Journal of Research in Science and Teaching*, *Science Education*, and *International Journal of Science Education*. Finally, we searched the ERIC and Taylor & Francis databases.

We used several search terms to find articles that studied epistemology in the context of chemistry courses. We used the search term "epistem*" to find articles containing words related to epistemology, such as "epistemic" and "epistemological." We also searched "nature of science," as this research often looks at how people perceive scientific knowledge and how that knowledge is obtained. We also used the more general search term "belief" because we anticipated that many articles would describe students' beliefs concerning chemistry knowledge and knowing without using the term "epistemic" or "epistemological" to describe these beliefs.

Since some of these searches yielded a large number of hits, we applied a few filters to narrow down the number of results. We decided to limit the scope of this review to papers published between 2000 and 2022. For journals or databases that included multiple types of publications, we restricted the search results to research articles (as opposed to publications describing activities, for example). For the science education journals and databases, "chemistry" was also entered as a search term. The ERIC database also contained the option to restrict results to those tagged as "Higher Education," which was helpful in eliminating articles focused on K-12 education.

Similarly, we used the tag “Education” to narrow the scope of results obtained by searching the Taylor & Francis database.

With these filters in place, the initial searches for “epistem*,” “nature of science,” and “belief” yielded 693 unique articles. From here, we screened the articles manually through the iterative process shown in Figure 3.1. Since the scope of this review is limited to undergraduate students’ epistemologies, articles that collected data from K-12 students, graduate students, and teachers or faculty were removed. The first author then performed a keyword search on each article using the terms “epistem,” “nature of science,” and “belief” to determine whether these were the focus of the article or simply mentioned them in passing. For many of the articles, these terms were only found in the titles of referenced articles or mentioned in passing. For example, several articles utilized the resources theoretical framework, which encompasses both conceptual and epistemic resources; if the article then went on to characterize only conceptual resources, it was not included in this review.



⬛ **Phase 1: Initial search**

- Search terms: epistem*, “nature of science”, belief*
- Journals/databases: JCE, CERP, JRST, Sci Ed, Int J Sci Ed, Taylor & Francis, ERIC
- Filters: research articles, 2000—2022, chemistry education

🟥 **First Pass: Quick skim**

- Excluded studies on faculty/instructors and students at the primary, secondary, or graduate education levels
- Excluded studies that did not characterize epistemology

🟧 **Second Pass: Scope refinement**

- Excluded studies on pre-service teachers’ epistemologies for *teaching* chemistry rather than *learning* chemistry
- Excluded studies that did not characterize epistemology

🟩 **Third Pass: Closer read**

- Excluded studies that did not characterize epistemology

🟦 **Fourth Pass: Detailed read while coding**

- Excluded studies that did not characterize epistemology
- Excluded studies that primarily focused on instrument development

Figure 3.1. Article selection process. [a] Because epistemic cognition was often intertwined with other aspects of learning (*e.g.*, conceptual learning, affective outcomes), we continuously refined our inclusion criteria, resulting in an iterative selection process. See Supporting Information for more details.

Following the initial screen of search results, 163 articles remained. The first author skimmed through these to determine if the studies sought to characterize or measure students’ epistemologies. Another 36 articles were removed during this phase, resulting in 127 articles. From here, the first author read through each article in full to determine if it met the criteria described above and reduced the sample down to 85 articles. From there, we started to code the articles, and during this process, we eliminated an additional 31 articles. Some were eliminated because, upon a closer read, they did not focus on characterizing students’ epistemic cognition. A few others were eliminated due to a primary focus on instrument development. The total number of articles included in the analysis for this review is 54. A complete list can be found in the Supporting Information.

Analysis of Articles—Descriptive Codes

The first part of the analysis involved coding for who was being studied and how they were being studied. Five categories of codes were developed: study population, unit of analysis, methodology, study design, and data collection. The codes within each category are summarized in Table 3.3 and described in more detail below.

Table 3.3. List of codes used to describe methodological aspects of the studies.

Code	Description
Sample	
Intro chem	Study participants were recruited from a first-year chemistry, introductory chemistry, or general chemistry course for majors or non-majors.
Intro chem lab	Study participants were recruited from a first-year chemistry, introductory chemistry, or general chemistry laboratory course for majors or non-majors.
OChem	Study participants were recruited from an organic chemistry course for majors or non-majors.
OChem lab	Study participants were recruited from an organic chemistry laboratory course for majors or non-majors.
PChem	Study participants were recruited from a physical chemistry course or physical chemistry laboratory course.
Non-course specific	Study participants were recruited from multiple chemistry courses or were studied as they progressed through multiple chemistry courses.
Chem for pre-service teachers	Study participants were recruited from chemistry or science classes designed for pre-service teachers.
Unit of Analysis	
Individual	Data was collected on individual students.
Group	Data was collected on groups of students, ranging from pairs of students to whole classes.
Individual and group	Data was collected on both individuals and groups.
Methodology	
Qualitative	Non-numerical data, such as words or images, was collected and analyzed for themes, patterns, or features relevant to the research question.
Quantitative	Numeric data was collected and analyzed via statistical methods to determine relationships among variables.
Mixed methods	A combination of numerical and non-numerical data was collected and analyzed using both quantitative and qualitative techniques.

Study Design	
Exploratory	A phenomenon was explored on a small sample with no comparison groups or treatments administered.
Quasi-experimental	The impact of an intervention was assessed by: • comparing data collected before and after implementation of the intervention on a treatment group; • evaluating data collected after implementation of the intervention on a treatment group; • Comparing data collected on two or more treatment groups before and after implementation of an intervention; • Comparing data collected on two or more treatment groups after implementation of an intervention.
Longitudinal	Data was collected at three or more timepoints over a period of time (semester or longer) to understand and/or measure how an outcome variable changes.
Correlational	The quantitative relationship between two or more variables was determined.
Data Collection Method	
Interview	Researchers met with students and asked students to respond to questions orally or complete tasks. Interviews could be conducted with individual students or groups of students.
Open-ended survey	Students were asked to respond to a written or electronic set of questions using their own words.
Written artifact	Written (or electronic) work that students created as part of their regular coursework. These included laboratory reports, essays, worksheets, and exams.
Classroom recording	Video and/or audio recordings of the whole class or small groups of students, typically used to collect classroom dialogue.
Selected-response survey	Students were asked to respond to a written or electronic set of questions and/or statements by selecting the response that best aligns with their thoughts.

Sample. A simple coding scheme was developed to summarize the different subsets of undergraduate chemistry students represented in these studies. The code *Intro chem* describes courses labeled as general, introductory, or first-year chemistry and encompasses variations for chemistry majors, STEM majors, and non-majors. A separate code, *Intro chem lab*, is used for general or introductory chemistry laboratory courses. The codes *OChem* and *OChem lab* describe the organic chemistry lecture and laboratory courses, respectively, that students typically take during their second year of college or university. A few studies recruited students from an upper-level physical chemistry lecture or lab course; these were labeled *PChem*. No other upper-level chemistry courses were represented in our sample. The code *Undergrad chem* was applied to studies that recruited students from across different chemistry courses. Finally, the code *Chem for pre-service teachers* was assigned to studies that sampled students from chemistry classes designed for pre-service teachers.

Unit of Analysis. The unit of analysis code was implemented to distinguish studies on personal epistemology from those on social epistemology. Studies that collected data on each student, consistent with personal epistemology research, were coded as *Individual*. Studies that collected data on groups of students, consistent with social epistemology research, were coded as *Group*. A third code, *Individual and Group*, was included to describe studies that collected data from both individuals and groups of students.

Methodology. The articles included in this study were also characterized according to their methodologies and study designs. Methodology was described as *Qualitative*, *Quantitative*, and *Mixed methods*. Qualitative studies collect non-numerical data, such as classroom dialogue, that are analyzed by looking for themes, patterns, or other features relevant to the research question (Johnson & Christensen, 2020). Quantitative studies collect numerical data, such as exam scores, that are analyzed via statistical methods (Johnson & Christensen, 2020). Mixed methods studies utilize both types of data and analyses (Johnson & Christensen, 2020).

Study Design. As these methodology categories are quite broad, more specific codes for study designs were employed. *Exploratory* studies were defined as those that collected data from a single group, absent a treatment, with the goal of understanding some aspect of students' epistemic cognition. *Quasi-experimental* studies (Campbell & Stanley, 1963), on the other hand, seek to determine the effects of a treatment on students' epistemic cognition. We chose not to distinguish between studies that involved a single treatment group versus those that included a control group or multiple treatment groups, nor did we separate studies that used a pre- and post-tests from those that used only post-tests. We reasoned that all of these studies were united by a common goal-assessing the impact of some curricular and/or pedagogical change and the particulars of how they did so were not crucial to this review. *Longitudinal* studies focus primarily on understanding or documenting change over time. White and Arzi (2005) define a longitudinal study as "one in which two or more measures or observations of a comparable form are made of the same individuals or entities over a period of at least a year." We modified their criteria slightly when coding. We required that a study collect data at more than two timepoints to distinguish longitudinal studies from the many studies that used pre- and post-tests but were focused on the impact of an intervention rather than the dynamics of epistemic cognition. We also lowered the duration to a semester given that most undergraduate courses operate over a semester rather than a year. Finally, *Correlational* studies sought to demonstrate a quantitative relationship, or lack thereof, between two or more variables, at least one of which was epistemic.

Data Collection. We expanded upon the codes used by Rodriguez et al. (2020) when describing the different ways in which data on students' epistemic cognition was collected. *Interviews* consist of verbal responses to questions posed by the interviewer. This code included interviews conducted with individuals, pairs of students, and focus groups as well as various types of interviews, such as think-aloud (Charters, 2003), stimulated-recall (Dempsey, 2010), and task-based, cognitive clinical interviews (Ginsburg, 1997, Russ et al., 2012). *Open-ended surveys* included any written or electronic form in which students responded to questions in their own words. *Selected-response surveys* asked students to select a response from a given list. The code *Written artifacts* pertains to any work students submitted as part of the course, such as exams or laboratory reports. Finally, *Classroom recording* describes any audio and/or video recordings of the whole class or small groups of students. Because each study can collect data via multiple avenues, a single study could receive multiple data collection codes.

Analysis—Model of Cognition Codes

Part two of the analysis focused on characterizing each article's treatment of epistemic cognition in terms of hierarchy, stability, and explicitness. Codes belonging to each category are shown in Table 3.4 and described in detail below. When assigning codes, all sections in an article were considered, but the sections that described data collection and data analysis proved especially useful.

Table 3.4. List of codes used to describe the assumptions about the nature of epistemic cognition.

Code	**Description**
Hierarchy	
Hierarchical	Certain epistemic ideas are considered more sophisticated, expert-like, or desirable than others.
Variable utility	Context determines the value of epistemic ideas.
Ambiguous	No indication of value or conflicting statements concerning the relative values of epistemic ideas.
Stability	
Stable	Epistemic cognition is treated as stable across contexts and/or relatively long periods of time (*e.g*., semester, year).
Unstable	Epistemic cognition is treated as unstable across context and/or time (on the scale of minutes).
Unclear	No indication as to whether epistemic cognition is considered stable or unstable across context and/or time.
Explicitness	
Explicit	Participants are asked to respond to statements/questions in which they would need to be consciously aware of their own epistemic ideas.
Implicit	Epistemic ideas are inferred from participants' responses and/or behavior.
Both	Data is collected in multiple ways, some of which require participants' awareness and some of which are inferred.

Hierarchy. The first category of codes concerns the hierarchical nature of epistemologies. Some articles characterized students' responses according to varying levels of sophistication in which some epistemologies were deemed better or more desirable than others. These articles were given the code *Hierarchical*. Typically, the most sophisticated descriptor was applied to responses that aligned with those an expert chemist or scientist gave. For example, when developing the

Colorado Learning Attitudes about Science Survey for chemistry, Adams et al. (2008) administered the survey to chemistry faculty in order to define the expert response for each survey item. Student responses were then scored in terms of how closely they matched the expert responses. In other studies, the relative rank of descriptors seems to have been determined according to the authors' judgement. Other articles described epistemologies as more or less useful based on the particular circumstance rather than being universally better or worse; these were coded as *Variable utility*. Viewing epistemologies as of variable utility does not preclude situated hierarchies. In a given context, one may find some ways of knowing and learning more useful than others in advancing a particular aim. For example, one may find accumulating information from public health authorities more useful than building a model of disease spread from experience if one is trying to figure out whether to wear a mask to the supermarket. Variable utility does not mean all epistemologies are equally useful across all contexts. Finally, the code *Ambiguous* was used for articles that did not discuss the value of particular epistemologies or contained contradictory statements regarding value. For example, an article could invoke a resources model in the theoretical framework section but report a hierarchical coding scheme in the analysis.

Stability. The Stability category of codes was developed to describe how dynamic epistemic cognition was assumed to be, especially as it related to the timescale of each study. The code *Stable* was used for studies that seemed to view epistemic beliefs as relatively unchanging over long periods of time or in the absence of an intervention (Table 3.3). None of the articles included in our study directly stated an assumption of stability; rather, we inferred this from the methods used. Many of these studies utilized a pre-post design in which pre- and post-measures were spaced days, weeks, or months apart. Study designs of this sort, we argue, suggest that researchers expected any changes to occur over a long time period (typically a semester). If they anticipated epistemic cognition was dynamic and subject to change over the course of a single class period, they would likely collect data more frequently to distinguish signal from noise.

Views on stability (whether implicit or explicit) could also be inferred from how authors treat the influence of context on epistemic cognition. If the authors described shifts in epistemic ideas in response to different prompts or comments from peers, the articles were coded as *Unstable*. Finally, if a study made no mention of how stable or unstable they considered epistemologies to be across time and context, it was coded as *Unclear*. In a few cases, epistemic cognition was treated as stable over time but unstable across contexts; these were also coded as *Unclear*.

Explicitness. The third category of codes, Explicitness, was used to describe how data on epistemic cognition was obtained and interpreted. Some methods asked the participants to respond to direct questions about their epistemic ideas, which required them to consciously consider their own epistemic beliefs. The prompts often took the form of declarative statements, such as "Knowledge is obtained from authority," which requires little interpretation on the part of the researcher. Other studies collected data in the form of written artifacts or classroom recordings, from which the researchers needed to infer epistemic cognition indirectly. Although the amount of inference required to make a claim about students' epistemic ideas varies considerably, for the sake of simplicity, each type of data was simply coded as *Explicit* or *Implicit* based on how it was collected and analyzed. Since some articles utilized multiple data strands, the code *Both* was applied if both explicit and implicit methods of data collection and analysis were used.

Inferred Model of Cognition. By considering the Hierarchy, Stability, and Explicitness codes together, we were able to distinguish studies consistent with a developmental or dimensional model from those that were consistent with a resources model. When assigning a model to each study, we required all three codes to be consistent with that model. In addition, we tentatively assigned models to studies whose codes were mostly aligned with that model. For studies in which two codes were consistent with a single model and one was not associated with any model (*i.e.*, *Ambiguous*, *Unclear*, or *Both*), we tentatively assigned the model based on the two matching codes. For example, a study coded as *Ambiguous*, *Unstable*, and *Implicit* was determined to be mostly aligned with a resources model. Studies that were coded as *Hierarchical*, *Stable*, and *Implicit* were considered mostly aligned with a developmental or dimensional model, since the association of these models with the *Explicit* code was based on literature trends rather than descriptions of the models themselves. For the remaining studies that did not fall into any of the categories described above, we were unable to assign a model as we did not have enough evidence or had contradictory evidence.

Establishing Trustworthiness

Initial coding was carried out by both authors. They individually coded ten articles at a time and then met to compare codes, resolve any discrepancies, and clarify the codebook. This was repeated three times. Subsequently, the first author read each article multiple times over the data analysis period, highlighting and annotating the parts of the text that provided evidence for each code assignment. The second author was brought in to discuss any codes the first author was unsure of.

There is no agreed upon method for establishing trustworthiness in qualitative research (Rolfe, 2006). We chose to establish trustworthiness primarily by recording a detailed decision trail, as recommended by Noble and Smith (2015). This was especially helpful for distinguishing codes based on the presence of evidence (*hierarchical, variable utility, stable, unstable, explicit, implicit, both*) from those based on lack of evidence (*ambiguous, unclear*). We chose not to report inter-rater agreement statistics because reaching an acceptable level of inter-rater agreement typically involves several rounds of code refinement on new subsets of data, which was impractical with our limited data set. Furthermore, this is consistent with several other chemistry education review articles (*e.g*., Flaherty, 2020; Hunter et al., 2022; Bain et al., 2014). The full coded data set with researcher notes may be found in the Supporting Information.

Findings

In the first part of our analysis, we sought to describe who was being studied and how they were studied in chemistry education epistemology research. We will report our findings for each category of codes and describe examples of each.

The majority of studies focused on students in first-year chemistry courses and pre-service teachers.

Figure 3.2 displays the distribution of Sample codes, which describe the chemistry courses from which study participants were drawn. First-year chemistry students were the most studied population among the studies included in this review. Twenty studies sampled students from lecture courses while seven sampled students from the lab component specifically. There was some variation as to the students who were taking these first-year chemistry classes. Some were courses designed for non-majors, some were designed for STEM majors, and others were open to all majors. Pre-service teachers enrolled in chemistry or science courses for pre-service teachers were the second-most common group studied ($N = 7$). Four studies recruited students who were studying to become chemistry teachers specifically (Venessa et al., 2019; Ağlarcı et al., 2016; Çelik, 2020; Sendur et al., 2017). The other three studies sampled pre-service elementary (or primary) school teachers enrolled in chemistry or science courses specifically designed for them (Crujeiras-Pérez & Brocos, 2021; McDonald, 2010; Çalik & Cobern, 2017). Several studies recruited participants from organic chemistry lecture ($N = 6$) or laboratory ($N = 3$) courses. Just three studies involved students from upper-level chemistry courses; interestingly, these were all physical chemistry lecture or laboratory courses. Other studies sought to understand chemistry students' epistemic cognition outside of the context of a specific class and recruited students from all levels of chemistry (*e.g.*, Sevian & Coutre, 2018; Li et al., 2013) or compared students enrolled in introductory and organic chemistry courses (*e.g.*, Hofer, 2004; Mazzarone & Grove, 2013).

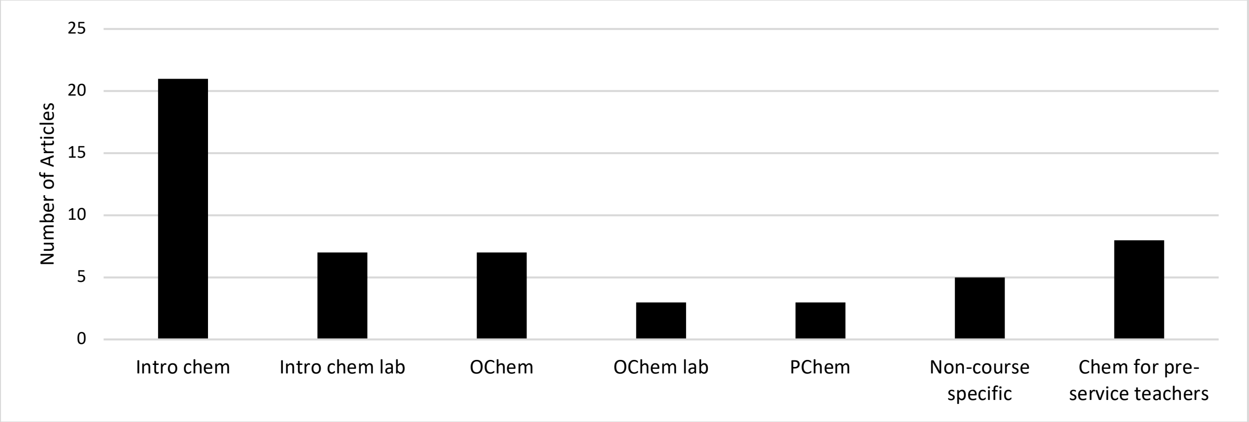


Figure 3.2. Distribution of Sample codes.

The reasons offered for choosing a particular course to study varied. For studies focused on assessing the impact of an intervention (see below), course selection depending on where the intervention was being implemented. Sometimes these interventions were initiated by an individual instructor; other times they were part of a larger department initiative (*e.g.*, Chopra et al., 2017). Some studies on pre-service teachers' epistemic cognition were in part motivated by the role of teachers in shaping students' perspectives on chemistry and science more broadly. As Ağlarcı et al. (2016) argue, "science education programs and teachers play a key role in [improving citizens' images of science], as they are mostly responsible for educating people."

Most studies examined personal epistemology rather than social epistemology.

As shown in Figure 3.3, most studies (N = 38, 72%) collected data from individual students. Individuals responded to surveys, participated in interviews, or wrote their own lab reports or essays. These were used to make inferences about the individual's epistemic cognition. Ten studies characterized aspects of epistemology that belonged to a group of students (Fig. 3.3). These tended to be studies that focused on group dialogue, especially argumentation. Six studies collected data on individuals as well as groups of students (Fig. 3.3). For example, Walker et al. (2019) studied argumentation and included in-person argumentation (group dialogue) as well as written arguments (individual lab reports). These results demonstrate that the interplay between students, instructors, and the cultural context in which learning occurs is understudied in chemistry education research on epistemology and may be a productive avenue for future research.

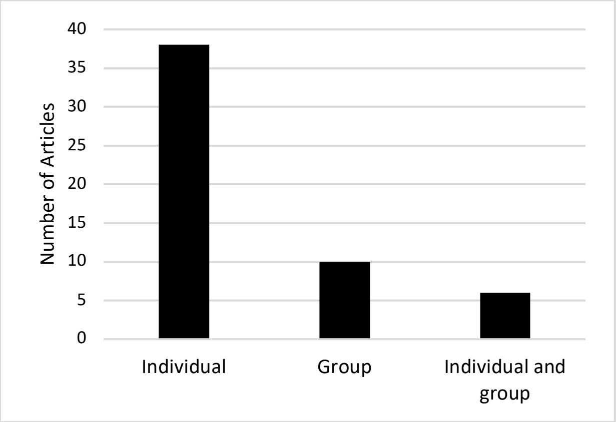


Figure 3.3. Distribution of Unit of study codes.

Exploratory qualitative studies are the most common types of studies on students' epistemic cognition.

Nearly half (*N* = 24) of the studies surveyed were classified as exploratory (Fig. 3.4). Twenty of these were qualitative while the remaining four used a mixed methods approach. The targets of exploration varied considerably. Some focused on specific aspects of epistemic cognition such as students' epistemic stances when evaluating models (Kelly et al., 2021) or what students considered acceptable justifications for their knowledge products (Crujeiras-Pérez & Brocos, 2021; Becker et al., 2013). Others characterized epistemic cognition more broadly by attending to students' perspectives on the nature of science (*e.g.*, Agustian, 2020, Venessa et al., 2019).

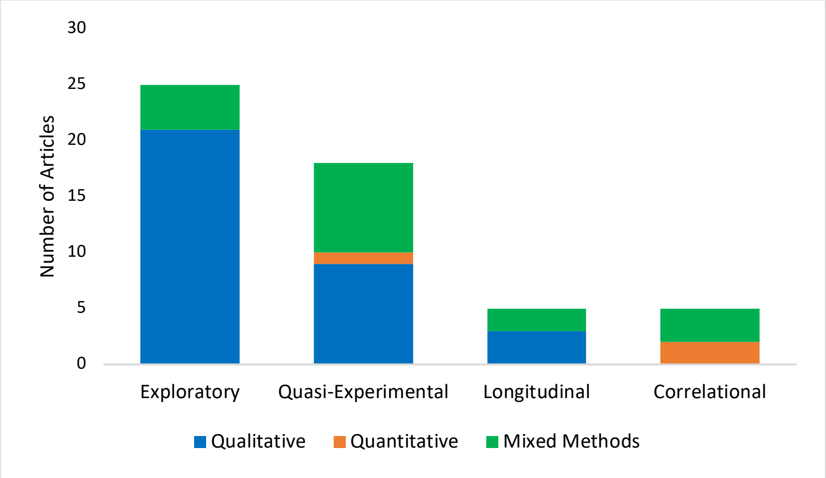


Figure 3.4. Distribution of Study Design codes, separated by Methodology code. Blue corresponds to Qualitative studies, yellow corresponds to Quantitative studies, and green corresponds to Mixed methods studies.

A large portion (N = 19) of studies were carried out to assess the impact of a curricular or pedagogical intervention (Fig. 3.4). Ten of these were qualitative, eight were mixed methods, and one was quantitative. Interventions included using explicit approaches to teaching nature of science (*e.g.*, Celik, 2020), incorporating inquiry-based laboratory experiments (*e.g.*, Russell & Weaver, 2021), and implementing a new curriculum (*e.g.*, Bowen et al., 2022). Most used a pre-post design in which data was collected from students before and after the intervention. Some of these involved a control group and a treatment group while some only had a treatment group. A few studies only collected data after the intervention.

Five longitudinal studies were present in the sample (Fig. 3.4). Three were qualitative, and two were mixed methods. Two of these sought to understand how students' understanding of argumentation, including epistemic aspects, changed as the students participated in an argument-driven inquiry general chemistry laboratory course (Hosbein et al., 2021; Walker & Sampson, 2013). The remaining three examined students' epistemic cognition more generally, either over two semesters of organic chemistry (Grove & Bretz, 2010; Grove & Bretz, 2012) or over the general chemistry and organic chemistry sequence (Mazzarone & Grove, 2013).

Correlational studies made up the remaining five studies (Fig. 3.4). In some, the epistemic variable was general. For example, Lee et al. (2022) looked at the relationships among epistemic beliefs, engagement in a flipped chemistry class, and learning outcomes. Other studies examined epistemic variables grounded in chemistry. Li et al. (2013) compared students' conceptions of learning chemistry to their approaches to learning chemistry and found some correlation. Aguirre-Mendez et al. (2020), measured the relationship between argumentation quality and gains in chemistry content knowledge. All of the correlational studies were quantitative or mixed methods.

Data on students' epistemic cognition was collected through a variety of methods.

No clear preferences for one method of data collection over others were found in the reviewed articles, as shown in Figure 3.5. Interviews (*N* = 25) and open-ended surveys (*N* = 21) were the most common approaches to eliciting data on epistemic cognition. Written artifacts (*N* = 14) and classroom recordings (*N* = 13) were slightly less common, although both were still used in over a quarter of the studies. Selected-response surveys (*N* = 6) were the least common means of data collection. Thirty-four studies relied on a single method of data collection while 19 elicited data using more than one method. Written artifacts were mostly used in combination with other data sources while selected-response surveys tended to be used alone. Interviews, open-ended surveys, and classroom recordings were used approximately evenly as the sole data source or in tandem with other sources. Given the complex nature of epistemic cognition, considering the pros and cons of what each data source can offer seems prudent when designing a study.

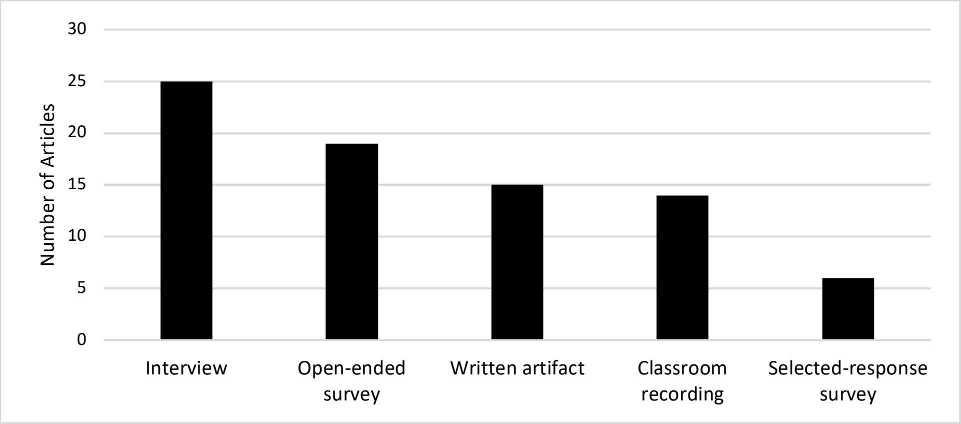


Figure 3.5. Methods of data collection. The total number exceeds the number of articles because some studies collected multiple types of data.

In the second phase of our analysis, we coded for assumptions about the hierarchy and stability of epistemic ideas as well as the extent to which epistemic ideas asked for directly during data collection or inferred during data analysis. We will report our findings for each category of codes and describe examples of each.

Most studies characterized students' epistemologies in a hierarchical manner.

The vast majority of articles (*N* = 39, 72%) interpreted students' epistemologies using hierarchical coding schemes (Fig. 3.6). The Views on the Nature of Science (VNOS) instrument (Lederman et al., 2002) used in nine of the studies interprets responses as naïve or informed (or somewhere in between, depending on the particular study). Other studies employed Likert scale survey items in which higher (or sometimes lower) scores were indicative of more expert-like responses. For example, Shultz and Gere (2015) asked general chemistry students to respond to the question, "When two different theories arise to explain the same phenomenon, what should scientists do?" The students were given several responses to this question which they were asked to rate on a five-point Likert scale from strongly disagree to strongly agree. Students who strongly agree with the statement, “Scientists should not accept any theory before distinguishing which is best through the scientific method because there is only one truth about phenomenon” were classified as having naïve views on the nature of science as it relates to the certainty of knowledge. Conversely, students who strongly disagreed with the statement were characterized as having sophisticated views related to certainty of knowledge.

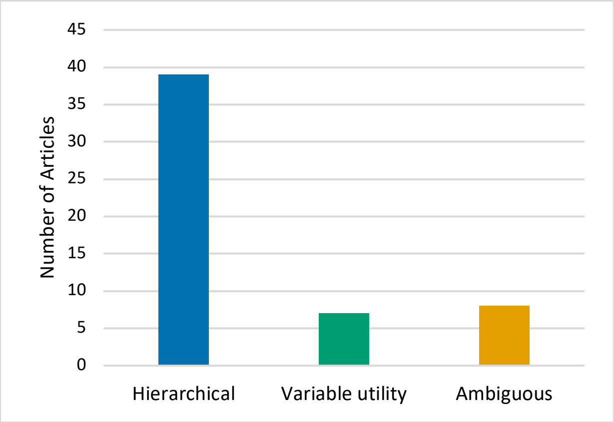


Figure 3.6. Distribution of hierarchy codes. Hierarchical (blue) aligns with a developmental or dimensional model. Variable utility (green) aligns with a resources model.

Far fewer studies (*N* = 7, 13%) explicitly stated that different epistemologies were useful in different circumstances. Three of these invoked a resources model of epistemic cognition as part of their theoretical frameworks. For example, Rodriguez et al. (2020) used the framework of epistemic games to analyze how students solve kinetics problems during think-aloud interviews. In their discussion of the results, the authors state, “Thus, specific epistemic games are not problematic on their own, but issues arise when students have difficulty switching between games.” One epistemic game is not inherently better than another, but one might be more appropriate for a specific use or type of problem than another. This attention to context and the productive use of knowledge results in quite different implications for teaching. Instead of channeling students toward a single epistemic belief or set of beliefs, the emphasis is on helping students recognize which are appropriate for the problem at hand.

Interestingly, all of the studies coded as *Variable utility* were qualitative, and most focused on a single episode of problem-solving, either in a classroom or interview setting. While it is difficult to imagine a quantitative study that would be coded as *Variable utility,* a mixed methods study could in principle be useful for synthesizing these individual moments over the course of a semester or comparing across classes.

There were also some studies (*N* = 8, 15%) in which no judgements were made regarding the value of the epistemic ideas elicited. For example, Talanquer (2010) categorized students' explanations in terms of their structures as non-causal, macrocausal (additive or interactive), and microcausal (additive or static). It was unclear, however, whether particular types of explanation structures were more desirable than others.

An approximately equal number of studies treated epistemology as stable versus unstable.

Stability was the second feature we looked for to aid in inferring the model of epistemic cognition for each paper. We found less evidence on which to make claims about stability than we found for hierarchy, resulting in 16 (30%) coded as *Unclear* (Fig. 3.7). Of the remaining studies, 19 (35%) were coded as *Stable* and 19 (35%) were coded as *Unstable* (Fig. 3.7).

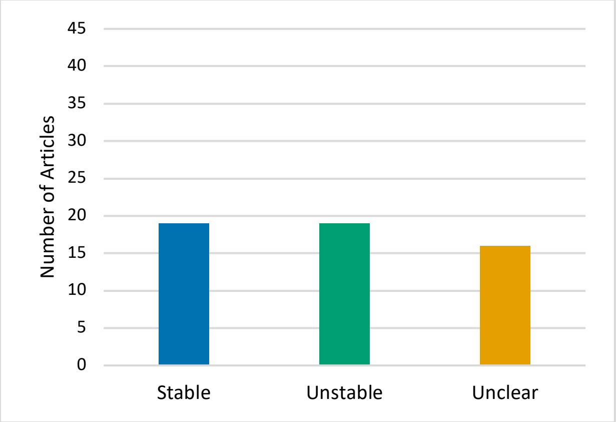


Figure 3.7. Distribution of Stability codes. Stable (blue) aligns with a developmental or dimensional model. Unstable (green) aligns with a resources model.

Studies that utilized a pre-/post-test design made up the majority in the *Stable* category. Often these kinds of study designs, which were common in our data set, are used to evaluate the effect(s) of an intervention. Examples of interventions related to epistemology or the nature of science include implementation of direct instruction on history of science or nature of science (*e.g*., Ağlarcı et al., 2016), participation in argument-driven inquiry labs (*e.g*., Hosbein et al., 2021), and metacognitive interventions (*e.g*., Saribas et al., 2013).

In the *Unstable* category, instability was often connected to the context sensitivity observed. For example, Lazenby et al. (2020) conducted a study on students' epistemic criteria for scientific models and found that "although students' conceptual resources are potentially productive, they are highly sensitive to context, as evidenced by the variation in themes across domain-general and chemistry-specific tasks." They noticed that students invoked different criteria depending on the type of model they were thinking about in that moment.

Studies that were coded as *Unclear* are somewhat difficult to describe as this code was largely based on the absence of evidence rather than the presence of certain features. Typically, these studies involved data collection at a single timepoint and did not discuss how context may have influenced the data. A few studies that were coded as *Unclear* described epistemic cognition as stable over time but unstable across contexts. For example, McDonald (2010) used a pre-/post-test experimental design to explore students' views on the nature of science, implying stability over time. However, she observed that asking about nature of science in the context of socioscientific versus scientific contexts elicited different responses from some participants. She interpreted this discrepancy as evidence that students possess multiple epistemologies, some general and some specific to science, implying instability across contexts.

Data on students' epistemologies was collected using an approximately equal number of explicit and implicit probes.

Finally, we examined how evidence of students' epistemologies was elicited to provide insight into the model of cognition used. We asked ourselves, when reading about the method of data collection, "Would a student have to think consciously about their own epistemology to provide this data?" If the answer was yes, we coded it as *Explicit* and if the answer was no, we coded it as *Implicit*. In doing so, it was helpful to consider the manner in which data was collected. Therefore, we present the results accordingly.

Approximately half of the studies included in our sample collected data using only explicit probes ($N = 25$, 46%). Most studies that collected data by administering surveys were coded as Explicit (Fig. 3.8). Many selected-response surveys, like the CHEMX survey used by Mazzarone & Grove (2013), asked students the extent to which they agreed with statements like "Knowledge in chemistry consists of many pieces of information, each of which applies primarily to a specific situation." Other selected-response surveys asked students to choose the option that most closely aligned with their views. For example, Venessa et al. (2019) used a mostly multiple-choice survey to ask students about purpose of science, the nature of scientific knowledge, and the relationship between science and technology.

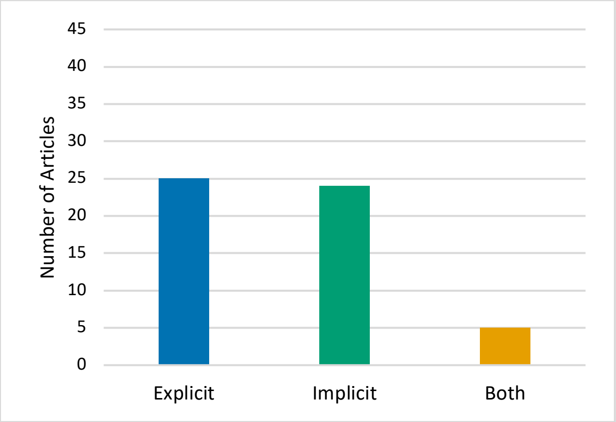


Figure 3.8. Distribution of Explicitness codes. Explicit (blue) aligns with a developmental or dimensional model. Implicit (green) aligns with a resources model.

Explicit items were often found on open-ended surveys and in interview protocols as well. The VNOS survey was used in nine different studies and contained explicit questions like, "What, in your view, is science? What makes science (or scientific discipline such as physics, biology, etc.) different from other disciplines of inquiry (*e.g.*, religion, philosophy)?" Similar kinds of questions were asked in interviews. For example, Havdala & Ashkenazi (2007) asked questions like "How would you define science?" and "Is there any way to find objectivity or certainty in science?" in their interviews with students enrolled in a general chemistry laboratory course. Like the selected-response surveys, these questions require a person to consciously consider the ways in which they think about and/or use knowledge.

Twenty-four studies (44%) were coded as *Implicit* (Fig. 3.8). These studies largely collected data in the form of written artifacts or classroom recordings. Written artifacts included reflective essays (*e.g.*, Grove & Bretz, 2012), argumentative writing assignments (*e.g.*, Moon et al., 2019), and lab reports (*e.g.*, Petritis et al., 2021). Classroom recordings typically captured student dialogue as they engaged in problem-solving (*e.g.*, Wickman, 2004) and/or argumentation (*e.g.*, Walker et al., 2019). These sources of data typically provided information about the structure of knowledge or justifications for knowledge. Scientific arguments, written or verbal, were commonly analyzed for the presence of and relationships between claims, evidence, and reasoning, originally derived from Toulmin's model of an argument (1958).

The *Implicit* code was also applied to some studies that used open-ended surveys and interviews. These typically contained questions that asked students to reflect on their experiences in class or engage in problem-solving. For example, one of the questions Bowen et al. (2022) asked students was "What would you tell [a student thinking about enrolling in organic chemistry] is the most difficult thing about organic chemistry?" Some of the responses to this question were epistemic in nature and revealed challenges related to obtaining or using knowledge. Kelly et al. (2021) conducted interviews in which they first asked students to watch a video on precipitation reactions, then think aloud as they completed a card sort and modeling task to describe the mechanism of precipitation, and finally critique three mechanistic animations for their scientific accuracy. From these interviews, Kelly et al. inferred students' epistemic activities, such as comparing and modeling, and their epistemic stances, such as doubting or puzzlement.

Twenty-one studies collected multiple strands of data, but only five studies (9%) used a combination of explicit and implicit methods of data collection and received the code *Both* (Fig. 3.8). For example, McDonald (2010) surveyed students about their views on the nature of science using the VNOS (*Explicit*) but also looked at students' written and verbal scientific arguments surrounding scientific issues (*Implicit*). Grooms (2020) also combined students' written scientific arguments (*Implicit*) with a survey that asked students explicitly about their epistemic ideas related to argumentation (*e.g.*, What is evidence?).

A model of epistemic cognition can be inferred for some studies.

As mentioned previously, most researchers did not frame their studies through the lens of a particular model of epistemic cognition. By considering the set of codes each study received, we acquired some evidence for what model was tacitly informing each study. Figure 3.9 depicts the relative number of co-occurrences for each coding combination to help visualize the relationships between codes.

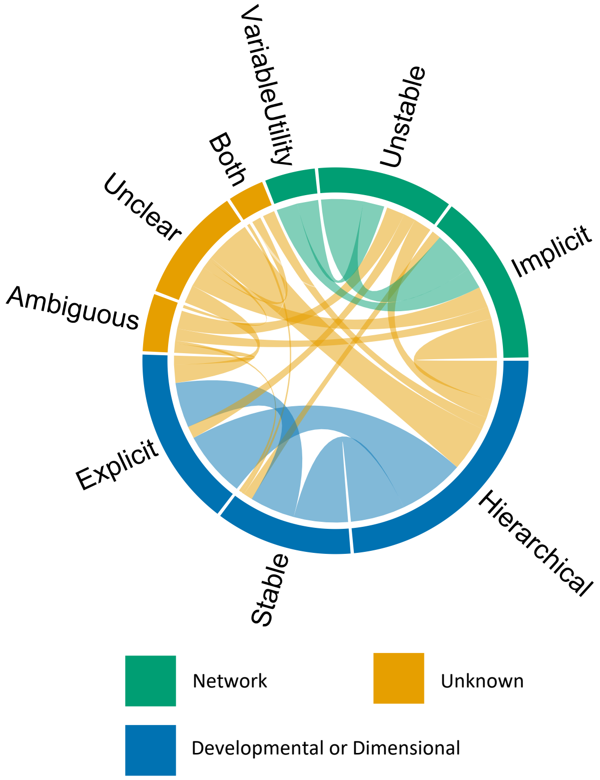


Figure 3.9. Chord Diagram depicting connections between codes. The width of each link is proportional to the number of studies that received the two codes connected by the link. Blue codes align with a developmental or dimensional model. Green codes align with a resources model. Yellow codes do not align with any model.

Developmental and dimensional models were characterized by *Hierarchical*, *Stable*, and *Explicit* codes (Fig. 3.8). Fifteen studies received these codes, completely aligning with a developmental or dimensional model. An additional eight mostly aligned with these models. Three were coded as *Hierarchical, Stable*, and *Implicit* or *Both*, and five were coded as *Hierarchical*, *Unclear*, and *Explicit*. In total, 23 studies seemed to be informed by a developmental or dimensional model (Fig. 3.9).

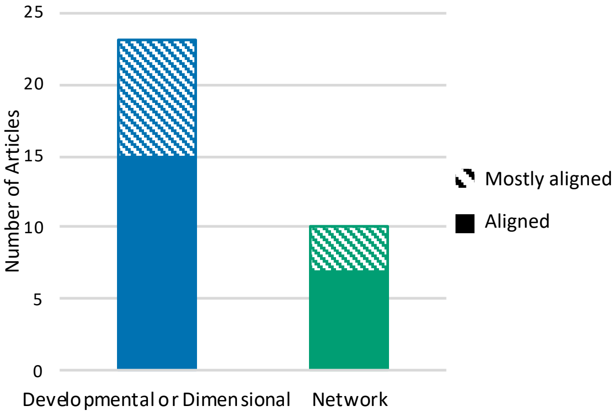


Figure 3.10. Models of epistemic cognition assigned based on Hierarchy, Stability, and Explicitness Codes. Solid bars represent articles that align with a model in all three codes. Striped bars represent articles that align with a model based on their Hierarchy and Stability codes.

The resources model asserts that epistemic cognition is context-dependent, dynamic, and largely implicit. Thus, an article using a resources model to frame the study should receive codes of Variable utility, Unstable, and Implicit (Fig. 3.9). Three of the seven studies that reported a resources model were assigned these codes. Of the remaining four studies, two were coded as *Ambiguous*, *Unstable*, and *Explicit*; one was coded as *Ambiguous*, *Unstable*, and *Both*; and one was coded as *Hierarchical*, *Unstable*, and *Explicit*. Four studies that did not report using a model of epistemic cognition received the codes aligned with a resources model (*i.e.*, *Variable utility*, *Unstable*, *Implicit)*. An additional three studies were coded as *Ambiguous*, *Unstable*, and *Implicit*, mostly aligning with a resources model. In total, seven studies aligned completely with a resources model and three studies mostly aligned (Fig. 3.10).

The remaining 26 studies received coding combinations that did not clearly align with any model of epistemic cognition. Fifteen of these studies were coded as *Hierarchical* and *Unclear* with regards to stability. These were split approximately evenly between *Explicit* and *Implicit* codes. Six studies were coded as *Hierarchical* and *Unstable*. The remaining studies were either coded as *Ambiguous* and *Unclear* (N = 1), *Ambiguous* and *Stable* (N = 1), or *Ambiguous* and

Unstable ($N = 3$). In total, a lack of evidence prevents us from inferring a model of cognition for nearly half of studies on student epistemologies in undergraduate chemistry education research published between 2000 and 2022.

Discussion and Implications

Some of the findings reported above are unsurprising. For example, our field tends to study the epistemic cognition of individual students enrolled in introductory chemistry classes within the confines of an academic semester. We suspect this tendency reflects some combination of access to student populations, project timelines, and educational traditions that focus on individual (vs. community) learning. However, just because these sorts of studies have been done in the past does not mean that they represent the only, or best, way to approach exploring epistemic cognition. One might persuasively argue for the importance of longitudinal studies of community epistemic cognition – after all, people reason about scientific questions as members of their social and cultural groups and with other members of those groups across many contexts (Feinstein & Waddington, 2020). Such studies would of course require sustained funding and diverse expertise (*e.g.*, science education, science and technology studies).

Regardless of the sample and duration of epistemology-focused studies, researchers will be faced with the choice of collecting data that requires conscious articulation/selection of epistemic ideas or observing behavior to infer the epistemic cognition underlying that behavior. The roughly equal distribution of *Explicit* and *Implicit* codes in our data suggests that researchers see advantages and disadvantages to both approaches. Explicit survey or interview questions provide information about the participants' perceptions of their own epistemic cognition. Data analysis is also relatively straightforward; it requires little interpretation of participants' responses. This allows researchers to collect data on large numbers of students and perform statistical analyses.

Sandoval & Çam (2011) argue, however, that students may experience the context of an interview or survey as substantially different from the context of behaviors researchers are interested in. As such, ideas about knowing and learning activated when responding to a survey or interview question may not map onto epistemologies that underlie behavior in class or in life. This makes it difficult to make reasonable claims about student behavior or suggest classroom interventions based on survey response data alone.

Implicit measurements inherently require the researcher to make more inferences. Rarely does a student say something like, "My epistemic aim at this moment is to avoid obtaining false beliefs." Instead, they might say something like, "I don't think that is right. Let's check with the professor," from which we can perhaps infer an epistemic aim of avoiding false beliefs and the professor as a source of knowledge. As a result, data analysis is more complicated and time-intensive for data collected via implicit measurements. This makes study designs employing these approaches to data collection and analysis less practical for large sample sizes and more difficult for a practitioner to use to evaluate their classes. However, they can offer a more nuanced, context-sensitive picture of students' epistemologies than explicit measurements because they can capture epistemology in use, *i.e.*, "practical epistemologies" (Sandoval, 2005; Berland et al., 2016).

As with the Explicitness codes, we saw an approximately even distribution of Stability codes. This category of codes was challenging to apply given that few researchers discussed their assumptions of stability or instability, resulting in a large portion of studies coded as *Unclear*. We argue, however, that this assumption influences the design of the study and the interpretation of data. Administering pre- and post-tests before and after an intervention, for example, would be reasonable if one assumes students' responses indicate relatively stable epistemic ideas that were expected to persist in the absence of the intervention. But if one assumes students' responses indicate epistemic ideas invoked in the moment, which may or may not be deeply held, then one would be cautious about attributing any changes in responses to the intervention. Thus, a major takeaway of this review is that future researchers should address assumptions of stability with regard to epistemic cognition.

The most striking finding, we claim, was that more than 70% of studies performed a hierarchical analysis of students' epistemic cognition data. We hypothesize that *Hierarchical* studies are so prevalent because creating hierarchies seems intuitive and the results of hierarchical analyses lend themselves to relatively straightforward interpretations. By placing students on a continuum from "naïve" to "expert" epistemic cognition, we can judge how/whether an intervention was successful in supporting hoped-for improvements. We have two major objections to employing context-invariant hierarchies: (1) The assumption that one set of epistemic ideas is best in all circumstances is not reasonable, and (2) assembled hierarchies nearly always position an idealized vision of White, Western norms as most sophisticated and de-value or ignore other powerful and legitimate ways of knowing and learning.

Overemphasis on students advancing toward and achieving the "best" epistemologies may overlook the ways in which other epistemologies could prove useful and act to marginalize whole groups of students. In some articles, students were considered naïve for thinking that there would be a single correct answer. A quick reflection on how we use knowledge, both in chemistry and everyday life, should reveal why equating binary thinking with epistemic immaturity is overly simplistic. Sometimes it is useful to adopt a binary perspective, such as when assessing if you made the desired pharmaceutical compound or its toxic enantiomer. The danger of an inflexible, hierarchical view is that descriptors or measurements of epistemic cognition may be interpreted as value-laden traits of the students themselves, creating difference among groups of students that can be used to justify inequitable treatments of those groups (Kirchgasler, 2017). One might, for example, make claims that students who are identified as “less sophisticated” dualistic thinkers are less capable of engaging in chemistry courses than students who are identified as “more sophisticated” relativist thinkers. This could be used to justify separate tracks whereby “less sophisticated” students are assigned to a “lower level” course. Thus, a rigidly hierarchical viewpoint of epistemology may in fact lead to educational policies that restrict who is allowed to continue studying science.

Furthermore, by defining a universal best epistemology, almost always based on White, Western norms, we ignore or devalue powerful and legitimate systems of knowing that exist in other cultures (Ladson-Billings, 2000). As a consequence, Bang and Medin (2010) assert, “In education, most epistemology research makes the assumption that the epistemologies students come to classrooms with are inferior, or less productive, compared with the one(s) that research and educators (for our purposes, science education) are trying to assist students in learning.” They go on to discuss how such a perspective devalues ways of knowing that Native American communities possess. The prevalent use of hierarchical coding schemes in chemistry education research is consistent with their assertion about science education research generally. Such a view discounts the productive resources that all students possess, but especially those who are already marginalized by our education systems and thus under-represented in science. Adopting a perspective that values multiple ways of constructing and evaluating knowledge is one way that chemistry educators can work toward creating more equitable learning environments.

Saying that “all epistemologies may have utility in some context” should not be read as implying “all epistemologies are equally useful in all contexts”. Most scholars who ascribe to a resources view of epistemic cognition acknowledge the existence of situated epistemological hierarchies, in which some ways of knowing and learning may be particularly useful in advancing toward a particular aim in a given moment. This means that "evaluations of epistemological sophistication must account for the appropriateness and utility of epistemological resources being used in the current context" (Berland and Crucet, 2016, p. 10; Elby and Hammer, 2001; Hammer & Elby, 2002). Theoretical and analytic work of this sort is far from straightforward. One must grapple with questions such as: How should we make arguments that epistemologies are more/less useful without an *a priori* set of "best epistemologies"? How might we define a "context" for the purposes of this sort of analysis, given that epistemologies can shift over a very short time scale? How should we think about the interplay between instructors' epistemic learning goals, the design of a learning environment, and ways of knowing and learning students experience as valuable in that environment? Conversations around questions such as these are ongoing in the science education community (*e.g.*, Pierson et al., 2023; Warren et al., 2020), but still fairly rare in the context of college chemistry learning. We are hopeful this article serves to spark more conversations around how and why we define "epistemological sophistication" in the ways that we do.

Limitations

In searching for and selecting articles to include in this review, it was necessary to make decisions to restrict the scope. The search terms "epistem*," "nature of science," and "beliefs" were used to find articles, but articles employing other terms to describe students' thinking about chemistry knowledge may have been missed. Searches were performed in prominent chemistry education and science education journals, as well as a few databases, but nevertheless, some relevant articles may have eluded us. Finally, we chose to restrict our analysis to articles involving students in undergraduate chemistry courses. We do not know if studies on other populations, such as secondary school students, secondary school teachers, graduate students, or college instructors, would exhibit the same trends.

Like all coding schemes, the coding scheme presented in Table 3.3 is a tool to summarize and interpret the data, albeit at the cost of some resolution. With only three codes per category, much of the variation is obscured. This was especially true for the coding category Explicitness. Among articles coded as *Implicit*, the degree of inference required ranges depending on the exact nature of the data collection methods. For example, an interview asking students to reflect on their experiences in a course is not the same as observing them as they solve problems in a small group.

Furthermore, in applying our coding scheme, we needed to make decisions based only on what was published in the articles. At the time of this writing, the field has not established agreed-upon guidelines regarding what information should be including in publications on students' epistemic cognition. For example, assumptions about the stability of epistemic ideas are not usually stated, resulting in a large number of articles coded as *Unclear*. Thus, much of our coding relied on inferences drawn from the theoretical framework invoked, the data collection and analysis methods used, and the conclusions drawn. It is possible our interpretations do not match the authors' intended meanings.

In developing our coding scheme, we chose to attend to what we perceive as some of the important assumptions embedded in the various theoretical models of epistemic cognition. Other assumptions were not operationalized in our coding scheme. An example is the extent to which epistemic cognition is domain general or domain specific. (These ideas were incorporated less rigorously into our discussion of stability.)

Conclusions and Future Directions

The model of epistemic cognition researchers employ informs all aspects of a study, from the research questions that can be asked to the implications and conclusions that can be drawn. We reviewed articles on students' epistemic cognition in undergraduate chemistry courses and found that very few articles described their model of epistemic cognition, so instead we looked for distinguishing characteristics of developmental or dimensional models versus resources models (*i.e.*, assumptions about hierarchy, stability, and explicitness). From this analysis, we were able to tentatively infer that one third of studies were informed by a developmental or dimensional model, one fifth were informed by a resources model, and the remaining half remained too ambiguous to infer a model.

Developmental and dimensional models have played an important role in enabling research and discussion regarding the epistemic aspect of students' education. We argue that a resources model takes into account many of the ideas put forward in these models (*e.g.*, kinds of epistemic knowledge) but incorporates them into a more modern understanding of the dynamic and highly context-dependent nature of cognition (diSessa, 1988; Hammer & Elby, 2005). Furthermore, not only has a resources model been shown to better account for data (Hammer & Elby, 2002), but it does not require the researcher to impose a Eurocentric value system when analyzing the data. Rather, a resources model allows researchers to treat ways of knowing from marginalized communities as valid and valuable. But employing a resources model brings its own set of theoretical questions and methodological challenges. How do we collect and analyze data on large samples in a nuanced way? How does individual resource activation influence collective ideas on knowledge construction and evaluation and vice versa? Who decides (and who should decide) which epistemic resources are useful in a particular context and to whom? These are some of the questions we hope future work on epistemic cognition will address.

Notes and References

1) King and Kitchener (2004) caution that participants' views can fall into multiple categories, and that each stage is best understood as where the majority of their views lie.
2) Note some scholars question whether the last two dimensions are actually epistemic.

Ağlarcı, O., Sarıçayır, H., & Şahin, M. (2016). Nature of science instruction to Turkish prospective chemistry teachers: The effect of explicit-reflective approach. *Cogent Education*, *3*(1), 1213350. https://doi.org/10.1080/2331186X.2016.1213350

Aguirre-Mendez, C., Chen, Y.-C., Terada, T., & Techawitthayachinda, R. (2020). Predicting components of argumentative writing and achievement gains in a general chemistry course for nonmajor college students. *Journal of Chemical Education*, *97*(8), 2045–2056. https://doi.org/10.1021/acs.jchemed.0c00042

Agustian, H. Y. (2020). Students' understanding of the nature of science in the context of an undergraduate chemistry laboratory. *Electronic Journal for Research in Science & Mathematics Education*, *24*(2), 56–85.

American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: A call to action*.

Bain, K., Moon, A., Mack, M. R., & Towns, M. H. (2014). A review of research on the teaching and learning of thermodynamics at the university level. *Chemistry Education Research and Practice*, *15*(3), 320–335. https://doi.org/10.1039/C4RP00011K

Bang, M., & Medin, D. (2010). Cultural processes in science education: Supporting the navigation of multiple epistemologies. *Science Education*, *94*(6), 1008–1026. https://doi.org/10.1002/sce.20392

Becker, N., Rasmussen, C., Sweeney, G., Wawro, M., Towns, M., & Cole, R. (2013). Reasoning using particulate nature of matter: An example of a sociochemical norm in a university-level physical chemistry class. *Chemistry Education Research and Practice*, *14*(1), 81–94. https://doi.org/10.1039/C2RP20085F

Berland, L. K., & Crucet, K. (2016). Epistemological trade-offs: Accounting for context when evaluating epistemological sophistication of student engagement in scientific practices. *Science Education*, *100*(1), 5–29. https://doi.org/10.1002/sce.21196

Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. J. (2016). Epistemologies in practice: Making scientific practices meaningful for students. *Journal of Research in Science Teaching*, *53*(7), 1082–1112. https://doi.org/10.1002/tea.21257

Bing, T. J., & Redish, E. F. (2009). Analyzing problem solving using math in physics: Epistemological framing via warrants. *Physical Review Special Topics - Physics Education Research*, *5*(2), 020108. https://doi.org/10.1103/PhysRevSTPER.5.020108

Bowen, R. S., Flaherty, A. A., & Cooper, M. M. (2022). Investigating student perceptions of transformational intent and classroom culture in organic chemistry courses. *Chemistry Education Research and Practice*, *23*(3), 560–581. https://doi.org/10.1039/D2RP00010E

Buehl, M. M., & Alexander, P. A. (2001). Beliefs about academic knowledge. *Educational Psychology Review*, *13*(4), 385–418. https://doi.org/10.1023/A:1011917914756

Çalik, M., & Cobern, W. W. (2017). A cross-cultural study of CKCM efficacy in an undergraduate chemistry classroom. *Chemistry Education Research and Practice*, *18*(4), 691–709. https://doi.org/10.1039/C7RP00016B

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin Company.

Çelik, S. (2020). Changes in nature of science understandings of preservice chemistry teachers in an explicit, reflective, and contextual nature of science teaching. *International Journal of Research in Education and Science*, *6*(2), 315–326. https://doi.org/10.46328/ijres.v6i2.892

Charters, E. (2003). The use of think aloud methods in qualitative research an introduction to think aloud methods. *Brock Education Journal*, *12*(2), 68–82.

Chinn, C. A., Buckland, L. A., & Samarapungavan, A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist*, *46*(3), 141–167. https://doi.org/10.1080/00461520.2011.587722

Chopra, I., O'Connor, J., Pancho, R., Chrzanowski, M., & Sandi-Urena, S. (2017). Reform in a general chemistry laboratory: How do students experience change in the instructional approach? *Chemistry Education Research and Practice*, *18*(1), 113–126. https://doi.org/10.1039/C6RP00082G

Collins, A., & Ferguson, W. (1993). Epistemic forms and epistemic games: Structures and strategies to guide inquiry. *Educational Psychologist*, *28*(1), 25–42. https://doi.org/10.1207/s15326985ep2801_3

Cooper, M., & Klymkowsky, M. (2013). Chemistry, life, the universe, and everything: A new approach to general chemistry, and a model for curriculum reform. *Journal of Chemical Education*, *90*(9), 1116–1122. https://doi.org/10.1021/ed300456y

Crujeiras-Pérez, B., & Brocos, P. (2021). Pre-service teachers' use of epistemic criteria in the assessment of scientific procedures for identifying microplastics in beach sand. *Chemistry Education Research and Practice*, *22*(2), 237–246. https://doi.org/10.1039/D0RP00176G

Dempsey, N. P. (2010). Stimulated recall interviews in ethnography. *Qualitative Sociology*, *33*(3), 349–367. https://doi.org/10.1007/s11133-010-9157-x

diSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. B. Pufall (Eds.), *Constructivism in the computer age* (pp. 49–70). Lawrence Erlbaum Associates, Inc.

Elby, A., & Hammer, D. (2001). On the substance of a sophisticated epistemology. *Science Education*, *85*(5), 554–567. https://doi.org/10.1002/sce.1023

Entwistle, N. J. (1991). Approaches to learning and perceptions of the learning environment. *Higher Education*, *22*(3), 201–204. https://doi.org/10.1007/BF00132287

Feinstein, N. W., & Waddington, D. I. (2020). Individual truth judgments or purposeful, collective sensemaking? Rethinking science education's response to the post-truth era. *Educational Psychologist*, *55*(3), 155–166. https://doi.org/10.1080/00461520.2020.1780130

Flaherty, A. A. (2020). A review of affective chemistry education research and its implications for future research. *Chemistry Education Research and Practice*, *21*(3), 698–713. https://doi.org/10.1039/C9RP00200F

Ginsburg, H. P. (1997). *Entering the child's mind: The clinical interview in psychological research and practice*. Cambridge University Press. https://doi-org/10.1017/CBO9780511527777

Greene, J. A., Azevedo, R., & Torney-Purta, J. (2008). Modeling epistemic and ontological cognition: Philosophical perspectives and methodological directions. *Educational Psychologist*, *43*(3), 142–160. https://doi.org/10.1080/00461520802178458

Grooms, J. (2020). A comparison of argument quality and students' conceptions of data and evidence for undergraduates experiencing two types of laboratory instruction. *Journal of Chemical Education*, *97*(8), 2057–2064. https://doi.org/10.1021/acs.jchemed.0c00026

Grove, N. P., & Bretz, S. L. (2010). Perry's Scheme of Intellectual and Epistemological Development as a framework for describing student difficulties in learning organic chemistry. *Chemistry Education Research and Practice*, *11*(3), 207–211. https://doi.org/10.1039/C005469K

Grove, N. P., & Bretz, S. L. (2012). A continuum of learning: From rote memorization to meaningful learning in organic chemistry. *Chemistry Education Research and Practice*, *13*(3), 201–208. https://doi.org/10.1039/C1RP90069B

Hammer, D., & Elby, A. (2002). On the form of a personal epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 169–190).

Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 89–119).

Havdala, R., & Ashkenazi, G. (2007). Coordination of theory and evidence: Effect of epistemological theories on students' laboratory practice. *Journal of Research in Science Teaching*, *44*(8), 1134–1159.

Hofer, B. K. (2004). Exploring the dimensions of personal epistemology in differing classroom contexts: Student interpretations during the first year of college. *Contemporary Educational Psychology*, *29*(2), 129–163. https://doi.org/10.1016/j.cedpsych.2004.01.002

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, *67*(1), 88–140. https://doi.org/10.3102/00346543067001088

Hosbein, K. N., Lower, M. A., & 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. *Journal of Chemical Education*, *98*(6), 1875–1887. https://doi.org/10.1021/acs.jchemed.0c01225

Hunter, K. H., Rodriguez, J.-M. G., & Becker, N. M. (2022). A review of research on the teaching and learning of chemical bonding. *Journal of Chemical Education*. https://doi.org/10.1021/acs.jchemed.2c00034

Johnson, R. B., & Christensen, L. (2020). *Educational research: Quantitative, qualitative, and mixed approaches* (7th ed.). SAGE Publications, Inc.

Kelly, R. M., Akaygun, S., Hansen, S. J. R., Villalta-Cerdas, A., & Adam, J. (2021). Examining learning of atomic level ideas about precipitation reactions with a resources framework. *Chemistry Education Research and Practice*, *22*, 886–904. https://doi.org/10.1039/D0RP00071J

King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgement: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. Jossey-Bass.

King, P. M., & Kitchener, K. S. (2004). Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist*, *39*(1), 5–18. https://doi.org/10.1207/s15326985ep3901_2

Kirchgasler, K. L. (2017). Scientific Americans: Historicizing the making of difference in early 20th century U.S. science education. In T. A. Popkewitz, J. Diaz, & C. Kirchgasler (Eds.), *A political sociology of educational knowledge* (1st ed., pp. 87–102). Routledge.

Kitchener, K. S. (1983). Cognition, metacognition, and epistemic cognition: A three-level model of cognitive processing. *Human Development*, *26*(4), 222–232. https://doi.org/10.1159/000272885

Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher*, *28*(2), 16–46. https://doi.org/10.3102/0013189X028002016

Ladson-Billings, G. (2000). Racialized discourses and ethnic epistemologies. In *Handbook of qualitative research* (2nd ed., pp. 257–277). SAGE Publications, Inc.

Lazenby, K., Stricker, A., Brandriet, A., Rupp, C. A., & Becker, N. M. (2020). Undergraduate chemistry students' epistemic criteria for scientific models. *Journal of Chemical Education*, *97*(1), 16–26. https://doi.org/10.1021/acs.jchemed.9b00505

Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching*, *29*(4), 331–359. https://doi.org/10.1002/tea.3660290404

Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. S. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching*, *39*(6), 497–521. https://doi.org/10.1002/tea.10034

Lee, J., Park, T., & Davis, R. O. (2022). What affects learner engagement in flipped learning and what predicts its outcomes? *British Journal of Educational Technology*, *53*(2), 211–228. https://doi.org/10.1111/bjet.12717

Li, W.-T., Liang, J.-C., & Tsai, C.-C. (2013). Relational analysis of college chemistry-major students' conceptions of and approaches to learning chemistry. *Chemistry Education Research and Practice*, *14*(4), 555–565. https://doi.org/10.1039/C3RP00034F

Lising, L., & Elby, A. (2005). The impact of epistemology on learning: A case study from introductory physics. *American Journal of Physics*, *73*(4), 372. https://doi.org/10.1119/1.1848115

Louca, L., Elby, A., Hammer, D., & Kagey, T. (2004). Epistemological resources: Applying a new epistemological framework to science instruction. *Educational Psychologist*, *39*(1), 57–68. https://doi.org/10.1207/s15326985ep3901_6

Mazzarone, K. M., & Grove, N. P. (2013). Understanding epistemological development in first- and second-year chemistry students. *Journal of Chemical Education*, *90*(8), 968–975. https://doi.org/10.1021/ed300655s

McDonald, C. V. (2010). The influence of explicit nature of science and argumentation instruction on preservice primary teachers' views of nature of science. *Journal of Research in Science Teaching*, *47*(9), 1137–1164. https://doi.org/10.1002/tea.20377

Momsen, J., Offerdahl, E., Kryjevskaia, M., Montplaisir, L., Anderson, E., & Grosz, N. (2013). Using assessments to investigate and compare the nature of learning in undergraduate science courses. *CBE—Life Sciences Education*, *12*(2), 239–249. https://doi.org/10.1187/cbe.12-08-0130

Moon, A., Moeller, R., Gere, A. R., & Shultz, G. V. (2019). Application and testing of a framework for characterizing the quality of scientific reasoning in chemistry students' writing on ocean acidification. *Chemistry Education Research and Practice*, *20*(3), 484–494. https://doi.org/10.1039/C9RP00005D

Muis, K. R., Bendixen, L. D., & Haerle, F. C. (2006). Domain-generality and domain-specificity in personal epistemology research: Philosophical and empirical reflections in the development of a theoretical framework. *Educational Psychology Review*, *18*(1), 3–54. https://doi.org/10.1007/s10648-006-9003-6

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas* (p. 13165). National Academies Press. https://doi.org/10.17226/13165

Noble, H., & Smith, J. (2015). Issues of validity and reliability in qualitative research. *Evidence Based Nursing*, *18*(2), 34–35. https://doi.org/10.1136/eb-2015-102054

Perry, W. J. Jr. (1970). *Forms of ethical and intellectual development in the college years: A scheme*. Holt, Rinehart and Winston.

Petritis, S. J., Kelley, C., & Talanquer, V. (2021). Exploring the impact of the framing of a laboratory experiment on the nature of student argumentation. *Chemistry Education Research and Practice*, *22*(1), 105–121. https://doi.org/10.1039/D0RP00268B

Pierson, A. E., Brady, C. E., Clark, D. B., & Sengupta, P. (2023). Students' epistemic commitments in a heterogeneity-seeking modeling curriculum. *Cognition and Instruction*, *41*(2), 125–157.

Rodriguez, J.-M. G., Bain, K., & Towns, M. H. (2020). The role of epistemology and epistemic games in mediating the use of mathematics in chemistry: Implications for mathematics instruction and research on undergraduate mathematics education. *International Journal of Research in Undergraduate Mathematics Education*, *6*(2), 279–301. https://doi.org/10.1007/s40753-019-00110-8

Rodriguez, J.-M. G., Hunter, K. H., Scharlott, L. J., & Becker, N. M. (2020). A review of research on process oriented guided inquiry learning: Implications for research and practice. *Journal of Chemical Education*. https://doi.org/10.1021/acs.jchemed.0c00355

Rolfe, G. (2006). Validity, trustworthiness and rigour: Quality and the idea of qualitative research. *Journal of Advanced Nursing*, *53*(3), 304–310. https://doi.org/10.1111/j.1365-2648.2006.03727.x

Rosenberg, S., Hammer, D., & Phelan, J. (2006). Multiple epistemological coherences in an eighth-grade discussion of the rock cycle. *The Journal of the Learning Sciences*, *15*(2), 261–292.

Russ, R. S. (2018). Characterizing teacher attention to student thinking: A role for epistemological messages. *Journal of Research in Science Teaching*, *55*(1), 94–120. https://doi.org/10.1002/tea.21414

Russ, R. S., Lee, V. R., & Sherin, B. L. (2012). Framing in cognitive clinical interviews about intuitive science knowledge: Dynamic student understandings of the discourse interaction. *Science Education*, *96*(4), 573–599. https://doi.org/10.1002/sce.21014

Russell, C. B., & Weaver, G. C. (2011). A comparative study of traditional, inquiry-based, and research-based laboratory curricula: Impacts on understanding of the nature of science. *Chemistry Education Research and Practice*, *12*(1), 57–67. https://doi.org/10.1039/C1RP90008K

Sandoval, W. A. (2005). Understanding students’ practical epistemologies and their influence on learning through inquiry. *Science Education*, *89*, 634–656.

Sandoval, W. A., & Çam, A. (2011). Elementary children’s judgments of the epistemic status of sources of justification. *Science Education*, *95*(3), 383–408. https://doi.org/10.1002/sce.20426

Sandoval, W. A., Greene, J. A., & Bråten, I. (2016). Understanding and promoting thinking about knowledge: Origins, issues, and future directions of research on epistemic cognition. *Review of Research in Education*, *40*, 457–496.

Saribas, D., Mugaloglu, E. Z., & Bayram, H. (2013). Creating metacognitive awareness in the lab: Outcomes for preservice science teachers. *Eurasia Journal of Mathematics, Science and Technology Education*, *9*(1), 83–88. https://doi.org/10.12973/eurasia.2013.918a

Schmitt, F. (2017). Social Epistemology. In J. Greco & E. Sosa (Eds.), *The Blackwell guide to epistemology* (pp. 354–382). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781405164863.ch15

Schommer-Aikins, M. (2004). Explaining the epistemological belief system: Introducing the embedded systemic model and coordinated research approach. *Educational Psychologist*, *39*(1), 19–29. https://doi.org/10.1207/s15326985ep3901_3

Scouller, K. M. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, *35*(4), 453–472. https://doi.org/10.1023/A:1003196224280

Scouller, K. M., & Prosser, M. (1994). Students' experiences in studying for multiple choice question examinations. *Studies in Higher Education*, *19*(3), 267–279. https://doi.org/10.1080/03075079412331381870

Sendur, G., Polat, M., & Kazancı, C. (2017). Does a course on the history and philosophy of chemistry have any effect on prospective chemistry teachers' perceptions? The case of chemistry and the chemist. *Chemistry Education Research and Practice*, *18*(4), 601–629. https://doi.org/10.1039/C7RP00054E

Sevian, H., & Couture, S. (2018). Epistemic games in substance characterization. *Chemistry Education Research and Practice*, *19*(4), 1029–1054. https://doi.org/10.1039/C8RP00047F

Shultz, G. V., & Gere, A. R. (2015). Writing-to-learn the Nature of Science in the context of the Lewis dot structure model. *Journal of Chemical Education*, *92*(8), 1325–1329. https://doi.org/10.1021/acs.jchemed.5b00064

Snyder, B. (1973). *The hidden curriculum*. The MIT Press: Cambridge, MA.

Talanquer, V. (2010). Exploring dominant types of explanations built by general chemistry students. *International Journal of Science Education*, *32*(18), 2393–2412. https://doi.org/10.1080/09500690903369662

Talanquer, V., Bolger, M., & Tomanek, D. (2015). Exploring prospective teachers' assessment practices: Noticing and interpreting student understanding in the assessment of written work. *Journal of Research in Science Teaching*, *52*(5), 585–609. https://doi.org/10.1002/tea.21209

Talanquer, V., & Pollard, J. (2017). Reforming a large foundational course: Successes and challenges. *Journal of Chemical Education*, *94*(12), 1844–1851. https://doi.org/10.1021/acs.jchemed.7b00397

Tiruneh, D. T., Verburgh, A., & Elen, J. (2014). Effectiveness of critical thinking instruction in higher education: A systematic review of intervention studies. *Higher Education Studies*, *4*(1), p1. https://doi.org/10.5539/hes.v4n1p1

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

Venessa, D. M., Hernani, H., & Halimatul, H. S. (2019). Exploring view of nature of science and technology pre-service chemistry teachers. *Journal of Science Learning*, *3*(1), 19–28. https://doi.org/10.17509/jsl.v3i1.17757

Walker, J. P., & 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. *Journal of Research in Science Teaching*, *50*(5), 561–596. https://doi.org/10.1002/tea.21082

Walker, J. P., Van Duzor, A. G., & Lower, M. A. (2019). Facilitating argumentation in the laboratory: The challenges of claim change and justification by theory. *Journal of Chemical Education*, *96*(3), 435–444. https://doi.org/10.1021/acs.jchemed.8b00745

Warren, B., Vossoughi, S., Rosebery, A. S., Bang, M., & Taylor, E. V. (2020). Multiple ways of knowing*: Re-imagining disciplinary learning. In *Handbook of the cultural foundations of learning* (pp. 277-294). Routledge.

White, R. T., & Arzi, H. J. (2005). Longitudinal studies: Designs, validity, practicality, and value. *Research in Science Education*, *35*(1), 137–149. https://doi.org/10.1007/s11165-004-3437-y

Wickman, P.-O. (2004). The practical epistemologies of the classroom: A study of laboratory work. *Science Education*, *88*(3), 325–344. https://doi.org/10.1002/sce.10129

Supporting Information

The supporting information contains details on the articles included in the sample and the codes assigned to them during the analysis phase. Table 3.S1 lists the articles in alphabetical order by last name of first author and provides the author names, publication year, title, journal, volume, issue, page numbers, and DOI where possible. Tables 3.S2, 3.S3, and 3.S4 provide details on the analysis. The study population, unit of analysis, methodology, and study design codes assigned to each article are recorded in Table 3.S2. Table 3.S3 lists the data sources and whether each explicitly or implicitly prompted for students' epistemic ideas. Table 3.S4 contains the codes from part two of the analysis: hierarchy, stability, explicitness, inferred model of epistemic cognition, and stated model of epistemic cognition (if applicable).

Table 1.S1. Bibliographic information for the articles selected for the review, arranged alphabetically by last name of first author.

Abbreviation	Authors	Year	Title	Journal	Volume	Issue	Page Numbers	DOI
Adams et al., 2008	Adams, W. K., Wieman, C. E., Perkins, K. K., & Barbera, J.	2008	Modifying and validating the Colorado Learning Attitudes about Science Survey for use in chemistry	*Journal of Chemical Education*	85	10	1435-1439	https://doi.org/10.1021/ed085p1435
Aglarci et al., 2016	Ağlarcı, O., Sarıçayır, H., & Şahin, M.	2016	Nature of science instruction to Turkish prospective chemistry teachers: The effect of explicit-reflective approach	*Cogent Education*	3	1	1213350	https://doi.org/10.1080/2331186X.2016.1213350
Aguirre-Mendez et al., 2020	Aguirre-Mendez, C., Chen, Y.-C., Terada, T., & Techawitthayachinda, R.	2020	Predicting components of argumentative writing and achievement gains in a general chemistry course for nonmajor college students	*Journal of Chemical Education*	97	8	2045–2056	https://doi.org/10.1021/acs.jchemed.0c00042
Agustian, 2020	Agustian, H. Y.	2020	Students' understanding of the nature of science in the context of an undergraduate chemistry laboratory	*Electronic Journal for Research in Science & Mathematics Education*	24	2	56–85	
Becker et al., 2013	Becker, N., Rasmussen, C., Sweeney, G., Wawro, M., Towns, M., & Cole, R.	2013	Reasoning using particulate nature of matter: An example of a sociochemical norm in a university-level physical chemistry class	*Chemistry Education Research and Practice*	14	1	81–94	https://doi.org/10.1039/C2RP20085F
Bowen et al., 2022	Bowen, R. S., Flaherty, A. A., & Cooper, M. M.	2022	Investigating student perceptions of transformational intent and classroom culture in organic chemistry courses	*Chemistry Education Research and Practice*	23	3	560–581	https://doi.org/10.1039/D2RP00010E

Table 1.S1. (Continued) Bibliographic information for the articles selected for the review, arranged alphabetically by last name of first author.

Abbreviation	Authors	Year	Title	Journal	Volume	Issue	Page Numbers	DOI
Burrows et al., 2017	Burrows, N. L., Nowak, M. K., & Mooring, S. R.	2017	Students' perceptions of a project-based organic chemistry laboratory environment: A phenomenographic approach	*Chemistry Education Research and Practice*	18	4	811–824	https://doi.org/10.1039/C7RP00064B
Calik & Cobern, 2017	Çalik, M., & Cobern, W. W.	2017	A cross-cultural study of CKCM efficacy in an undergraduate chemistry classroom	*Chemistry Education Research and Practice*	18	4	691–709	https://doi.org/10.1039/C7RP00016B
Celik, 2020	Celik, S.	2020	Changes in nature of science understandings of preservice chemistry teachers in an explicit, reflective, and contextual nature of science teaching	*International Journal of Research in Education and Science*	6	2	315–326	https://doi.org/10.46328/ijres.v6i2.892
Chopra et al., 2017	Chopra, I., O' Connor, J., Pancho, R., Chrzanowski, M., & Sandi-Urena, S.	2017	Reform in a general chemistry laboratory: How do students experience change in the instructional approach?	*Chemistry Education Research and Practice*	18	1	113–126	https://doi.org/10.1039/C6RP00082G
Christian & Talanquer, 2012	Christian, K., & Talanquer, V.	2012	Content-related interactions in self-initiated study groups	*International Journal of Science Education*	34	14	2231–2255	https://doi.org/10.1080/09500693.2012.708064
Cigdemoglu et al., 2017	Cigdemoglu, C., Arslan, H. O., & Cam, A.	2017	Argumentation to foster pre-service science teachers' knowledge, competency, and attitude on the domains of chemical literacy of acids and bases	*Chemistry Education Research and Practice*	18	2	288–303	https://doi.org/10.1039/C6RP00167J

Table 1.S1. (Continued) Bibliographic information for the articles selected for the review, arranged alphabetically by last name of first author.

Abbreviation	Authors	Year	Title	Journal	Volume	Issue	Page Numbers	DOI
Crujeiras-Perez & Brocos, 2021	Crujeiras-Pérez, B., & Brocos, P.	2021	Pre-service teachers' use of epistemic criteria in the assessment of scientific procedures for identifying microplastics in beach sand	*Chemistry Education Research and Practice*	22	2	237–246	https://doi.org/10.1039/D0RP00176G
Flaherty, 2020	Flaherty, A. A.	2020	Investigating perceptions of the structure and development of scientific knowledge in the context of a transformed organic chemistry lecture course	*Chemistry Education Research and Practice*	21	2	570–581	https://doi.org/10.1039/C9RP00201D
Garcia-Carmona et al., 2017	García-Carmona, A., & Acevedo-Díaz, J. A.	2017	Understanding the nature of science through a critical and reflective analysis of the controversy between Pasteur and Liebig on fermentation	*Science & Education*	26	1	65–91	https://doi.org/10.1007/s11191-017-9876-4
Gardner, 2017	Gardner, M. A.	2017	Developing connections between integrated laboratory practices and first-year undergraduate Nature of Science (NOS) understanding	*International Journal of Environmental and Science Education*	12	5	1299–1315	
Goff et al., 2012	Goff, P., Boesdorfer, S. B., & Hunter, W.	2012	Using a multicultural approach to teach chemistry and the nature of science to undergraduate non-majors	*Cultural Studies of Science Education*	7	3	631–651	https://doi.org/10.1007/s11422-012-9382-6
Grooms, 2020	Grooms, J.	2020	A comparison of argument quality and students' conceptions of data and evidence for undergraduates experiencing two types of laboratory instruction	*Journal of Chemical Education*	97	8	2057–2064	https://doi.org/10.1021/acs.jchemed.0c00026

Table 1.S1. (Continued) Bibliographic information for the articles selected for the review, arranged alphabetically by last name of first author.

Abbreviation	Authors	Year	Title	Journal	Volume	Issue	Page Numbers	DOI
Grove & Bretz, 2012	Grove, N. P., & Bretz, S. L.	2010	Perry's Scheme of Intellectual and Epistemological Development as a framework for describing student difficulties in learning	*Chemistry Education Research and Practice*	11	3	207–211	https://doi.org/10.1039/C005469K
Grove & Bretz, 2010	Grove, N. P., & Bretz, S. L.	2012	A continuum of learning: From rote memorization to meaningful learning in organic chemistry	*Chemistry Education Research and Practice*	13	3	201–208	https://doi.org/10.1039/C1RP90069B
Havdala & Ashkenazi, 2007	Havdala, R., & Ashkenazi, G.	2007	Coordination of theory and evidence: Effect of epistemological theories on students' laboratory practice	*Journal of Research in Science Teaching*	44	8	1134–1159	https://doi.org/10.1002/tea.20215
Hofer, 2004	Hofer, B. K.	2004	Exploring the dimensions of personal epistemology in differing classroom contexts: Student interpretations during the first year of college	*Contemporary Educational Psychology*	29	2	129–163	https://doi.org/10.1016/j.cedpsych.2004.01.002
Hosbein et al., 2021	Hosbein, K. N., Lower, M. A., & Walker, J. P.	2021	Tracking student argumentation skills across general chemistry through Argument-Driven Inquiry using the Assessment of Scientific Argumentation in	*Journal of Chemical Education*	98	6	1875–1887	https://doi.org/10.1021/acs.jchemed.0c01225
Hunter et al., 2021	Hunter, K. H., Rodriguez, J.-M. G., & Becker, N. M.	2021	Making sense of sensemaking: Using the sensemaking epistemic game to investigate student discourse during a collaborative gas law activity	*Chemistry Education Research and Practice*	22	2	328–346	https://doi.org/10.1039/D0RP00290A

Table 1.S1. (Continued) Bibliographic information for the articles selected for the review, arranged alphabetically by last name of first author.

Abbreviation	Authors	Year	Title	Journal	Volume	Issue	Page Numbers	DOI
Karch & Sevian, 2022	Karch, J. M., & Sevian, H.	2022	Development of a framework to capture abstraction in physical chemistry problem solving	*Chemistry Education Research and Practice*	23	1	55–77	https://doi.org/10.1039/D1RP00119A
Keen & Sevian, 2021	Keen, C., & Sevian, H.	2021	Qualifying domains of student struggle in undergraduate general chemistry laboratory	*Chemistry Education Research and Practice*	23	1	12–37	https://doi.org/10.1039/D1RP00051A
Kelly et al., 2021	Kelly, R. M., Akaygun, S., Hansen, S. J. R., Villalta-Cerdas, A., & Adam, J.	2021	Examining learning of atomic level ideas about precipitation reactions with a resources framework	*Chemistry Education Research and Practice*	22	4	886–904	https://doi.org/10.1039/D0RP00071J
Kulatunga et al., 2013	Kulatunga, U., Moog, R. S., & Lewis, J. E.	2013	Argumentation and participation patterns in general chemistry peer-led sessions	*Journal of Research in Science Teaching*	50	10	1207–1231	https://doi.org/10.1002/tea.21107
Lazenby et al., 2019	Lazenby, K., Rupp, C. A., Brandriet, A., Mauger-Sonnek, K., & Becker, N. M.	2019	Undergraduate chemistry students' conceptualization of models in general chemistry.	*Journal of Chemical Education*	96	3	455–468	https://doi.org/10.1021/acs.jchemed.8b00813
Lazenby et al., 2020a	Lazenby, K., Stricker, A., Brandriet, A., Rupp, C. A., & Becker, N. M.	2020	Undergraduate chemistry students' epistemic criteria for scientific models	*Journal of Chemical Education*	97	1	16–26	https://doi.org/10.1021/acs.jchemed.9b00505

Table 1.S1. (Continued) Bibliographic information for the articles selected for the review, arranged alphabetically by last name of first author.

Abbreviation	Authors	Year	Title	Journal	Volume	Issue	Page Numbers	DOI
Lazenby et al., 2020b	Lazenby, K., Stricker, A., Brandriet, A., Rupp, C. A., Mauger-Sonnek, K., & Becker	2020	Mapping undergraduate chemistry students' epistemic ideas about models and modeling	*Journal of Research in Science Teaching*	57	5	794–824	https://doi.org/10.1002/tea.21614
Lee et al., 2022	Lee, J., Park, T., & Davis, R. O.	2022	What affects learner engagement in flipped learning and what predicts its outcomes?	*British Journal of Educational Technology*	53	2	211–228	https://doi.org/10.1111/bjet.12717
Li et al., 2013	Li, W.-T., Liang, J.-C., & Tsai, C.-C.	2013	Relational analysis of college chemistry-major students' conceptions of and approaches to learning chemistry	*Chemistry Education Research and Practice*	14	4	555–565	https://doi.org/10.1039/C3RP00034F
Lopez et al., 2014	Lopez, E. J., Shavelson, R. J., Nandagopal, K., Szu, E., & Penn, J.	2014	Ethnically diverse students' knowledge structures in first-semester organic chemistry	*Journal of Research in Science Teaching*	51	6	741–758	https://doi.org/10.1002/tea.21160
Manneh et al., 2018	Manneh, I. A., Rundgren, C.-J., Hamza, K. M., & Eriksson, L.	2018	Tutor-student interaction in undergraduate chemistry: A case of learning to make relevant distinctions of molecular structures for determining oxidation states	*International Journal of Science Education*	40	16	2023–2043	https://doi.org/10.1080/09500693.2018.1517423
Marchlewicz & Wink, 2011	Marchlewicz, S. C., & Wink, D. J.	2011	Using the activity model of inquiry to enhance general chemistry students' understanding of Nature of Science	*Journal of Chemical Education*	88	8	1041–1047	https://doi.org/10.1021/ed100363n

Table 1.S1. **(Continued)** Bibliographic information for the articles selected for the review, arranged alphabetically by last name of first author.

Abbreviation	Authors	Year	Title	Journal	Volume	Issue	Page Numbers	DOI
Mazzarone & Grove, 2013	Mazzarone, K. M., & Grove, N. P.	2013	Understanding epistemological development in first- and second-year chemistry students	*Journal of Chemical Education*	90	8	968–975	https://doi.org/10.1021/ed300655s
McDonald, 2010	McDonald, C. V.	2010	The influence of explicit nature of science and argumentation instruction on preservice primary teachers' views of nature of science	*Journal of Research in Science Teaching*	47	9	1137–1164	https://doi.org/10.1002/tea.20377
Moon et al., 2019	Moon, A., Moeller, R., Gere, A. R., & Shultz, G. V.	2019	Application and testing of a framework for characterizing the quality of scientific reasoning in chemistry students' writing on ocean acidification	*Chemistry Education Research and Practice*	20	3	484–494	https://doi.org/10.1039/C9RP00005D
Pabuccu & Erduran, 2017	Pabuccu, A., & Erduran, S.	2017	Beyond rote learning in organic chemistry: The infusion and impact of argumentation in tertiary education	*International Journal of Science Education*	39	9	1154–1172	https://doi.org/10.1080/09500693.2017.1319988
Petritis et al., 2021	Petritis, S. J., Kelley, C., & Talanquer, V.	2021	Exploring the impact of the framing of a laboratory experiment on the nature of student argumentation	*Chemistry Education Research and Practice*	22	1	105–121	https://doi.org/10.1039/D0RP00268B
Petritis et al., 2022	Petritis, S. J., Kelley, C., & Talanquer, V.	2022	Analysis of factors that affect the nature and quality of student laboratory argumentation	*Chemistry Education Research and Practice*	23	1	257–274	https://doi.org/10.1039/D1RP00298H

Table 1.S1. (Continued) Bibliographic information for the articles selected for the review, arranged alphabetically by last name of first author.

Abbreviation	Authors	Year	Title	Journal	Volume	Issue	Page Numbers	DOI
Rodriguez et al., 2020	Rodriguez, J.-M. G., Bain, K., & Towns, M. H.	2020	The role of epistemology and epistemic games in mediating the use of mathematics in chemistry: Implications for mathematics instruction and research on undergraduate	*International Journal of Research in Undergraduate Mathematics Education*	6	2	279–301	https://doi.org/10.1007/s40753-019-00110-8
Russell & Weaver, 2011	Russell, C. B., & Weaver, G. C.	2011	A comparative study of traditional, inquiry-based, and research-based laboratory curricula: Impacts on understanding of the nature of science	*Chemistry Education Research and Practice*	12	1	57–67	https://doi.org/10.1039/C1RP90008K
Saribas et al., 2013	Saribas, D., Mugaloglu, E. Z., & Bayram, H.	2013	Creating metacognitive awareness in the lab: Outcomes for preservice science teachers	*Eurasia Journal of Mathematics, Science and Technology Education*	9	1	83–88	https://doi.org/10.12973/eurasia.2013.918a
Sendur et al., 2017	Sendur, G., Polat, M., & Kazancı, C.	2017	Does a course on the history and philosophy of chemistry have any effect on prospective chemistry teachers' perceptions? The case of chemistry and the	*Chemistry Education Research and Practice*	18	4	601–629	https://doi.org/10.1039/C7RP00054E
Sevian & Coutre, 2018	Sevian, H., & Couture, S.	2018	Epistemic games in substance characterization	*Chemistry Education Research and Practice*	19	4	1029–1054	https://doi.org/10.1039/C8RP00047F
Shultz & Gere, 2015	Shultz, G. V., & Gere, A. R.	2015	Writing-to-learn the Nature of Science in the context of the Lewis dot structure model	*Journal of Chemical Education*	92	8	1325–1329	https://doi.org/10.1021/acs.jchemed.5b00064

Table 1.S1. (Continued) Bibliographic information for the articles selected for the review, arranged alphabetically by last name of first author.

Abbreviation	Authors	Year	Title	Journal	Volume	Issue	Page Numbers	DOI
Sinapuelas & Stacy, 2015	Sinapuelas, M. L. S., & Stacy, A. M.	2015	The relationship between student success in introductory university chemistry and approaches to learning outside of the classroom	*Journal of Research in Science Teaching*	52	6	790–815	https://doi.org/10.1002/tea.21215
Talanquer, 2010	Talanquer, V.	2010	Exploring dominant types of explanations built by general chemistry students	*International Journal of Science Education*	32	18	2393–2412	https://doi.org/10.1080/09500690903369662
Venessa et al., 2019	Venessa, D. M., Hernani, H., & Halimatul, H. S.	2019	Exploring view of nature of science and technology pre-service chemistry teachers	*Journal of Science Learning*	3	1	19–28	https://doi.org/10.17509/jsl.v3i1.17757
Walker & Sampson, 2013	Walker, J. P., & 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	*Journal of Research in Science Teaching*	50	5	561–596	https://doi.org/10.1002/tea.21082
Walker et al., 2019	Walker, J. P., Van Duzor, A. G., & Lower, M. A.	2019	Facilitating argumentation in the laboratory: The challenges of claim change and justification by theory	*Journal of Chemical Education*	96	3	435–444	https://doi.org/10.1021/acs.jchemed.8b00745
Wickman, 2004	Wickman, P.-O.	2004	The practical epistemologies of the classroom: A study of laboratory work	*Science Education*	88	3	325–344	https://doi.org/10.1002/sce.10129

Table 1.S2. Study population, unit of analysis, methodology, and study design codes assigned to each article.

Article	Study Population	Unit of Analysis	Methodology	Study Design
Aglarci et al., 2016	Chem for pre-service	individual	qualitative	intervention
Aguirre-Mendez et al., 2020	Intro chem	individual	mixed methods	correlational
Agustian, 2020	PChem	individual	qualitative	exploratory
Becker et al., 2013	PChem	group	qualitative	exploratory
Bowen et al., 2022	OChem	individual	mixed methods	intervention
Burrows et al., 2017	OChem lab	individual	qualitative	exploratory
Calik & Cobern, 2017	Chem for pre-service	individual	mixed methods	intervention
Celik, 2020	Chem for pre-service	individual	qualitative	intervention
Chopra et al., 2017	Intro chem lab	individual	qualitative	intervention
Christian & Talanquer, 2012	OChem	group	qualitative	exploratory
Cigdemoglu et al., 2017	Intro chem	individual and group	mixed methods	intervention
Crujeiras-Perez & Brocos,	Chem for pre-service	group	qualitative	exploratory
Flaherty, 2020	OChem	individual	qualitative	exploratory
Gardner, 2017	Intro chem	individual	qualitative	intervention
Goff et al., 2012	Intro chem	individual	mixed methods	intervention
Grooms, 2020	Intro chem lab	individual	mixed methods	intervention
Grove & Bretz, 2010	OChem	individual	qualitative	longitudinal
Grove & Bretz, 2012	OChem	individual	qualitative	longitudinal
Havdala & Ashkenazi, 2007	Intro chem lab	individual	qualitative	exploratory
Hofer, 2004	Non-course specific	individual	qualitative	exploratory
Hosbein et al., 2021	Intro chem lab	group	mixed methods	longitudinal
Hunter et al., 2021	Intro chem	group	qualitative	exploratory
Karch & Sevian, 2022	PChem	individual and group	qualitative	exploratory
Keen & Sevian, 2021	Intro chem	individual	qualitative	exploratory
Kelly et al., 2021	Intro chem	individual	qualitative	exploratory
Kulatunga et al., 2013	Intro chem	group	mixed methods	exploratory

Table 1.S2. (Continued) Study population, unit of analysis, methodology, and study design codes assigned to each article.

Article	Study Population	Unit of Analysis	Methodology	Study Design
Lazenby et al., 2019	Intro chem	individual	mixed methods	exploratory
Lazenby et al., 2020a	Intro chem	individual	qualitative	exploratory
Lazenby et al., 2020b	Intro chem	individual	qualitative	exploratory
Lee et al., 2022	Intro chem	individual	quantitative	correlational
Li et al., 2013	Non-course specific	individual	quantitative	correlational
Lopez et al., 2014	OChem	individual	mixed methods	correlational
Manneh et al., 2018	Intro chem	group	qualitative	exploratory
Marchlewicz & Wink, 2011	Intro chem	individual	qualitative	intervention
Mazzarone & Grove, 2013	Non-course specific	individual	qualitative	longitudinal
McDonald, 2010	Chem for pre-service	individual	qualitative	intervention
Moon et al., 2019	Intro chem	individual	mixed methods	exploratory
Pabuccu & Erduran, 2017	OChem	group	qualitative	intervention
Petritis et al., 2021	OChem lab	individual and group	mixed methods	intervention
Petritis et al., 2022	OChem lab	individual and group	mixed methods	exploratory
Rodriguez et al., 2020	Intro chem	individual	qualitative	exploratory
Russell & Weaver, 2011	Intro chem	individual	qualitative	intervention
Saribas et al., 2013	Intro chem lab	individual	mixed methods	intervention
Sendur et al., 2017	Chem for pre-service	individual	qualitative	intervention
Sevian & Coutre, 2018	Non-course specific	individual	qualitative	exploratory
Shultz & Gere, 2015	Intro chem	individual	mixed methods	intervention
Sinapuelas & Stacy, 2015	Intro chem	individual	mixed methods	correlational
Talanquer, 2010	Intro chem	individual	qualitative	exploratory
Venessa et al., 2019	Chem for pre-service	individual	qualitative	exploratory
Walker & Sampson, 2013	Intro chem lab	individual and group	mixed methods	longitudinal
Walker et al., 2019	Intro chem lab	individual and group	qualitative	exploratory
Wickman, 2004	Intro chem	group	qualitative	exploratory

Table 1.S3. Methods of data collection and whether they probe students' epistemic ideas directly (explicit) or indirectly (implicit).

Article	Data Sources	Explicitness
Adams et al., 2008	Selected-response survey	Explicit
Aglarci et al., 2016	Open-ended survey	Explicit
	Interview	Explicit
Aguirre-Mendez et al., 2020	Written artifact	Explicit
	Open-ended survey	Explicit
Agustian, 2020	Open-ended survey	Explicit
	Interview	Explicit
Becker et al., 2013	Classroom recording	Implicit
Bowen et al., 2022	Open-ended survey	Implicit
Burrows et al., 2017	Interview	Implicit
Calik & Cobern, 2017	Selected-response survey	Explicit
Celik, 2020	Open-ended survey	Explicit
Chopra et al., 2017	Interview	Implicit
Christian & Talanquer, 2012	Classroom recording	Implicit
Cigdemoglu et al., 2017	Open-ended survey	Implicit
	Classroom recording	Implicit
	Written artifact	Implicit
Crujeiras-Perez & Brocos, 2021	Classroom recording	Implicit
	Written artifact	Implicit
Flaherty, 2020	Interview	Explicit
Gardner, 2017	Open-ended survey	Explicit
	Interview	Explicit
Goff et al., 2012	Open-ended survey	Explicit
	Interview	Explicit

Table 1.S3. (Continued) Methods of data collection and whether they probe students' epistemic ideas directly (explicit) or indirectly (implicit).

Article	Data Sources	Explicitness
Grove & Bretz, 2010	Written artifact	Implicit
	Interview	Implicit
Grove & Bretz, 2012	Written artifact	Explicit
	Interview	Explicit
Havdala & Ashkenazi, 2007	Interview	Explicit
	Written artifact	Implicit
Hofer, 2004	Interview	Explicit
Hosbein et al., 2021	Classroom recording	Implicit
Hunter et al., 2021	Interview	Implicit
Karch & Sevian, 2022	Interview	Implicit
Kaufman et al., 2017	Interview	Implicit
Keen & Sevian, 2021	Classroom recording	Implicit
	Open-ended survey	Implicit
	Interview	Implicit
Kelly et al., 2021	Interview	Implicit
Kulatunga et al., 2013	Classroom recording	Implicit
Lazenby et al., 2019	Open-ended survey	Explicit
Lazenby et al., 2020a	Open-ended survey	Explicit
Lazenby et al., 2020b	Open-ended survey	Explicit
Lee et al., 2022	Selected-response survey	Explicit
Li et al., 2013	Selected-response survey	Explicit
Lopez et al., 2014	Interview	Explicit
Manneh et al., 2018	Classroom recording	Implicit
Marchlewicz & Wink, 2011	Open-ended survey	Explicit
	Written artifact	Explicit

Table 1.S3. (Continued) Methods of data collection and whether they probe students' epistemic ideas directly (explicit) or indirectly (implicit).

Article	Data Sources	Explicitness
Mazzarone & Grove, 2013	Selected-response survey	Explicit
	Interview	Explicit
McDonald, 2010	Open-ended survey	Explicit
	Open-ended survey	Implicit
	Interview	Explicit
	Classroom recording	Implicit
	Written artifact	Implicit
Moon et al., 2019	Written artifact	Implicit
Pabuccu & Erduran, 2017	Written artifact	Implicit
	Classroom recording	Implicit
Petritis et al., 2021	Written artifact	Implicit
Petritis et al., 2022	Written artifact	Implicit
Rodriguez et al., 2020	Interview	Implicit
Russell & Weaver, 2011	Interview	Explicit
Saribas et al., 2013	Open-ended survey	Explicit
Sendur et al., 2017	Open-ended survey	Implicit
Sendur et al., 2017	Interview	Explicit
Sevian & Coutre, 2018	Open-ended survey	Implicit
	Interview	Implicit
Shultz & Gere, 2015	Open-ended survey	Explicit
Sinapuelas & Stacy, 2015	Interview	Explicit
Talanquer, 2010	Open-ended survey	Implicit
Venessa et al., 2019	Selected-response survey	Explicit
Walker & Sampson, 2013	Classroom recording	Implicit
	Written artifact	Implicit
Walker et al., 2019	Classroom recording	Implicit
	Interview	Explicit
Wickman, 2004	Classroom recording	Implicit

Table 1.S4. Hierarchy, stability, explicitness, inferred model of epistemic cognition, and stated model of epistemic cognition codes assigned to each article.

Article	Hierarchy Code	Stability Code	Explicitness Code	Inferred Model of Epistemic Cognition	Stated Model of Epistemic Cognition
Adams et al., 2008	Hierarchical	Stable	Explicit	Developmental or Dimensional	
Aglarci et al., 2016	Hierarchical	Stable	Explicit	Developmental or Dimensional	
Aguirre-Mendez et al., 2020	Hierarchical	Unclear	Explicit	Developmental or Dimensional	
Agustian, 2020	Hierarchical	Unclear	Explicit	Developmental or Dimensional	
Becker et al., 2013	Variable utility	Unstable	Implicit	Resources	
Bowen et al., 2022	Ambiguous	Unstable	Both	Unknown	Resources
Burrows et al., 2017	Hierarchical	Unclear	Implicit	Unknown	
Calik & Cobern, 2017	Hierarchical	Stable	Explicit	Developmental or Dimensional	
Celik, 2020	Hierarchical	Stable	Explicit	Developmental or Dimensional	
Chopra et al., 2017	Ambiguous	Unclear	Explicit	Unknown	
Christian & Talanquer, 2012	Hierarchical	Unstable	Implicit	Unknown	
Cigdemoglu et al., 2017	Hierarchical	Stable	Implicit	Developmental or Dimensional	
Crujeiras-Perez & Brocos, 2021	Hierarchical	Unclear	Implicit	Unknown	
Flaherty, 2020	Hierarchical	Unstable	Explicit	Unknown	

Table 1.S4. (Continued) Hierarchy, stability, explicitness, inferred model of epistemic cognition, and stated model of epistemic cognition codes assigned to each article.

Article	Hierarchy Code	Stability Code	Explicitness Code	Inferred Model of Epistemic Cognition	Stated Model of Epistemic Cognition
Garcia-Carmona et al., 2017	Hierarchical	Stable	Explicit	Developmental or Dimensional	
Gardner, 2017	Hierarchical	Stable	Explicit	Developmental or Dimensional	
Goff et al., 2012	Hierarchical	Stable	Explicit	Developmental or Dimensional	
Grooms, 2020	Hierarchical	Stable	Both	Developmental or Dimensional	
Grove & Bretz, 2010	Hierarchical	Stable	Implicit	Developmental or Dimensional	Developmental
Grove & Bretz, 2012	Hierarchical	Unclear	Implicit	Unknown	
Havdala & Ashkenazi, 2007	Hierarchical	Unclear	Both	Unknown	Dimensional
Hofer, 2004	Hierarchical	Stable	Explicit	Developmental or Dimensional	Dimensional
Hosbein et al., 2021	Hierarchical	Unclear	Implicit	Unknown	
Hunter et al., 2021	Ambiguous	Unstable	Implicit	Resources	
Karch & Sevian, 2022	Ambiguous	Unstable	Implicit	Resources	
Keen & Sevian, 2021	Variable utility	Unstable	Implicit	Resources	
Kelly et al., 2021	Variable utility	Unstable	Implicit	Resources	Resources
Kulatunga et al., 2013	Hierarchical	Unclear	Implicit	Unknown	

Table 1.S4. (Continued) Hierarchy, stability, explicitness, inferred model of epistemic cognition, and stated model of epistemic cognition codes assigned to each article.

Article	**Hierarchy Code**	**Stability Code**	**Explicitness Code**	**Inferred Model of Epistemic Cognition**	**Stated Model of Epistemic Cognition**
Lazenby et al., 2019	Ambiguous	Unstable	Explicit	Unknown	Resources
Lazenby et al., 2020a	Ambiguous	Unstable	Explicit	Unknown	Resources
Lazenby et al., 2020b	Hierarchical	Unstable	Explicit	Unknown	Resources
Lee et al., 2022	Hierarchical	Unclear	Explicit	Developmental or Dimensional	
Li et al., 2013	Hierarchical	Unclear	Explicit	Developmental or Dimensional	
Lopez et al., 2014	Hierarchical	Stable	Explicit	Developmental or Dimensional	
Manneh et al., 2018	Variable utility	Unstable	Implicit	Resources	
Marchlewicz & Wink, 2011	Hierarchical	Stable	Explicit	Developmental or Dimensional	
Mazzarone & Grove, 2013	Hierarchical	Stable	Explicit	Developmental or Dimensional	
McDonald, 2010	Hierarchical	Unclear	Both	Unknown	
Moon et al., 2019	Hierarchical	Unclear	Implicit	Unknown	
Pabuccu & Erduran, 2017	Hierarchical	Unstable	Implicit	Unknown	
Petritis et al., 2021	Hierarchical	Unstable	Implicit	Unknown	
Petritis et al., 2022	Hierarchical	Unstable	Implicit	Unknown	

Table 1.S4. (Continued) Hierarchy, stability, explicitness, inferred model of epistemic cognition, and stated model of epistemic cognition codes assigned to each article.

Article	**Hierarchy Code**	**Stability Code**	**Explicitness Code**	**Inferred Model of Epistemic Cognition**	**Stated Model of Epistemic Cognition**
Rodriguez et al., 2020	Variable utility	Unstable	Implicit	Resources	Resources
Russell & Weaver, 2011	Hierarchical	Stable	Explicit	Developmental or Dimensional	
Saribas et al., 2013	Hierarchical	Stable	Explicit	Developmental or Dimensional	
Sendur et al., 2017	Ambiguous	Stable	Implicit	Unknown	
Sevian & Coutre, 2018	Variable utility	Unstable	Implicit	Resources	Resources
Shultz & Gere, 2015	Hierarchical	Stable	Explicit	Developmental or Dimensional	
Sinapuelas & Stacy, 2015	Hierarchical	Stable	Explicit	Developmental or Dimensional	
Talanquer, 2010	Ambiguous	Unstable	Implicit	Resources	
Venessa et al., 2019	Hierarchical	Unclear	Explicit	Developmental or Dimensional	
Walker & Sampson, 2013	Hierarchical	Unclear	Implicit	Unknown	
Walker et al., 2019	Hierarchical	Unclear	Both	Unknown	
Wickman, 2004	Variable utility	Unstable	Implicit	Resources	

Chapter 4

Beliefs versus Resources: A Tale of Two Models of Epistemology

This work was conducted in collaboration with Rosemary S. Russ, Prayas Sutar, and Ryan L. Stowe.

Abstract

Compelling evidence, from multiple levels of schooling, suggests that teachers' knowledge and beliefs about knowledge, knowing, and learning (*i.e.*, epistemologies) play a strong role in shaping their approaches to teaching and learning. Given the importance of epistemologies in science teaching, we as researchers must pay careful attention to how we model them in our work. That is, we must work to explicitly and cogently develop theoretical models of epistemology that account for the learning phenomena we observe in classrooms and other settings. Here, we use interpretation of instructor interview data to explore the constraints and affordances of two models of epistemology common in chemistry and science education scholarship: epistemological beliefs and epistemological resources. Epistemological beliefs are typically assumed to be stable across time and place and to lie somewhere on a continuum from "instructor-centered" (worse) to "student-centered" (better). By contrast, a resources model of epistemology contends that one's view on knowledge and knowing is compiled in-the-moment from small-grain units of cognition called resources. Thus, one's epistemology may change one moment to the next. Further, the resources model explicitly rejects the notion that there is one "best" epistemology, instead positing that different epistemologies are useful in different contexts. Using both epistemological models to infer instructors' epistemologies from dialogue about their approaches to teaching and learning, we demonstrate that how one models epistemology impacts the kind of analyses possible as well as reasonable implications for supporting instructor learning. Adoption of a beliefs model enables claims about which instructors have "better" or "worse" beliefs and suggests the value of interventions aimed at shifting toward "better" beliefs. By contrast, modeling epistemology as in

situ activation of resources enables us to explain observed instability in instructors' views on knowing and learning, surface and describe potentially productive epistemological resources, and consider instructor learning as refining valuable intuition rather than "fixing" "wrong beliefs.

Introduction

It goes without saying that chemistry instructors at the undergraduate level have a great deal of knowledge about chemistry. The content they teach is rich and complex and requires nuanced understandings of an incredible array of phenomena. However, in addition to this knowledge of chemistry, instructors also have a great deal of knowledge and beliefs—albeit potentially tacit—about teaching and learning (Hora, 2014; Gibbons et al., 2018; Popova et al., 2020). For example, consider two different instructors' understandings of teaching and learning chemistry.

> One of my most important roles as an instructor was to show people how the ideas interconnected... I should be doing something that goes, I guess, beyond just following the textbook because that's information they already can get. –Liam
>
> The process of learning what a model is, what it applies to, and going through the practice of application of that model to explain an outcome and seeing that those things can be connected is the powerful thing we want our science students to do. –James

From these quotes, we might infer that Liam conceptualizes knowledge as consisting of many pieces of information that must be connected and that James sees learning as constructing, applying, and connecting models to explain phenomena. But what can these quotes tell us about their teaching?

Research in teaching and teacher education demonstrates that teacher thinking about teaching and learning has a substantial impact on teacher practice (*e.g.*, Pajares, 1992; Clark & Peterson, 1986; Mansour, 2009). Teachers' implementations of curricular reforms are influenced by beliefs about teaching and learning (Haney et al.,1996; Wallace & Kang, 2004; Roehrig et al., 2007) as are smaller day-to-day decisions like how much time to spend on a particular topic (Cronin-Jones,

1991) or their interaction with curricular materials (Remillard, 2005). The relationship between beliefs and practice is complex and its strength may vary depending on contextual factors (Fang, 1996). Nevertheless, if we wish to support instructors in improving their teaching practices, the literature suggests we should attend to instructor thinking.

In this work then, we examine and unpack existing research on instructors' knowledge about teaching and learning in chemistry. First, we recast that work in terms of what has been referred to elsewhere in the science education literature as epistemologies (Hofer & Pintrich, 1997; Smith & Wenk, 2006; Havdala & Ashkenazi, 2007; Lising & Elby, 2005; Oliveira et al., 2012; Sandoval, 2005) or, more recently, epistemic cognitions (Greene et al., 2016). Specifically, epistemologies "consist of [people's] systems of beliefs [tacit or explicit] about (1) the nature of knowledge and (2) the processes of knowing" (Hofer & Pintrich, 1997). Second, we compare and contrast two models of instructor thinking, particularly in regard to their underlying assumptions about the stability and hierarchy of beliefs. We then analyze our interview data according to each model and discuss affordances and limitations of each. Finally, we consider the implications of each model on instructor professional development.

Literature Background

Education researchers have long sought to understand aspects of instructors' thinking that give rise to their teaching practice (Abell, 2008; Clark and Peterson, 1986; Kagan, 1992; Schoenfeld, 1998; Shulman, 1986). This approach to studying teaching practice is rooted in a cognitive paradigm that "conceptualizes teaching largely in terms of [teachers'] mental life and focuses on teaching as a way of thinking with a particular set of specialized knowledge and cognitive processes" (Russ et al., 2016). Within this tradition, scholars of science education have examined teacher's knowledge, beliefs, identities, and goals in an attempt to get "under the hood" of teacher practice (*e.g*., Abd-El-Khalick et al., 1998; Avraamidou & Zembal-Saul, 2010; Lederman, 1999; Loughran et al., 2004; Pajares, 1992; Remillard, 2005). Within chemistry education, scholars have similarly focused on instructors' attitudes, beliefs, and orientations toward teaching (*e.g*., Gibbons et al., 2018; Mack & Towns, 2016; Popova et al., 2020).

Of specific concern within science education has been the set of knowledge and beliefs that teachers possess that is associated with knowledge, knowing, and learning. For example, participants may view knowledge as constructed from things they already know or knowledge as transferred from authority. Further, they may view science learning as either an opportunity to make sense of phenomena or to memorize information. Although researchers use a range of constructs to conceptualize these knowledge and beliefs, here we follow work in science education that characterizes them as epistemologies (Hofer & Pintrich, 1997) or, more recently, epistemic cognitions (Greene et al., 2016).

Tracing back to the 1970s, scholars have worked both to identify participants' epistemologies and also to tie those views to classroom practices of teaching and learning. Both correlational and case-study evidence suggests that epistemologies play an important role in school settings (Greene et al., 2018; Liang & Tsai, 2010; Rosenberg et al., 2006). A range of researchers across both K-12 and undergraduate settings have explored how instructors' tacit views of knowledge and knowing impact the ways they engage in teaching (Wendell et al., 2019). For example, Russ and Luna (2013) followed a high school teacher across multiple class sessions to identify how her teaching practice shifted depending on whether she viewed teaching as an opportunity to Connect Biological Ideas or Use Procedural Knowledge. Similarly, Chari and her colleagues (2019) analyzed 50 episodes of upper-division, undergraduate physics instruction to demonstrate how differing behavior of instructors was shaped by their two-dimensional epistemological understanding of problem-solving as being algorithmic/conceptual and mathematics/physics. Likewise, within chemistry education, researchers have probed the link between instructor thinking and practice. Gibbons et al. (2018) conducted a large-scale study of chemistry instructors and found correlations between the instructors' beliefs about teaching and learning and reported pedagogical practices. Popova et al. (2020) focused specifically on assistant chemistry professors and similarly found some alignment between beliefs and practices.

These findings from across science education bear out the assumption that epistemology plays a strong role in shaping the teaching practices of instructors in science courses. As such, here we take as a given that epistemologies are an important piece of what lies "under the hood" in chemistry instructors' approaches to teaching and learning. Further, given the importance of epistemologies in science teaching, we as researchers must pay careful attention to how we model them in our work. That is, we must work to explicitly and cogently develop theoretical models of epistemology that account for the learning phenomena we observe in classrooms.

Theoretical Framework

In our review of the science education literature, we identified two distinct approaches to modeling epistemology. In one approach, epistemologies are seen as "theories" that people consciously possess and apply in their lives (Hofer & Pintrich, 1997; Hashweh, 1996; Davis, 2003; Kittleson, 2011; Havdala & Ashkenazi, 2007). These are often referred to as "epistemological beliefs" (Hofer & Pintrich, 1997; Schommer-Aikins, 2004). The other approach views epistemology as constructed in-the-moment from "epistemological resources" —fine-grained knowledge elements concerning knowledge and the nature of knowing (Hammer & Elby, 2002). These models differ from each other in two key aspects: the extent to which epistemologies are assumed to be stable and whether or not epistemologies develop hierarchically over time. We describe each model and its underlying assumptions in more detail below.

A Focus on Beliefs

Modeling epistemologies as beliefs is common across science education literature and is especially prominent in chemistry education research. For example, Popova et al. (2020) interviewed assistant chemistry professors about their beliefs and checked in two years later to see how these beliefs changed (Popova et al., 2021). Mack and Towns (2016) focused on physical chemistry instructors and interviewed them about their approach to teaching, which revealed beliefs about the purpose of their courses and the nature of knowledge in their discipline. Other studies have described instructors' beliefs in the context of specific topics, such as systems thinking (Szozda et al., 2022) and grading (Mutambuki & Fynewever, 2012).

Although studies on instructor beliefs have uncovered a variety of beliefs regarding teaching and learning, many further classify their beliefs (and/or practices) as instructor-centered or student-centered (*e.g*., Gibbons et al., 2018; Popova et al., 2020; Popova et al., 2021). Instructor-centered beliefs are associated with a transmission view of learning and include beliefs that students learn chemistry most effectively by taking notes during lecture or doing homework problems. In contrast, believing that students learn chemistry most effectively by working in groups or making connections between chemistry and everyday life is considered student-centered and is associated with a constructivist view of learning. In their implications, the authors of these studies discussed ways to shift instructors' epistemological beliefs and their practice from instructor-centered to student-centered.

Modeling epistemologies as beliefs brings with it a set of common, if tacit, features. These features include: 1) beliefs are stable and 2) beliefs develop hierarchically over time. We detail these features below with examples from the literature.

Beliefs are stable.

Chemistry instructor beliefs are often treated as stable over time. We can infer this feature from the methodologies – commonly longitudinal studies – used to study these beliefs. If beliefs are assumed to be unstable over the period of minutes or hours, we would expect to see studies looking at changes during this time scale. However, if beliefs are assumed to be stable over longer periods of time (*e.g.*, months or years), then it would be logical to collect data less frequently, perhaps once a semester or once a year. In the chemistry education literature, we mostly observe the latter. For example, Popova et al. (2021) conducted a study on assistant chemistry professors in which they compared participants' initial beliefs to their beliefs two years later, implying that changes were expected to occur on a longer time scale. Similarly, using a pre/post study design, in which beliefs are measured before and after an intervention, is reasonable if one assumes that the participants' beliefs would be essentially unchanged in the absence of the intervention for the duration of the study. Stains et al. (2015) have conducted such a study to measure the impact of a professional development program on assistant chemistry professors' beliefs. Conversely, we are not aware of any studies that characterize how chemistry instructors' thinking changes moment-to-moment.

Beliefs develop hierarchically over time.

In the tradition of Piagetian stages of the 1960s (Piaget, 1969; Piaget 1970) or the Expert-Novice studies of the 1980s (Chi et al., 1988), beliefs are often modeled as moving through a progression in which they become more sophisticated over long periods of time (Perry, 1970; King

& Kitchener, 1994; Kuhn, 1991; Hofer & Pintrich, 1997). For example, in order to develop a chemistry version of the Colorado Learning Attitudes about Science Survey (CLASS), originally developed for physics education research, Adams et al. (2006) interviewed non-major introductory chemistry students and chemistry faculty to establish the novice and expert responses, respectively, for survey items. This method makes sense if one expects differences in beliefs between these populations and similarities within each population. Furthermore, using their survey, the authors observed a "regression in beliefs" over a semester of general chemistry. The use of the term "regression" is consistent with a hierarchical, developmental model. Returning to the example of student-centered and instructor-centered beliefs, Popova et al. (2020) identified a cluster of beliefs they labeled "transitional and consistent," which contained a mixture of student-centered and instructor-centered beliefs. The label "transitional" implies an intermediate stage within a progression. While this continuum could be utilized in a purely descriptive manner, it has typically been presented in an evaluative manner. In their implication sections, the authors of these studies discuss ways to shift instructors from instructor-centered to student-centered, communicating that the latter is more desirable than the former.

A Focus on Epistemological Resources

In contrast to the model of epistemological beliefs commonly used in the chemistry education literature, another model of epistemology contends that it is made up of a range of smaller units of cognition known as resources (Hammer, 2000). Below we detail the features of this model, presenting them in contrast to the features embedded in a beliefs model of epistemology.

Epistemological resources are unstable.

Rather than understanding epistemologies as beliefs that are relatively stable across time and place, epistemological resources are taken to be unstable across contexts. As in the case with beliefs, this assumption shows up in the methods researchers use to study and document epistemologies. Specifically, researchers use methods that allow them to capture rich data over relatively short time spans on the order of minutes. For example, in a case study of a group of 8th graders reasoning about the rock cycle, Rosenberg and his colleagues (2006) use classroom video to demonstrate how students transition from one epistemology to another in a matter of moments based on a single comment from their teacher. Similarly, transitions in epistemologies that occur over minutes (rather than the hours, days, or years assumed in more stable models of cognition) have been documented in short excerpts (as few as 5–10 lines of transcript) in college physics classes (Scherr & Hammer, 2009; Modir et al., 2017; Irving et al., 2013; Dini & Hammer, 2017). The "framework of epistemological resources, smaller and more general than theories or traits" accommodates this dynamic contextual dependence (Hammer & Elby, 2002).

This unstable model of epistemology is rooted in a similar model of mind for conceptual understanding that may be more familiar to the reader (diSessa, 1993). Although science education began by comparing student thinking to scientific paradigms or robust scientific theories (McCloskey, 1983; Strike & Posner, 1985; Hewson & Hewson, 1984), a commitment to the notion of constructivism has demanded a move away from this (mis)conceptions model (Smith III et al., 1994). Specifically, the field is now "skeptical of treating knowledge or abilities as things one acquires and manipulates as intact units" (Hammer et al., 2005). Instead, we now think of conceptual knowledge as a complex system of many "pieces" (diSessa, 1993) which students unconsciously and dynamically assemble and disassemble in moments of thinking (Sherin, 2006; Philip, 2011; Minstrell, 1989). An epistemological resources model assumes the same is true for epistemology (Hammer, 2000; Hammer & Elby, 2002). Instead of people having "pre-compiled" (Hammer et al., 2005) views of knowledge that they call up in learning situations, an epistemological resource model assumes people compile their view of knowledge dynamically in real time by drawing on many small epistemological elements.

Epistemological resources are differentially useful in different contexts.

One of the key premises of a model of epistemological resources is that different situations call for different epistemologies (Elby & Hammer, 2001). For example, while the NGSS (NGSS Lead States, 2013) may encourage us to have students construct their own models for phenomena, we do not necessarily want the lay public to construct their own models for the spread of COVID (in fact the state of our public health may be drastically different if fewer people had done so!). In the former context (the classroom) we may want students to adopt a view that they can be the authority on knowledge, whereas in the sphere of COVID we want people to adopt a view that the scientific community is the authority. But even this grain size is not sufficient; it is not the case that the NGSS always wants students to believe they are the knowledge authority in classrooms. There are times in which we want students to adopt a view of learning where their teachers, or the textbook, are the authority—for example, when they are told a value like Avogadro's number.

Given the diversity and variability of epistemological resources that can be useful across the contexts of teaching and learning, researchers that adopt this model of epistemology explicitly reject a hierarchical model of progressive sophistication. Instead, this model assumes that there is no "more correct" or "more expert" epistemology but that instead epistemological resources are differentially productive for learning in context. Sophistication then is not merely adopting a set of expert views but is instead the ability to "explore and discuss the differences between knowledge in multiple contexts" (Elby & Hammer, 2001). In the case of teachers, epistemological expertise involves the "awareness and judicious use of" (Russ, 2018) a range of epistemological resources. Stated differently, epistemological sophistication means possessing a suite of epistemological resources as well as a finely tuned mechanism for identifying which contexts call for which resources.

Research Questions

In the proceeding sections, we have described assumptions that underpin two common models for epistemology (epistemological beliefs and epistemological resources). Here, we take a look at what these models let us infer about chemistry instructors' epistemologies from dialogue about their approaches to teaching and learning. Specifically, we examine whether modeling instructors' epistemologies as resources supports different implications for instructor learning than modeling instructors' epistemologies as hierarchical, stable beliefs. The following research questions guided our efforts:

1) What epistemologies do chemistry instructors articulate when talking about their approaches to teaching and learning in undergraduate organic chemistry?
2) What are the affordances and limitations of modeling instructor thinking as beliefs and as epistemological resources?

Our purpose here is to show that the model of epistemology researchers choose powerfully influences the kind of analysis they conduct on their data and what they can infer about useful approaches to supporting instructor learning.

Methods

Context and Participants

This study focused on introductory organic chemistry instructors at a large public university in the Midwest. Although much of the chemistry education research focuses on general chemistry, here we choose to focus on organic chemistry for two reasons – one opportunistic and one substantive. First, many discussions were taking place in the department regarding changing and/or unifying the course. As a result, there was a pre-existing need to understand the goals instructors have for their students' knowledge construction and the means by which they believe these goals can be achieved. Second, and perhaps more importantly for our argument here, organic chemistry instructors have considerable autonomy in how they teach. Thus, we expected that more of their decisions would be based on their own epistemologies rather than institutional constraints (*e.g.*, "I do this because my department says I have to"). This autonomy allows us to examine epistemologies more directly.

The introductory organic chemistry course at this university consists of two semesters (OChem I and OChem II). As this is a required course for chemistry, biology, and chemical engineering majors and anyone intending to pursue a career in the health field, it serves approximately 1,000 students each semester. Organic chemistry instruction is divided among tenured professors, pre-tenure professors, and non-tenure track professors. The non-tenure track professors typically teach both OChem I and OChem II while most tenured and pre-tenure professors teach only one of these courses. All instructors use the same textbook and there is general agreement regarding the content that should be covered in each course, but each instructor has the freedom to choose their own teaching practices, author their own exams, and determine how points are allocated in their course. Some instructors have chosen to teach jointly with shared course materials and exams.

Interview requests were sent to everyone involved in teaching introductory organic chemistry over the last five years. We chose to restrict invitations to instructors who taught in the last five years because presumably these people would still remember details of how they approach(ed) teaching the course and would be involved in teaching it for several more years. Ten organic chemistry instructors responded and consented to be interviewed. They included tenured professors, pre-tenure professors, and non-tenure track professors with teaching experience ranging from one year to approximately thirty years. Four of these instructors teach both OChem I and OChem II while the other six typically only teach one of these courses.

During data analysis, we utilized an intensity sampling approach (Creswell, 2007) to select "information-rich cases that manifest [teacher beliefs] intensely but not extremely" (p. 159). This approach allowed us to select a relatively small number of cases that provided in depth information for analysis; here we focus on three of the ten professors interviewed (Table 4.1). These instructors represent different roles within the department and exhibit a range of epistemological resources. James is a non-tenure track professor whose interview elicited fairly frequent and consistent epistemological ideas. Liam is a pre-tenure professor who demonstrated more inconsistency in his epistemic cognition. Mark is a tenured professor whose interview was most notable for the focus on logistical aspects of teaching rather than epistemological aspects.

Table 4.1. Relevant characteristics of instructors at the time they were interviewed.

Instructor[a]	Position	Courses Taught	Years of Teaching Experience
James	Non-tenure track	OChem I, OChem II	8
Liam	Pre-tenure	OChem I	1
Mark	Tenured	OChem II	15

[a] Actual names have been replaced with pseudonyms to protect participants' identities.

Data Collection

We chose to use interviews to infer instructor epistemologies. Interviews allowed the instructors to respond to the questions in their own words and in a more detailed manner than surveys typically allow. In recognition of the context-dependency of epistemic cognition, the interview questions were written to elicit reflections on the instructors' particular courses rather than general thoughts on teaching. Instructors could also supply context through the use of anecdotes and examples from their experiences. Additionally, the interview questions probed a range of teaching activities and contexts, from planning to assessment to student performance. However, such reflections are still filtered through the perceptions of the interviewees; thus, they are not equivalent to direct observations of the instructors as they lecture or author assessments (Alshenqeeti, 2014). Ideally, the interviews would be coupled with observations of the instructors as they taught, wrote assessments, graded assessments, etc. In the future, we hope to collect this data. Nevertheless, we believe that interviews can help us figure out productive ways to model epistemology and can prompt instructors to consider multiple contexts for their teaching practices.

Semi-structured interviews were conducted over Zoom by the third author and lasted approximately one hour. The interviews began with questions regarding how the instructor got interested in chemistry and why they chose to stay in academia following graduate school (Q1 & Q2). Then the instructors were asked why students should take organic chemistry, what the students should learn from the course, and how the students can maximize their learning (Q3, Q4, Q8). The interview also included a discussion of assessment: how the instructors evaluate learning, what they aim to assess, and how they interpret assessment responses (Q6 & Q9). The interviews concluded with questions about if and/or how the instructors make use of teaching resources, including advice from peers and chemistry education research, and the role of evidence in changing teaching practices (Q11–Q13). The full interview protocol can be found in the appendix.

The interviews were transcribed by Zoom, and the first author corrected these transcriptions as needed to ensure they were accurate. The first author also broke up longer sections of dialogue into utterances that focused on a particular idea. These served as the units of analysis while coding.

Data Analysis

Strand 1: Analyzing instructor beliefs

In our first strand of analysis, we sought to understand instructors' epistemologies using a beliefs model. To do so we developed an analytic scheme by looking across the work of multiple authors who seek to characterize teacher beliefs from interviews or from their practice (Luft &

Roehrig, 2007; Popova et al., 2021; Simmons et al., 1999). Although all of the authors identify a range of beliefs (*e.g.*, beliefs about student learning/actions, beliefs about the role/actions of teachers, beliefs about content, etc), their analyses ultimately cluster teachers by patterns of responses. Further, although they each have several different clusters, ultimately the clusters are placed along a continuum where the two ends are student-centered and instructor-centered beliefs. Synthesizing across the papers we identified some common key elements of these two ends of the spectrum.

- **Student-Centered**

  Instructors believe students learn by doing and not by listening; thus, the role of instructors involves collaborating with, facilitating, and guiding students as they construct ideas that are relevant to their lived experience from their prior knowledge.

- **Instructor-Centered**

  Instructors believe that students learn by paying attention and listening to the instructor; thus, the role of the instructor is to provide content and experiences so they can assess if students know a set of pre-defined facts.

In addition to these two ends of the spectrum, researchers typically also included a transitional or inconsistent category when an instructor evidenced beliefs from both ends of the spectrum.

In our work here, we used the two ends as a guide for our analysis; the first two authors read through the transcripts and together assigned a code of “student-centered” or “instructor-centered” to each utterance. (Recall that an utterance was a section of dialogue concerning a single topic, typically 5–8 sentences). We restricted our analysis to questions 3–10 of the interview protocol because these questions surfaced reflections on their own teaching rather than their perceptions of the department and the field of chemistry education. Furthermore, we ignored utterances that were not epistemic in nature (*e.g*., "Say that one more time."). For this analysis, we relied heavily on which pronouns ("I" versus "they/them") were used in the active voice and which were used in the passive voice when referencing teaching and learning. If an utterance included both student-centered and instructor-centered beliefs, it was labeled as "both." Table 4.2 provides some examples from our data for each of the two clusters.

Table 4.2. Examples of utterances coded as student-centered and instructor-centered.

Code	Example
Student-centered	What we need to be as educators, as teachers, is people who set up the students to have those experiences I just described. We need to be creating environments where somehow students are engaged in thinking about models, using models, writing about them to explain why something happens.
Instructor-centered	I felt like one of my most important roles as an instructor was to show people how the ideas interconnected. So whenever we introduce a new idea, be very clear about what is new in this idea… with kind of a very brief review of whatever that concept is. So I think that's something that is much harder to do when you're kind of working through, um, something kind of on your own.

Strand 2: Analyzing instructor resources

To describe the epistemological resources of chemistry instructors, we ground our analysis in the five-dimensional model proposed by Chinn and his colleagues (2011). We chose this model because it is, for us, the most comprehensive of all the existing models and is consistent with insights and components from other prominent scholars of epistemology (Hammer and Elby, 2002; Hofer and Pintrich, 1997; Schommer-Aikins, 2004). Below we will briefly describe each of the five dimensions.

1) **Epistemic Aims and Values**

   Epistemic aims are the goals relating to inquiry, and epistemic values describe the relative worth of particular aims. Aims, or what others call goals (Berland et al., 2016) are an important part of characterizing a person's epistemology because they are the ends to which other aspects of epistemic cognition are directed.

2) Structure of Knowledge

The structure of knowledge refers to how knowledge is organized and answers questions like “What kind of answer should our [learning] provide?” (Berland et al., 2016). They are akin to epistemic forms (Collins & Ferguson, 1993; Hammer & Elby, 2002) which are “target structures that guide inquiry.”

3) Reliable and Unreliable Processes for Achieving Epistemic Aims

Processes refer to the actions one takes to achieve one’s epistemic aims. Epistemic processes are similar to Hammer and colleagues’ (Hammer & Elby, 2002; Rosenberg et al., 2006) epistemological activities that help people (tacitly!) answer the question, “What are you doing?” in terms of knowledge construction or use. Processes are also consistent with what researchers in undergraduate physics education (Odden & Russ, 2018; Chen et al., 2013; Tuminaro & Redish, 2007) have called the moves in an epistemic game (Collins & Ferguson, 1993).

4) Sources, Justifications, and Stances

Sources of knowledge refers to where knowledge was obtained from, such as an expert, authority figure, textbook, or one’s direct experience. Justifications for knowledge are the criteria by which a person evaluates knowledge, such as coherence with prior knowledge, logical consistency, or support with acceptable evidence. Stances toward knowledge describe a person’s view on a given knowledge claim. Although Chinn et al. (2011) put these together because they are tightly linked in practice, other scholars treat these dimensions independently (Berland et al., 2016; Hammer and Elby; 2002; Tuminaro & Reddish, 2007).

5) Virtues and Vices

Epistemic virtues and vices encompass personal characteristics that either support or hinder epistemic endeavors. Few other scholars in science education discuss this dimension.

Using these categories as a guide for our coding process, the authors read each utterance of dialogue and discussed whether they saw anything that would fall into the categories. Once this was completed for the three interviews, the authors summarized their observations for each category into a succinct list of resource codes. Once individual codes were defined, the authors used the constant comparison method (Glaser & Strauss, 1967) to confirm that all utterances were coded with a stable codebook (Table 4.3).

Table 4.3. Epistemic Resource Codes and Examples

Epistemic Resource Codes	**Examples**
Aims and Values	
Memorization	The students then have to memorize these factoids and memorize these patterns instead of understanding the model where they don't have to memorize anything.
Explanation	I'd like to do a better job of assessing, um, is, um, actually getting some feedback myself about where they're deriving their explanations. So like when they say this would go through SN2, um, basically how can it be explained, um, like why, why did you say SN2, or what sort of factors do you think are at play here?
Problem-solving	So, uh, for somebody interested in, um, medicine, um, first of all, I guess like a large fraction of people taking the class, I think that, um, there are sort of aspects of the, the type of problem solving we do in organic chemistry that's really important. So, um, and sort of as specifically as I can, I guess what I feel like we're talking about is, uh, taking like a set of, uh, I guess, kind of starting criteria, like sort of the simple ideas, like steric bulk, um, electronic sort of perturbations, that have these principles and then trying to figure out how to sort of interconnect them to come up with an answer to a new sort of problem.
Usefulness	Now I have no belief that most of my students will do a distillation again after they leave my class. And I do not care if they ever do a distillation again, that's irrelevant. But I know that 100% of my students are going to apply models

	to explain systems. They're going to use models to predict outcomes. They're gonna use models to rationalize outcomes. We should have them engaged in doing that.
Structure of Knowledge	
Pieces	I think the important things for us to be actually getting from [the students] are like connecting concepts and that's not connecting any concepts. Um, but I think hitting at some of the individual concepts on their own is also important. Um, so making sure that they're getting those building blocks and that we're not only assessing them on connecting the building blocks. I find that’s also important.
Connections	And so I felt like one of my most important roles as an instructor was to show people how the ideas interconnected. So whenever we introduce a new idea, be very clear about what is new in this idea and what is drawing on things they've already learned, with kind of a very brief review of whatever that concept is.
Hierarchy (building up)	I think being able to connect independent concepts to address a more complex question, um, I think that's sort of a fundamental learning objective for organic chemistry.
Hierarchy (underlying)	And so, especially for these pre-professional students who may never take another science class beyond second semester organic chemistry, um, this teaches them how you master a complicated topic that demands more than just rote memorization, right? This, it really does kind of, uh, teach you that, um, cramming isn't feasible at, um, you do have to understand underlying mechanisms to really succeed in a class like this.
Reliable Processes	
Accumulating	But one of the challenges to, um, doing formative assessment, in my view, is that because we put so much content in the class, I think it, I found it very difficult to adjust, to sort of respond to the students. Um, ‘cause I would like to, if they're really struggling with the question, be able to dig in a little bit more, um, and sort of give that a little bit more time, and on some of the times that was okay and possible. Um, there certainly were other times where that wasn't going to be feasible because I had to get to the next sort of set of content.
Connecting (structural)	But even for those [students] who are reading the book, I think that my job as an instructor is to, uh, put all of this information into a, into a package that's digestible so that they can see how the inferences get drawn, to see analogies from one unit to another one.
Connecting (functional)	But I'm trying to assess, uh, whether [students] can predict reactivity or properties like acidity from molecular structure.

	Uh, and there, the sort of like a sub version of that, that's sort of predicting relative behavior of different structures, so being able to predict how two different mol-, how two different structure will result in two different activities.
Forming	Um, and that, I think, there's sort of a trap in organic chemistry for those students because our content is, the learning objectives are about figuring out which of these principles to be thinking about and then thinking about them properly. And stuff can seem clear when you have the answer, where you really wouldn't be able to derive that answer yourself… learning the process of actually solving the problem is, I think, the most important thing for being successful
Sources	
Instructor	In practice, I think most of [the students] use the lecture as the main source of information, so, you know, as much as I would like for them all to be reading the book, I think that I am a primary source of information for them.
Textbook/online	So I didn't feel like my role was to define what the content was. And also there are very good resources; online textbooks are pretty good. There are lots of places [students] can get kind of that most basic information.
Data	And so I try to, I try to convey the point that this class is such a conceptual one, such a theoretical one that that learning modality [of cramming] is going to fail, and I show [the students] data from previous years to show exactly why this fails.
Justifications	
Correctness	And so then if they get [the question] right, or by and large get it right as a class, um, then I feel like I'm safe to move on [to the next topic]. If not, then that means I devote a little bit more time in the lecture to trying to clarify whatever that specific problem was.

Results and Discussion

In the section that follows, we consider the affordances and limitations of beliefs and resources models of epistemology in describing instructors' views on knowing and learning manifest during our interviews. We begin by briefly describing the instructor epistemologies elicited during the interviews, first using instructor-centered and student-centered descriptors. Then we will summarize the epistemological resources we observed using Chinn et al.'s (2011) multidimensional framework. Then we will consider the extent to which epistemologies embedded in interview dialogue were stable and the implications of treating these epistemologies as hierarchical.

RQ1: What epistemologies do chemistry instructors articulate when talking about their approaches to teaching and learning in undergraduate chemistry?

When we coded our instructors' beliefs as student-centered, instructor-centered, or both, we observed three qualitatively distinct profiles for our three instructors (Fig. 4.1). Approximately three quarters of Mark's beliefs were deemed instructor-centered while the remaining were student-centered. The reverse was observed for James; the vast majority of his beliefs were student-centered while a few were instructor-centered. Liam's beliefs were distributed almost equally among student-centered and instructor-centered. Therefore, if we were to adopt this model of describing instructor thinking, we would label Mark as instructor-centered, James as student-centered, and Liam as transitional.

When we coded for epistemological resources and organized them according to Chinn et al.'s (2011) multidimensional model, we identified several aims, reliable processes, sources, etc. These epistemological resources are summarized, along with examples from our data, in Table 4.3. The epistemic aims expressed by our instructors included memorization, explanations, and problem-solving, along with the value of usefulness. Reliable (or unreliable) processes for achieving these aims included forming (*i.e.*, constructing one's own knowledge based on prior knowledge), accumulating, and connecting. Connecting could be further described based on whether the instructor described how different topics relate (structural) or how causes give rise to effects (functional). These different ways of connecting knowledge were closely related to how the instructors discussed the structure of knowledge in their courses. They referenced "pieces" or "building blocks" of knowledge and articulated how making connections between them could result in more complex knowledge structures. Other times they described how the complexity could be reduced down to a few underlying pieces or fundamental ideas. Sources of knowledge referenced included the instructors themselves, the textbook, and data. We identified correctness as a commonly invoked justification for whether or not an aim had been achieved. Stances toward knowledge and virtues and vices were not observed in our dataset. Although we have described the epistemological resources observed amongst our instructors in aggregate, each of our instructors activated a variety of resources over the course of the interview. We will characterize each instructor's ideas in more detail when we explore the extent to which they were stable in the following section.

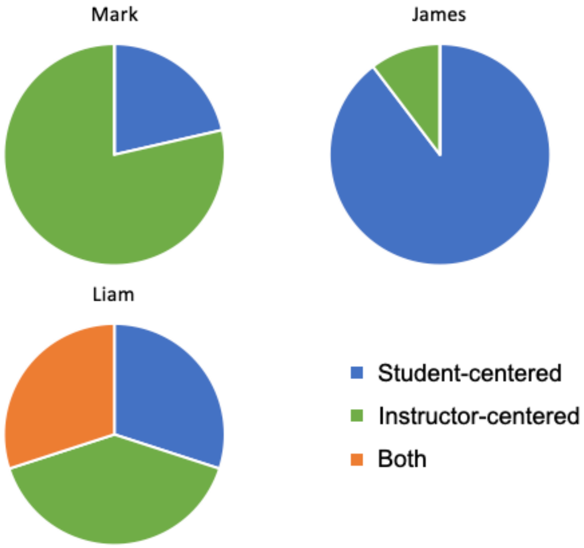


Figure 4.1. Pie charts showing the proportions of student-centered and instructor-centered beliefs expressed by the instructors in this study.

RQ2: What are the affordances and limitations of modeling instructor thinking as beliefs and as epistemological resources?

Throughout our interviews, instructors expressed epistemological ideas relating to course design, lecture practices, assessment strategies, and interactions with students. This enabled us to explore instructor epistemologies across different topics over the span of the interview. For this sort of analysis, we focus on each instructor individually rather than looking across them.

Stability

We begin with James, who repeatedly espoused student-centered beliefs during his interview. James' primary goal for his students is that they engage with the practice of scientific modeling.

> So like the process of learning what a model is, what it applies to, and going through the practice of application of that model to explain an outcome and seeing that those things can be connected is the powerful thing we want our science students to do. Because what it finally does is it gets people thinking scientifically in a meaningful way, in that they understand, "Oh, people have seen data. People have generated models that explain those data. They have then tested those models and refined them over time. And here's the best understanding we have right now. Now, there might need to be some tweaks to that down the road, but this is the best understanding we have right now. And I can take that understanding and apply it to these cases and work out what's likely to happen. And I can then test that with spectroscopy. I can test that with some tool.

In this description of modeling, James positions his students as the constructors of knowledge–a hallmark of student-centered instruction. He wants his students to "think scientifically," "to see that… things can be connected," and "take that understanding and apply it."

When looking across James' interview, he repeatedly expresses this goal for his students' learning. When asked why students should take organic chemistry James' answer mirrors the one above.

> The reason you should take organic chemistry…is that taking organic chemistry, if taught right, will help you understand that we can use some very simple, straightforward models that are accessible to students and to experts and they're the same model… We can use those same models to explain why chemical reactions happen. We can use those same simple models to rationalize why you get a particular regiochemistry or particular stereochemistry, why one product is major and one's minor, why one is seen and one is not observed in the data… That's incredibly empowering use of models to explain outcomes.

James centralizes modeling again when referencing his role as an instructor.

> What we need to be as educators, as teachers, is people who set up the students to have those experiences I just described. We need to be creating environments where somehow students

> are engaged in thinking about models, using models, writing about them to explain why something happens.

Both a beliefs model and a resources model of epistemology work quite well for making sense of James' thinking. James' thinking about teaching appears to be consistently, and stably, student-centered since students are positioned as the modelers. Alternatively, we can state that the epistemic aim of modelling was repeatedly activated by James when reflecting on his course.

However, the other two instructors' interviews demonstrate more instability. Liam said he did not think it was his job to determine the course content and that students could obtain content from a variety of sources.

> We have, although we have some differences in what we teach across the different instructors, we teach a lot of the same reactions and basic principles. So I didn't feel like my role was to define what the content was. And also there are very good resources; online textbooks are pretty good. There are lots of places [the students] can get kind of that most basic information.

Rather, he described his job as follows:

> So my role was one, I guess, make sure [the students] got some of the basic information, so sort of reviewing it a little bit, but more so than that, I thought my role was to show them how to connect concepts.

His primary goal was not to deliver knowledge but to have students connect and use that knowledge. This goal would be considered student-centered. However, moments later in the interview, Liam shared the challenges he experienced with implementing clicker questions in lecture. He said,

> But one of the challenges to doing formative assessment, in my view, is that because we put so much content in the class, I think it, I found it very difficult to adjust, to sort of respond to the students. 'Cause I would like to, if they're really struggling with the question, be able to dig in a little bit more and sort of give that a little bit more time, and on some of the times that was okay and possible. There certainly were other times where that wasn't going to be feasible because I had to get to the next sort of set of content.

Liam felt pressure to—and in fact states that he does—cover the content in lecture. That goal is part of an instructor-centered mindset. As a result of these two different statements, the beliefs model of epistemology might place Liam into a "transitional" or "inconsistent" category of beliefs. Using the lens of a resources model, we would account for the instability by noting that Liam possesses epistemic resources for understanding both himself and outside resources as sources of knowledge and that these were activated at different times.

Liam is not an anomaly in terms of this instability. Mark similarly demonstrates instability in his thinking about teaching and learning. For example, when talking about how students can maximize their learning in his course, Mark defaults to a practical, rather than knowledge-based, perspective.

> And so I try to, I try to convey the point that this class is such a conceptual one, such a theoretical one that that [cramming] is going to fail, and I show them data from previous years to show exactly why this fails... And I also try to tell them that, you know, that they should gear their, their studying around what the assessment is. And so if the assessment is, um, some kind of problem and a certain kinds of problems, then they should be doing those problems and those kinds of problems as part of their studying.

In this quote, learning looks like "time on task;" his focus is on students engaging in activities (*e.g.*, assessments) that he has designed for them, a hallmark of instructor-centered teaching (Simmons et al., 1999; Popova et al., 2020; Luft & Roehrig, 2007). However, when asked why students should take organic chemistry, Mark articulates the following:

> ...the reason that I think everyone ought to take it is that it teaches you how to deal in a more sophisticated way with drawing influences, uh, inferences from data, uh, from using data to support an argument...And so if everybody took organic chemistry, then it would sort of help them to think about how you, um, how you, uh, use data to make informed decisions, which seems like a really important thing just in general.

He believes organic chemistry enables students to make decisions outside of the classroom by teaching them reliable processes characteristic of science such as using data. Further, he positions students as authorities who can interpret data, craft arguments, and make decisions. Making the class "relevant" for students who then engage in substantive intellectual work is a hallmark of student-centered instruction (Popova et al, 2021). Like Liam, these two quotes of Mark's would lead beliefs researchers to characterize him as "inconsistent" or "transitional."

The resources model on the other hand expects this variability and treats it as a feature that can provide insight into Mark's teaching practice rather than a bug. For example, later in the interview Mark describes his approach to writing assessments which he summarizes as consisting of three general types of questions.

> The lowest level [type of question] is simply, you know, if you have some starting molecule, um, what reagents do you use to do some kind of a transformation?
>
> … [The second type of question] is what I call circle-square, um, kinds of questions, which is circle the most acidic compound, square the least, uh, acidic compound, circle the most nucleophilic compound, square the least nucleophilic compound. Right? And so, so these sorts of questions are trying to get students to think through structure-reactivity principles, to get a sense of the character of the compounds.
>
> …And then there are compound, there are, um, uh, questions that put everything together that has people to, uh, essentially explain an observable phenomenon, whether that is showing them a reaction and asking them to propose an arrow-pushing mechanism or giving them a phenomenon and asking them to explain why that phenomenon occurs or to rationalize the outcome. So it is very much along the lines of trying to model what a scientist does, right? If you are given an observable piece of data, how do you use theoretical constructs to rationalize that outcome?

The first two types of questions simply ask students for a claim, whether it's providing the correct reagents or circling the correct molecule. In the last type of question, students are asked to do intellectual work of using data, drawing inferences, and making arguments. What we see here is again instability in his epistemologies; he has resources both for seeing the epistemic aim of knowing facts (*i.e.*, obtaining true beliefs) and the epistemic aim of having a rational model for how the world works. Rather than categorizing Mark as merely "inconsistent," the resources model encourages us to explore the range and depth of his thinking and to view that range as potentially productive for his teaching practice (see below).

Hierarchy

Recall that if we examine the instructors' beliefs in aggregate, we see that James has mostly student-centered beliefs, Mark has mostly instructor-centered beliefs, and Liam has a mix of both. An implication of this analysis might be that James is the best instructor and that interventions are needed to shift Mark and Liam towards more student-centered beliefs.

Using an epistemological resources model of instructor thinking, we would come to a different conclusion. Since activation of epistemological resources is assumed to be context dependent, we would not treat individual resources as "good" or "bad." Rather, we would recognize situations in which they might be more or less productive. For example, consider the following quote from Mark:

> In practice, I think most of [the students] use the lecture as the main source of information, so, you know, as much as I would like for them all to be reading the book, I think that I am a primary source of information for them. But even for those who are reading the book, I think that my job as an instructor is to put all of this information into a package that's digestible so that they can see how the inferences get drawn, to see analogies from one unit to another one.

Since Mark positioned himself as the source of knowledge in the course, he would be described as instructor-centered and therefore less epistemologically sophisticated. Alternatively, we might notice that Mark possessed an epistemological resource for the instructor as a source of knowledge. Depending on the particular information Mark wanted to impart, this resource may be considered productive or unproductive. For example, creating space for students to "figure out" correspondence between features of spectroscopic traces and molecular structure may not be a good use of time. Organic chemists and organic chemistry learners need to be able to effectively analyze and interpret spectroscopic data (Stowe & Cooper, 2019), but they can do so by using skills and rules they are told (*e.g.*, the n+1 rule). The goal of pulling information from spectra is to inform arguments about component(s) of a system under study. It would therefore be better to spend more class time considering consistency between possible claims and spectroscopic evidence rather than, for example, "figuring out" the n+1 rule via numerous pattern matching exercises. From this and related examples, we can conclude that viewing the instructor as a source of knowledge is neither good nor bad but more or less appropriate depending on the particular circumstances.

An epistemological resource model also allows for a much more detailed characterization of instructor ideas, which enables us to recognize the variety of ideas each individual instructor holds, rather than reduce them to a single dimension. Even though James would overall be considered student-centered, some of his beliefs are more instructor-centered. For example, he states "[The students] have to be explaining chemical phenomena using correct models, those models have to be based on core ideas, it all has to tie together, they have to be able to do that on course-wide assessments." The standard of justification conveyed here is correctness, which based on our knowledge of his course, means agreement with scientific canon (*i.e.*, authority). A student-centered approach to justifying models might be consistency with data as judged by the classroom community. By labeling James as student-centered, we might not recognize the aspects of his teaching that could still be improved.

On the other hand, consider Mark, who expressed mostly instructor-centered beliefs. A closer examination reveals some student-centered beliefs. For example, when he articulated how he thinks organic chemistry aids pre-med students, he said:

> And so, especially for these pre-professional students who may never take another science class beyond second semester organic chemistry, this teaches them how you master a complicated topic that demands more than just rote memorization, right? This, it really does kind of teach you that cramming isn't feasible, you do have to understand underlying mechanisms to really succeed in a class like this. And I think that's important, especially for the people who are going on to these higher education where they are going to have to start learning things like medicine, where, you know, simply memorizing a list of, you know,

> characteristics of a disease is much less important than understanding the underlying mechanism. So it really is the same kind of thought process.

In this response, Mark stressed the importance of understanding rather than simply memorizing information. A resources model allows us to attend to these ideas.

The example of Liam arguably provides the most interesting case. Recall that Liam exhibited a mix of student-centered and instructor-centered beliefs. One method of analysis might be to place him in a "transitional" category. But treating the variability as "noise" ignores the interesting tensions Liam himself identified and prevents us from gaining insight into how we could support his teaching. For example, consider this quote from Liam.

> I guess something I do a bad job, I think, of assessing, but I'd like to do a better job of assessing is actually getting some feedback myself about where they're deriving their explanations. So like when they say this would go through SN2, basically how can it be explained, like why did you say SN2, or what sort of factors do you think are at play here? Again, I worry about grading burden.

Because he framed assessment improvement as a feedback tool for himself as the instructor and noted the implication in terms of the grading burden for himself and his TAs, we coded this as instructor-centered. But if we look from the perspective of the resources model, we can infer that Liam was not satisfied with the epistemic aim of correct answers for his students. Rather, he wanted to know that his students understood the "why." Liam's desire to improve assessment practices to gain more insight into students' thinking and extend justifications beyond simply correct claims is an excellent starting point for improving his teaching. In this case, the barriers that might lead him to what a beliefs model would characterize as an instructor centered approach are not epistemic but are instead logistical. A supportive approach for Liam would be to reinforce the aim of understanding for students and provide additional graders to help him assess understanding.

Implications and Conclusion

In this study, we have demonstrated the affordances and constraints of the two models of epistemology. Our findings suggest that each model allows us to capture some aspects of instructor thinking about instruction. However, below we argue for adopting a resources model that shifts away from the standard, tacit epistemology model used in chemistry education in which teachers slowly develop "better" (*i.e.*, more student-centered) beliefs. Doing so (1) allows us to approach instructor learning as constructivists and (2) suggests productive paths toward understanding and, ultimately, influencing instructors' views on knowledge and knowing in-the-moment and across moments.

A Resources Model of Epistemic Cognition Enables Us to Approach Instructor Learning as Constructivists

If we assume instructor beliefs are more-or-less stable across time and place and fall on a continuum from "worse" (*i.e.*, instructor-centered beliefs) to "better" (*i.e.*, student-centered beliefs), a focus on characterizing and improving "bad" beliefs makes a great deal of sense. Indeed, it is common for scholars who assume beliefs are stable to categorize these beliefs via self-report surveys (Gibbons et al., 2018), concept maps (Fletcher & Luft, 2011; Lee, 2019), or interviews (Luft & Roehrig, 2007; Popova et al., 2020), and propose interventions meant to support shifts toward "better" beliefs (Mattheis and Jensen, 2014; Moore et al., 2015; Pelch and McConnell, 2016; Czajka and McConnell, 2019; Fletcher and Luft, 2011; Lee, 2019). Unfortunately, this approach potentially positions instructors as having "wrong" beliefs which require "fixing" and largely ignores potentially productive, if nascent, ways instructors have for thinking about teaching and learning. Taking such an evaluative approach can have unwanted implications for supporting instructors. Specifically, treating their thinking as "wrong" and in need of fixing can elicit defensive behavior from instructors with whom we work, making them unreceptive to our suggestions.

By contrast, if we view instructor epistemic cognition as in-the-moment compilation of small-grain "pieces" related to knowledge and knowing, it becomes clear that no "piece" is inherently "right" or "wrong" (diSessa, 1993; Hammer, 1996; Hammer, 2000; Hammer et al., 2005; Smith III et al., 1994). Instead, clusters of epistemological resources may be more or less productive in progressing toward certain knowledge construction aims in a given moment; the resources model considers the context when assigning value to an idea. By attending to the context, we can avoid labeling instructors' beliefs as good or bad and perpetuating a deficit view. A focus on instructors' epistemological resources allows us to shift toward surfacing potentially productive resources and connections and creating contexts that signal the utility of desirable in-the-moment epistemologies. Stated differently, a resources perspective allows us to approach instructor learning as constructivists (Schafer et al., 2022). In doing so, we place instructors firmly in the role of having some expertise and ways of thinking that contribute to new ways of teaching, putting them in a position of power rather than a position of defensiveness. Our analysis of Liam and Mark in particular points to the "nuggets" of productive epistemologies that we could draw on in professional development to support their learning to teach.

Further, we know that, in principle, instructors possess productive epistemological resources for doing science that they could bring to the classroom. The instructors we interviewed are all practicing scientists with years of experience constructing, revising and communicating evidence-based causal accounts for phenomena they care about. While our interviews suggest that instructors activate some epistemological resources for doing science in the context of teaching, they do not activate others. For example, in a research setting, a model of how and why a reaction occurs is typically evaluated by consistency with experimental data, but in the classroom, such models are typically evaluated by alignment with expert models and deemed “correct” or “incorrect.” We hypothesize that supporting instructors in adopting doing science epistemologies in school contexts could lead to enactment of more authentic, meaningful chemistry learning environments. As a potential first step toward supporting epistemologies for science (Russ, 2014) in the classroom, we advocate for providing opportunities for instructors to reflect on how they approach science and how they approach teaching. Importantly, the goal should be to make use of their experiences as scientists rather than “fix” their teaching.

Attending to How and When Instructors Compile Their Epistemologies in Certain Ways Opens Interesting Paths for Research

A host of intriguing questions come into focus when one adopts a resources perspective on epistemic cognition, including: what leads instructors to compile their epistemologies in a certain way in a given moment? How can we influence those resources instructors (tacitly) see as productive? How does activation of certain epistemologies influence instructor decisions about curriculum and assessment? Seeking answers to these and related questions will allow us to understand mechanisms by which instructor epistemologies evolve, which will in turn support approaches to instructor epistemological learning that surface and build on productive resources. Such approaches would focus on helping instructors identify which epistemological resources are productive in which contexts.

Understanding, and ultimately influencing, instructor epistemologies in-the-moment and across moments is non-trivial. Epistemologies arise from a dynamic and complex system of interactions between people and materials inside and outside of the classroom. Furthermore, epistemologies cannot be understood in terms of discrete levels a person progresses through but rather as in-the-moment confluences of epistemological resources pertaining to aims for knowledge use, processes for achieving aims, sources of knowledge, and justifications for evaluating knowledge. We think the analysis described in this paper is a useful means of characterizing these resources, and the interviews we conducted surfaced some of the specific resources that might be observed. However, more research is needed to understand how instructor epistemologies arise and how they influence the design of course materials and evaluation of student knowledge products. Once we generate a working model of the relationship between instructor epistemologies and the actions they take in the context of teaching, we can study strategies for productively "tipping" instructors toward activating epistemological resources that have the potential to support students in engaging with science for the purpose of making sense of phenomena.

Limitations

We conceive of this work as the beginning of an investigation into instructor epistemologies, and there is still much we plan to explore. The data we analyzed were collected through interviews and therefore are filtered through the perceptions of the instructors. We do not know the extent to which the epistemologies elicited through our interview protocol align with the epistemologies which shape in-the-moment instructional or assessment decisions. We would need to observe instructor's behavior as they talk to students or grade exams in order to infer these epistemologies. Furthermore, we do not claim that the epistemologies we have identified are representative of chemistry instructors everywhere nor do we claim to have uncovered all the epistemologies for teaching and learning that the interviewed instructors possess. Rather, we offer this analysis as an illustration of how one might elicit and characterize instructor ideas.

Acknowledgements

The authors wish to thank the organic instructors who generously shared their time and expertise with us. This study was supported by the U.S. National Science Foundation under Grant No. DUE-2225025. Funding was also provided by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison with funding from the Wisconsin Alumni Research Foundation.

References

Abd-El-Khalick, F., Bell, R. L., & Lederman, N. G. (1998). The nature of science and instructional practice: Making the unnatural natural. Science Education, 82(4), 417–436. https://doi.org/10.1002/(SICI)1098-237X(199807)82:4<417::AID-SCE1>3.0.CO;2-E

Abell, S. K. (2008). Twenty years later: Does pedagogical content knowledge remain a useful idea?. International Journal of Science Education, 30(10), 1405–1416. https://doi.org/10.1080/09500690802187041

Adams, W. K., Perkins, K. K., Podolefsky, N. S., Dubson, M., Finkelstein, N. D., & Wieman, C. E. (2006). New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey. Physical Review Special Topics - Physics Education Research, 2(1), 010101. https://doi.org/10.1103/PhysRevSTPER.2.010101

Alshenqeeti, H. (2014). Interviewing as a data collection method: A critical review. English Linguistics Research, 3(1), 39–45. http://dx.doi.org/10.5430/elr.v3nlp39

Avraamidou, L., & Zembal-Saul, C. (2010). In search of well-started beginning science teachers: Insights from two first-year elementary teachers. Journal of Research in Science Teaching, 47(6), 661–686. https://doi.org/10.1002/tea.20359

Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. J. (2016). Epistemologies in practice: Making scientific practices meaningful for students. Journal of Research in Science Teaching, 53(7), 1082–1112. https://doi.org/10.1002/tea.21257

Chari, D. N., Nguyen, H. D., Zollman, D. A., & Sayre, E. C. (2019). Student and instructor framing in upper-division physics. American Journal of Physics, 87(11), 875–884. https://doi.org/10.1119/1.5120392

Chen, Y., Irving, P. W., & Sayre, E. C. (2013). Epistemic game for answer making in learning about hydrostatics. Physical Review Special Topics - Physics Education Research, 9(1), 010108. https://doi.org/10.1103/PhysRevSTPER.9.010108

Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.). (1988). The nature of expertise. Psychology Press. https://doi.org/10.4324/9781315799681

Chinn, C. A., Buckland, L. A., & Samarapungavan, A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. Educational Psychologist, 46(3), 141–167. https://doi.org/10.1080/00461520.2011.587722

Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.) Handbook on research in teaching (pp. 255–296). Macmillan Publishing Co

Collins, A., & Ferguson, W. (1993). Epistemic forms and epistemic games: Structures and strategies to guide inquiry. Educational Psychologist, 28(1), 25–42. https://doi.org/10.1207/s15326985ep2801_3

Creswell, J. W. (2007). Qualitative Inquiry and Research Design: Choosing Among Five Approaches. SAGE Publications.

Cronin-Jones, L. L. (1991). Science teacher beliefs and their influence on curriculum implementation: Two case studies. Journal of Research in Science Teaching, 28(3), 235–250. https://doi.org/10.1002/tea.3660280305

Czajika, C. D., & McConnell, D. (2019). The adoption of student-centered teaching materials as a professional development experience for college faculty. International Journal of Science Education, 41(5), 693–711. https://doi.org/10.1080/09500693.2019.1578908

Davis, E. (2003). Untangling dimensions of middle school students' beliefs about scientific knowledge and science learning. International Journal of Science Education, 25(4), 439–468. https://doi.org/10.1080/09500690210145765

Dini, V., & Hammer, D. (2017). Case study of a successful learner's epistemological framings of quantum mechanics. Physical Review Physics Education Research, 13(1), 010124. https://doi.org/10.1103/PhysRevPhysEducRes.13.010124

diSessa, A. A. (1993). Toward an epistemology of physics. Cognition and Instruction, 10(2–3), 105–225. https://doi.org/10.1080/07370008.1985.9649008

Elby, A., & Hammer, D. (2001). On the substance of a sophisticated epistemology. Science Education, 85(5), 554–567. https://doi.org/10.1002/sce.1023

Fang, Z. (1996). A review of research on teacher beliefs and practices. Educational Research, 38(1), 47–64. https://doi.org/10.1080/0013188960380104

Fletcher, S. S., & Luft, J. A. (2011). Early career secondary science teachers: A longitudinal study of beliefs in relation to field experiences. Science Education, 95(6), 1124–1146. https://doi.org/10.1002/sce.20450

Gibbons, R. E., Villafañe, S. M., Stains, M., Murphy, K. L., & Raker, J. R. (2018). Beliefs about learning and enacted instructional practices: An investigation in postsecondary chemistry education. Journal of Research in Science Teaching, 55(8), 1111–1133. https://doi.org/10.1002/tea.21444

Glaser, B., & Strauss, A. (1967). The Discovery of Grounded Theory. Aldine Publishing Company.

Greene, J. A., Cartiff, B. M., & Duke, R. F. (2018). A meta-analytic review of the relationship between epistemic cognition and academic achievement. Journal of Educational Psychology, 110(8), 1084–1111. https://doi/10.1037/edu0000263

Greene, J. A., Sandoval, W. A., & Bråten, I. (2016). An introduction to epistemic cognition. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), Handbook of Epistemic Cognition (pp. 1-15). Routledge. https://doi.org/10.4324/9781315795225

Hammer, D. (1996). Misconceptions or p-prims: How may alternative perspectives of cognitive structure influence instructional perceptions and intentions. Journal of the Learning Sciences, 5(2), 97–127. https://doi.org/10.1207/s15327809jls0302_1

Hammer, D. (2000). Student resources for learning introductory physics. American Journal of Physics, 68(S1), S52–S59. https://doi.org/10.1119/1.19520

Hammer, D., & Elby, A. (2002). On the form of a personal epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), Personal epistemology: The psychology of beliefs about knowledge and knowing (pp. 169–190). Lawrence Erlbaum Associates Publishers.

Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. In J. P. Mestre (Ed.), Transfer of Learning from a Modern Multidisciplinary Perspective (pp. 89–119). Information Age Publishing.

Haney, J. J., Czerniak, C. M., & Lumpe, A. T. (1996). Teacher beliefs and intentions regarding the implementation of science education reform strands. Journal of Research in Science Teaching, 33(9), 971–993. https://doi.org/10.1002/(SICI)1098-2736(199611)33:9<971::AID-TEA2>3.0.CI;2-S

Hashweh, M. Z. (1996). Effects of science teachers’ epistemological beliefs in teaching. Journal of Research in Science Teaching, 33(1), 47–63. https://doi.org/10.1002/(SICI)1098-2736(199601)33:1<47::AID-TEA3>3.0.CO;2-P

Havdala, R., & Ashkenazi, G. (2007). Coordination of theory and evidence: Effect of epistemological theories on students’ laboratory practice. Journal of Research in Science Teaching, 44(8), 1134–1159. https://doi.org/10.1002/tea.20215

Hewson, P. W., & Hewson, M. G. A. (1984). The role of conceptual conflict in conceptual change and the design of science instruction. Instructional Science, 13(1), 1–13. https://doi.org/10.1007/BF00051837

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. Review of Educational Research, 67(1), 88–140. https://doi.org/10.3102/00346543067001088

Hora, M. T. (2014). Exploring faculty beliefs about student learning and their role in instructional decision-making. The Review of Higher Education, 18(1), 37–70. https://doi.org/10.1353/rhe.2014.0047

Irving, P. W., Martinuk, M. S., & Sayre, E. C. (2013). Transitions in students’ epistemic framing along two axes. Physical Review Special Topics - Physics Education Research, 9(1), 010111. https://doi.org/10.1103/PhysRevSTPER.9.010111

Kagan, D. M. (1992). Implication of research on teacher belief. Educational Psychologist, 27(1), 65–90. https://doi.org/10.1207/s15326985ep2701_6

King, P. M., & Kitchener, K. S. (1994). Developing reflective judgement: Understanding and promoting intellectual growth and critical thinking in adolescents and adults. Jossey-Bass.

Kittleson, J. M. (2011). Epistemological beliefs of third-grade students in an investigation-rich classroom. Science Education, 95(6), 1026–1048. https://doi.org/10.1002/sce.20457

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

Lederman, N. G. (1999). Teachers' understanding of the nature of science and classroom practice: Factors that facilitate or impede the relationship. Journal of Research in Science Teaching, 36(8), 916–929. https://doi.org/10.1002/(SICI)1098-2736(199910)36:8<916::AID-TEA2>3.0.CO;2-A

Lee, S. W. (2019) The impact of a pedagogy course on the teaching beliefs of inexperienced graduate teaching assistants. CBE—Life Sciences Education, 18(1), 18:ar5, 1–18:ar5, 12. https://doi.org/10.1187/cbe.18-07-0137

Liang, J., & Tsai, C. (2010). Relational analysis of college science-major students' epistemological beliefs toward science and conceptions of learning science. International Journal of Science Education, 32(17), 2273–2289. https://doi.org/10.1080/09500690903397796

Lising, L., & Elby, A. (2005). The impact of epistemology on learning: A case study from introductory physics. American Journal of Physics, 73(4), 372. https://doi.org/10.1119/1.1848115

Loughran, J., Mulhall, P., & Berry, A. (2004). In search of pedagogical content knowledge in science: Developing ways of articulating and documenting professional practice. Journal of Research in Science Teaching, 41(4), 370–391. https://doi.org/10.1002/tea.20007

Luft, J. A., & Roehrig, G. H. (2004). Inquiry teaching in high school chemistry classrooms: The role of knowledge and beliefs. Journal of Chemical Education, 81(10), 1510–1516. https://doi.org/10.1021/ed081p1510

Luft, J. A., & Roehrig, G. H. (2007). Capturing science teachers' epistemological beliefs: The development of the teacher beliefs interview. The Electronic Journal for Research in Science & Mathematics Education, 11(2), 38–63.

Mack, M. R., & Towns, M. H. (2016). Faculty beliefs about the purposes for teaching undergraduate physical chemistry courses. Chemistry Education Research and Practice, 17(1), 80–99. https://doi.org/10.1039/C5RP00148J

Mansour, N. (2009). Science teachers' beliefs and practices: Issues, implications and research agenda. International Journal of Environmental and Science Education, 4(1), 25–48.

Mattheis, A., & Jensen, M. (2014). Fostering improved anatomy and physiology instructor pedagogy. Advances in Physiology Education, 38(4), 321–329. http://doi.org/10.1152/advan.00061.2014

McCloskey, M. (1983). Naive theories of motion. In D. Gentner & A. L. Stevens (Eds.), Mental Models (pp. 299–323). Lawrence Erlbaum Associates, Inc.

Minstrell, J. (1989). Teaching science for understanding. In L. B. O. Resnick & L. Klopfer (Eds.), Toward the thinking curriculum: Current cognitive research (pp. 133–149). Association of Super-vision and Curriculum Development.

Modir, B., Thompson, J. D., & Sayre, E. C. (2017). Students' epistemological framing in quantum mechanics problem solving. Physical Review Physics Education Research, 13(2), 020108. https://doi.org/10.1103/PhysRevPhysEducRes.13.020108

Moore, T. J., Guzey, S. S., Roehrig, G. H., Stohlmann, M., Park, M. S., Kim, Y. R., Callender, H. L., & Teo, H. J. (2015). Changes in faculty members' instructional beliefs while implementing model-eliciting activities. Journal of Engineering Education, 104(3), 279–302. https://doi.org/10.1002/jee.20081

Mutambuki, J., & Fynewever, H. (2012). Comparing chemistry faculty beliefs about grading with grading practices. Journal of Chemical Education, 89(3), 326–334. https://doi.org/10.1021/ed1000284

NGSS Lead States. (2013). Next generation science standards: For states, by states. The National Academies Press.

Odden, T. Ole. B., & Russ, R. S. (2018). Sensemaking epistemic game: A model of student sensemaking processes in introductory physics. Physical Review Physics Education Research, 14(2), 020122. https://doi.org/10.1103/PhysRevPhysEducRes.14.020122

Oliveira, A. W., Akerson, V. L., Colak, H., Pongsanon, K., & Genel, A. (2012). The implicit communication of nature of science and epistemology during inquiry discussion. Science Education, 96(4), 652–684. https://doi.org/10.1002/sce.21005

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. Review of Educational Research, 62(3), 307–332. https://doi.org/10.3102/00346543062003307

Pelch, M. A., & McConnell, D. A. (2016). Challenging instructors to change: A mixed methods investigation on the effects of material development on the pedagogical beliefs of geoscience instructors. International Journal of STEM Education, 3(5), 1–18. https://doi.org/10.1186/s40594-016-0039-y

Perry, W. G. (1970). Forms of intellectual and ethical development in the college years: A scheme. Holt, Rinehart, and Winston.

Philip, T. M. (2011). An “ideology in pieces” approach to studying change in teachers’ sensemaking about race, racism, and racial justice. Cognition and Instruction, 29(3), 297–329. https://doi.org/10.1080/07370008.2011.583369

Piaget, J. (1969). Psychology of intelligence, Littlefield Adams.

Piaget, J. (1970). Piaget’s theory. In P. Mussen (Ed.), Carmichael’s Handbook of child psychology (pp. 703–732), Wiley.

Popova, M., Kraft, A., Harshman, J., & Stains, M. (2021). Changes in teaching beliefs of early-career chemistry faculty: A longitudinal investigation. Chemistry Education Research and Practice, 22(2), 431–442. https://doi.org/10.1039/D0RP00313A

Popova, M., Shi, L., Harshman, J., Kraft, A., & Stains, M. (2020). Untangling a complex relationship: Teaching beliefs and instructional practices of assistant chemistry faculty at research-intensive institutions. Chemistry Education Research and Practice, 21(2), 513–527. https://doi.org/10.1039/C9RP00217K

Remillard, J. T. (2005). Examining key concepts in research on teachers’ use of mathematics curricula. Review of Educational Research, 75(2), 211–246. https://doi.org/10.3102/00346543075002211

Roehrig, G. H., Kruse, R. A., & Kern, A. (2007). Teacher and school characteristics and their influence on curriculum implementation. Journal of Research in Science Teaching, 44(7), 883–907. https://doi.org//10.1002/tea.20180

Rosenberg, S., Hammer, D., & Phelan, J. (2006). Multiple epistemological coherences in an eighth-grade discussion of the rock cycle. The Journal of the Learning Sciences, 15(2), 261–292. https://doi.org/10.1207/s15327809jls1502_4

Rupnow, R. L., LaDue, N. D., James, N. M., & Bergan-Roller, H. E. (2020). A perturbed system: How tenured faculty responded to the COVID-19 shift to remote instruction. Journal of Chemical Education, 97(9), 2397–2407. https://doi.org/10.1021/acs.jchemed.0c00802

Russ, R. S. (2014). Epistemology of science vs. epistemology for science. Science Education, 98(3), 388–396. https://doi.org/10.1002/sce.21106

Russ, R. S. (2018). Characterizing teacher attention to student thinking: A role for epistemological messages. Journal of Research in Science Teaching, 55(1), 94–120. https://doi.org/10.1002/tea.21414

Russ, R. S., & Luna, M. J. (2013). Inferring teacher epistemological framing from local patterns in teacher noticing. Journal of Research in Science Teaching, 50(3), 284–314. https://doi.org/10.1002/tea.21063

Russ, R. S., Sherin, B. L., & Sherin, M. G. (2016). What constitutes teacher learning? In D. H. Gitomer & C. A. Bell (Eds.), Handbook of Research on Teaching (5th ed., pp. 391–438). American Educational Research Association.

Sandoval, W. A. (2005). Understanding students' practical epistemologies and their influence on learning through inquiry. Science Education, 89, 634–656. https://doi.org/10.1002/sce.20065

Schafer, A. G. L., Kuborn, T. M., Schwarz, C. E., Deshaye, M. Y., & Stowe, R. L. (2022). Messages about valued knowledge products and processes embedded within a suite of transformed high school chemistry curricular materials. Chemistry Education Research and Practice, advance article. https://doi.org/10.1039/D2RP00124A

Scherr, R. E., & Hammer, D. (2009). Student behavior and epistemological framing: Examples from collaborative active-learning activities in physics. Cognition and Instruction, 27(2), 147–174. https://doi.org/10.1080/07370000902797379

Schoenfeld, A. H. (1998). Toward a theory of teaching-in-context. Issues in Education, 4(1), 1–94. https://doi.org/10.1016/S1080-9724(99)80076-7

Schommer-Aikins, M. (2004). Explaining the epistemological belief system: Introducing the embedded systemic model and coordinated research approach. Educational Psychologist, 39(1), 19–29. https://doi.org/10.1207/s15326985ep3901_3

Sevian, H., & Couture, S. (2018). Epistemic games in substance characterization. Chemistry Education Research and Practice, 19(4), 1029–1054. https://doi.org/10.1039/C8RP00047F

Sherin, B. (2006). Common sense clarified: The role of intuitive knowledge in physics problem solving. Journal of Research in Science Teaching, 43(6), 535–555. https://doi.org/10.1002/tea.20136

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. Educational Researcher, 15(2), 4–14. https://doi.org/10.3102/0013189X015002004

Simmons, P. E., Emory, A., Carter, T., Coker, T., Finnegan, B., Crockett, D., Richardson, L., Yager, R., Craven, J., Tillotson, J., Brunkhorst, H., Twiest, M., Hossain, K., Gallagher, J., Duggan-Haas, D., Parker, J., Cajas, F., Alshannag, Q., McGlamery, S., … Labuda, K. (1999). Beginning teachers: Beliefs and classroom actions. Journal of Research in Science Teaching, 36(8), 930–954. https://doi.org/10.1002/(SICI)1098-2736(199910)36:8<930::AID-TEA3>3.0.CO;2-N

Smith, C. L., & Wenk, L. (2006). Relations among three aspects of first-year college students' epistemologies of science. Journal of Research in Science Teaching, 43(8), 747–785. https://doi.org/10.1002/tea.20113

Smith III, J. P., diSessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. Journal of the Learning Sciences, 3(2), 115–163. https://doi.org/10.1207/s15327809jls0302_1

Stains, M., Pilarz, M., & Chakraverty, D. (2015). Short and long-term impacts of the Cottrell Scholars Collaborative New Faculty Workshop. Journal of Chemical Education, 92(9), 1466–1476. https://doi.org/10.1021/acs.jchemed.5b00324

Stowe, R. L., & Cooper, M. M. (2019). Arguing from spectroscopic evidence. Journal of Chemical Education, 96(10), 2072–2085. https://doi.org/10.1021/acs.jchemed.9b00550

Strike, K. A., & Posner, G. J. (1985). A conceptual change view of learning and understanding. In L. H. T. West & A. L. Pines (Eds.), Cognitive structure and conceptual change (pp. 211–231). Academic.

Szozda, A. R., Bruyere, K., Lee, H., Mahaffy, P. G., & Flynn, A. B. (2022). Investigating educators' perspectives toward systems thinking in chemistry education from international contexts. Journal of Chemical Education, 99(7), 2474–2483. https://doi.org/10.1021/acs.jchemed.2c00138

Tuminaro, J., & Redish, E. F. (2007). Elements of a cognitive model of physics problem solving: Epistemic games. Physical Review Special Topics - Physics Education Research, 3(2), 020101. https://doi.org/10.1103/PhysRevSTPER.3.020101

Wallace, C. S., & Kang, N. (2004). An investigation of experienced secondary science teachers' beliefs about inquiry: An examination of competing belief sets. Journal of Research in Science Teaching, 41, 936–960. https://doi.org/10.1002/tea.20032

Wendell, K. B., Swenson, J. E. S., & Dalvi, T. S. (2019). Epistemological framing and novice elementary teachers' approaches to learning and teaching engineering design. Journal of Research in Science Teaching, 56(7), 956–982. https://doi.org/10.1002/tea.21541

Appendix A

Sensemaking Tipping

This work was conducted in collaboration with Ryan L. Stowe.

This work is currently in progress. Preliminary analysis and results are presented.

Abstract

Engaging students in the process of doing science is thought to help them learn scientific knowledge and gain a more authentic understanding of how scientific knowledge is developed. However, students can participate in scientific practices without seeing them as valuable ways of figuring out the world around them. To support students in sensemaking—using mechanistic reasoning to figure out how a phenomenon arises because of a desire to understand—we need a better understanding of what factors influence how they approach opportunities to reason about phenomena, especially those that relate to chemistry. To that end, we interviewed pairs of students and asked them to decide on the best yeast substitute for baking. The preliminary analysis is presented here. We observed that students tended to answer quickly without spending much time discussing how the yeast substitutes work, but if given enough time and prompting, most transitioned into sensemaking for at least part of the interview. We also observed how interactions between students influenced their approaches to answering the interview questions. Interestingly, we found that students did not employ much mechanistic reasoning at the molecular level in their discussions but focused more on brainstorming ideas to test empirically. The future directions for this work and how they connect to other ongoing work on epistemic messaging are discussed.

Introduction

Science education reform efforts have prioritized engaging students in *doing* science rather than merely learning *about* science. *A Framework for K-12 Science Education: Practices,* *Crosscutting Concepts, and Core Ideas*, upon which the Next Generation Science Standards were built, describes eight scientific practices that students should engage in as they learn science (National Research Council, 2012). The *Framework* articulates three reasons for emphasizing practices: they help students understand how scientific knowledge develops, improve their understanding of scientific content, and they can help generate interest in science. The *Framework* states, “Students cannot comprehend scientific practices, nor fully appreciate the nature of scientific knowledge itself, without directly experiencing those practices for themselves” (p. 30).

Embedded in these arguments is the assumption that by participating in scientific practices, students will understand their purposes and place value on them in the same way that scientists do. However, a student might engage in data analysis or scientific argumentation not in order to figure out a puzzling phenomenon, as a scientist would, but in order to earn points or please their teacher. Berland and Hammer (2012) observed some of this performative engagement, which they termed “pseudoargumentation,” among a class of sixth graders and argued for greater attention to how instructors frame activities in which students participate in scientific practices.

In contrast to pseudoargumentation, *sensemaking* describes moments where students are engaged in figuring out how or why a phenomenon occurs and are motivated to do so by their own interest (Odden & Russ, 2019a). Studies on sensemaking are more prevalent in the K–12 literature, but a notable example of sensemaking in higher education was reported by Odden and Russ (2019b). They interviewed pairs of students from a physics course and asked them to determine whether or not it is safe to get out of a car that has been struck by lightning. They observed that reoccurring “vexing questions”—questions in which students seemed frustrated or irritated by the lack of knowing—seemed to promote or stabilize sensemaking. They suggested that these questions serve to draw students’ attention (back) to what they needed to figure out.

A promising example of sensemaking with undergraduate students in the context of chemistry was reported by Hunter et al. (2021). Using a POGIL-style activity, they prompted groups of introductory chemistry students to discuss the controversy surrounding deflated footballs at an extremely cold professional football game and, through a series of guiding questions, construct the Ideal Gas Law and use it to argue whether the deflation was intentional or caused by temperature. Using an adaption of the sensemaking epistemic game framework (Odden & Russ, 2018; Collins & Ferguson 1993), the authors obtained evidence that some of the groups engaged in sensemaking while other groups' dialogue was more consistent with an "answer-making game."

A few other examples of sensemaking in chemistry have been reported. Haraldsrud and Odden (2024) also studied sensemaking related to behavior of gases but with students from a physical chemistry course. They explored the role of computer stimulations in supporting sensemaking and found that stimulations can initiate or sustain sensemaking by giving students the ability to explore a phenomenon by adjusting parameters, engaging them with immediate feedback, aiding them with visualization, and producing results that contradict students' prior experiences. In a study that encompassed students from middle school to upper-level undergraduate chemistry courses, Sevian and Couture characterized the epistemic games students used to identify substances (2018). They observed two general approaches to the questions posed by the interviewer: one consistent with what Jiménez-Aleixandre et al. (2000) call "doing science" and one consistent with "doing the lesson."

These studies demonstrate that sensemaking in the context of chemistry is possible, but many questions remain. What chemistry phenomena are interesting to students and complex enough to challenge them intellectually but be at least partially understandable with the knowledge and resources they have available to them? In what ways can the instructor promote sensemaking?

What impact does opportunities for sensemaking have on students' learning, their understanding of the nature of science, their sense of belonging in science, or other student outcome measures of interest? In this study, we seek to address some of these questions in order to gain insight into how we can design more meaningful and authentic chemistry classes.

Theoretical Framework

This study employs the model of epistemology articulated by Hammer and Elby (2002). They proposed that a person's epistemic cognition consists of epistemic resources—small-grained ideas related to knowledge and knowing—that are activated and connected in the moment. The specific resources that are activated may change in response to new cues, such as a comment from another person or a change in setting. In a classroom setting, a conversation with a teacher can move students away from focusing on what scientific vocabulary to use in their explanation toward what ideas make sense to them (Rosenberg et al., 2006). Some resources may be co-activated repeatedly in a given situation, leading to a relatively consistent epistemic response. For example, most students walk into lecture halls expecting to acquire knowledge by listening to the professor because this is consistent with their past experiences in school. The resources model accounts for both observations of rapid changes in epistemic cognition (via shifts in resource activation) and consistencies in epistemic cognition (via repeated use of a set of resources).

The concept of framing provides a useful means of characterizing a person's epistemology at a larger grain size than epistemic dimensions or epistemic resources. A frame is one's sense of what is going on in a situation (Hammer et al., 2005; Tannen, 1993; MacLachlan & Reid, 1994). For example, a chemistry graduate student working in a lab might try to propose a mechanism that is consistent with the data they have collected in order to explain an unexpected reaction outcome they obtained. In their class, that same student may engage in the same activity (proposing a reaction mechanism) but with the goal of getting the right answer on their exam. In both scenarios, the person is engaged in the same scientific practice, but the epistemologies guiding their action are quite different. The former may be described as an inquiry frame, in which the goal is to figure something out, while the latter could be described as an oral examination frame, in which the goal is to produce the right answer (Russ et al., 2012). We conceive of frames as composed of epistemic resources and like epistemic resources, frames are not necessarily stable and may last only a few minutes (Hammer et al., 2005). In our analysis, we primarily attended to students' frames, but we sometimes noted smaller-grained epistemic resources, such as those pertaining to source of knowledge, when they were helpful in providing evidence for identifying the larger frame.

The term "sensemaking" has been used in various ways in the education literature, sometimes as a cognitive process, sometimes as a discourse practice, and sometimes as a frame. Here, we use the definition offered by Odden and Russ (2019a), which encompasses all three aspects. They define sensemaking as "a dynamic process of building or revising an explanation in order to 'figure something out'—to ascertain the mechanism underlying a phenomenon in order to resolve a gap or inconsistency in one's understanding" (p. 191–192). It is important to note that there is a clear epistemic aim in this definition—the goal of sensemaking is to "resolve a gap or inconsistency in one's understanding." Thus, attempting to create an explanation in order to satisfy the instructor would not be considered sensemaking.

Research Question

Inspired by the study done by Odden and Russ (2019b), we wanted to explore sensemaking in the context of a chemistry-related phenomenon. We designed a study to address the following research question: What initiates or sustains a sensemaking frame and what discourages a sensemaking frame when students are asked to reason about a chemical phenomenon from everyday life?

Methods

Participants

We decided to interview students in pairs rather than individually in order to encourage more dialogue, as recommended by prior research on sensemaking (Odden & Russ, 2019b). Pairs of students were recruited from an introductory organic chemistry course at a large research-intensive university. Six pairs of students were interviewed. Five pairs consisted of two women while the other consisted of two men. IRB approval was obtained, and written consent from each student was obtained before the interviews began.

Interview Protocol

Because we expected it to be difficult to induce or sustain a sensemaking frame (Lemke, 1990), we designed the interview to promote sensemaking as much as we could. First, we needed to identify a suitable phenomenon. Given that we wanted to stimulate sensemaking in the context of chemistry, we needed a phenomenon that could be potentially explained by invoking interactions at the particle level. Furthermore, we hypothesized that because sensemaking requires students to be motivated to figure something out, the phenomenon needed to be potentially relatable and interesting to students. These constraints led us to consider baking, which we thought most students would have some experience with. We ultimately chose the phenomenon of yeast substitutes for making dough rise. As it consumes sugar, yeast produces carbon dioxide, which aerates and raises the dough. Yeast substitutes, such as vinegar mixed with baking soda, also produce carbon dioxide but through an acid-base reaction. Having taken general chemistry and some organic chemistry, we expected students to be familiar with thinking through acid-base reactions.

Before beginning the interviews, the first author explained to students that the goal was to gain insight into their ideas and their thought processes rather than to evaluate their answers. To signal to students that their everyday experiences were of interest and an appropriate source of knowledge to use, the first questions concerned students' experiences with baking, such as whether they liked to bake, what they liked to bake, whether they liked baking shows, etc. Throughout the interview, the interviewer used phrases such as "I don't know the answers either" and "Just explain it in a way that makes sense to you."

The topic of yeast substitutes was introduced to the students through an anecdotal story from the interviewer. She shared an experience in which her homemade cinnamon rolls did not rise, which she attributed to expired yeast. She then shared the internet suggestion of using baking soda and vinegar as a substitute and asked students if they thought it would be effective. To support their sensemaking, baking soda and vinegar were brought to the interview, and students were offered the opportunity to mix the two substances together and observe the bubbling that occurred (Fig. A.1). In subsequent interviews, lemon juice was brought in to mix with baking soda as well, and we found that prompting students to compare the two combinations and choose the better one (however they defined better) elicited more discussion than asking them to reason about vinegar and baking soda alone. As needed, the interviewer also asked questions like, "What other yeast substitutes might you try?" and "What do you think is in the bubbles?"

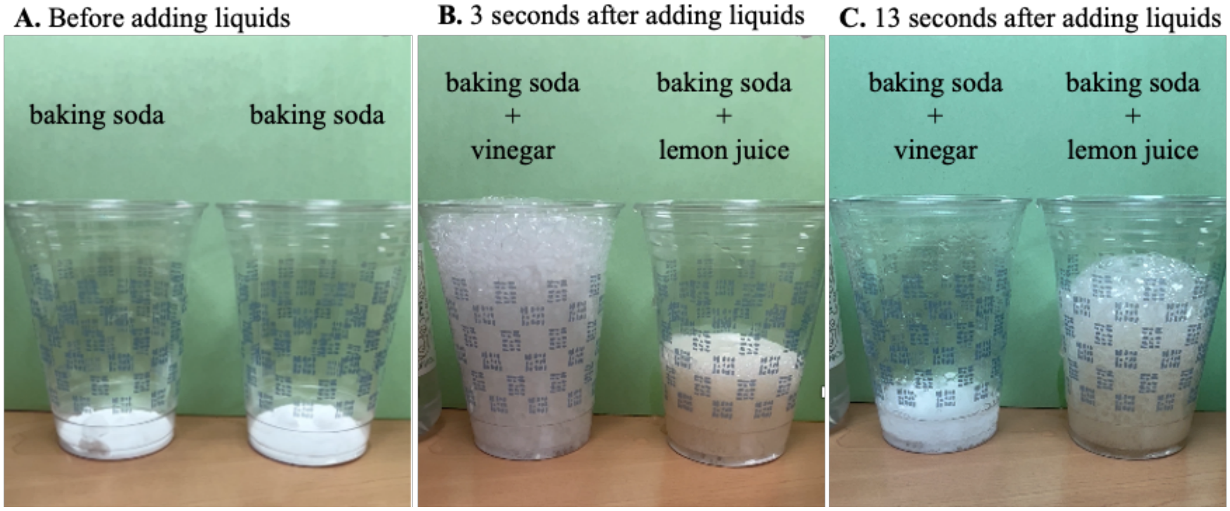


Figure A.1. Reactions of baking soda with vinegar (left cup) and lemon juice (right cup). **A** shows the two cups with baking soda before the liquids were added. **B** shows the two reactions a few seconds after the liquids were added simultaneously. **C** shows the two reactions after approximately 13 seconds had passed.

Data Analysis

The interviews were transcribed by an automatic transcription service, Temi, and edited by the first author for accuracy. Special attention was paid to capturing filler words (*e.g.*, um, uh) and pauses, both of which can be indicative of a person's frame (Tannen, 1993). Any identifying information was removed, and participants were given pseudonyms.

The transcripts were analyzed in the qualitative analysis program NVivo. In the first pass of analysis, the first author noted any instances that seemed to fit into one of the frames described by Russ et al. (2012). These included an oral examination frame, an expert interview frame, a brainstorming frame, and a sensemaking frame (Table A.1). In addition, the first author looked for moments where there seemed to be shifts in frame and moments where each student in a pair seemed to be adopting different, somewhat conflicting frames. The first author also noted explicit mentions of where students were drawing their knowledge from (*e.g.*, class, past experience) since we hypothesized these would be useful in inferring student epistemologies.

Table A.1. Descriptions and behavioral markers for oral examination, expert interview, brainstorming, and sensemaking frames, as described by Odden and Russ (2019b).

Frame	**Description**	**Dialogic Markers**
Oral Examination Frame	Students see their role as producing the correct responses to the interview questions, as evaluated by the interviewer.	• Lack of hedging language • Eye contact • Limited use of gesture • Use of scientific vocabulary
Expert Interview Frame	Students view their role as explaining their own thinking in a way that helps the interview understand their reasoning.	• Lack of hesitation • Eye contact • Prolific gesturing • Use of colloquial terminology
Brainstorming Frame	Students perceive the need to recall prior knowledge in order to figure out an explanation in the moment.	• Long pauses in speech • Restarts during explanations • Little eye contact • Prolific gesturing
Sensemaking Frame	Students work together to construct an explanation to a gap or inconsistency in their knowledge that they are motivated to resolve.	• Argumentation • Long pauses in speech • Restarts during explanations • Prolific gesturing • Vexing questions

The next step of analysis will involve formalizing a coding scheme for identifying students' frames. The descriptions and markers of various frames offered by Odden and Russ (2019b) will serve as the starting point for coding (Table A.1). If patterns of student dialogue do not fit into the categories described in Table A.1, inductive coding will be used to expand upon the initial coding scheme. To establish reliability, a second researcher will review the codebook and apply it to a subset of the data. Cohen's kappa will be used to measure interrater agreement, and rounds of coding will continue until a value of 0.7 or higher for Cohen's kappa is obtained (Cohen, 1960). After each round, coders will discuss any disagreements in the coding and update the codebook as needed.

Once the interview transcripts have been coded for students' frames, analysis will focus on determining the factors that may have led to shifts in framing. For example, Odden and Russ (2019b) saw that vexing questions served to initiate sensemaking frames in their interviews. We will look for vexing questions and other factors, such as comments from the interviewer signaling the type of response expected or the sources of knowledge they may use. Through a constant comparative approach (Glaser & Strauss, 1967), we will develop a coding scheme for the data. Reliability will be established through the same process described for the framing analysis.

Preliminary Results and Discussion

The initial analysis found evidence for oral examination and sensemaking frames, as well as instances where students seemed to approach the situation from different frames. Most students made a decision on which yeast substitute was better relatively quickly into the interview, usually only a few minutes after mixing the ingredients and discussing. This behavior is consistent with an oral examination frame. Sensemaking was observed during later portions of the interviews, but in general, students did not engage in mechanistic reasoning at the molecular level while in a sensemaking frame. These themes are discussed in more detail below with excerpts from the interviews.

Default Framing: Oral Examination

When prompted to consider which yeast substitute was better, most students offered an answer quite quickly, usually only a few minutes after mixing together the ingredients. For example, consider the excerpt shown in Table A.2. This conversation occurred immediately after the two participants, Jenn and Victoria, had mixed together the baking soda and acids. Jenn began by observing the bubbles (line 1), and Victoria jumped in to try to identify the gas within the bubbles (line 2). This cued Jenn into thinking about what she learned in her microbiology class, and she attempted to identify which type of microbe yeast is so that she could figure out what the gas is (lines 3 and 5). She drew a parallel between protists excreting “something” to grow and the reactions releasing carbon dioxide, but she was hesitant and looked to Victoria for affirmation, asking “Right?” (line 5). Victoria ignored Jenn and stated, “Self purifies, gets rid of... the gas.” It was not evident at this point in the interview what this phrase means, but later, she elaborated on where this idea came from. In her organic chemistry class, they discussed a reaction in which the carbon dioxide byproduct conveniently separates from the desired product by bubbling out of solution. In this moment, Victoria seemed to be in a brainstorming frame in which she hazily recalled this phrase from class. Jenn affirmed this idea with a “yeah” but then resumed her own line of reasoning, saying that the reaction continues until it can’t and that’s what makes the dough rise (line 7). At that point, Victoria and Jenn both decided they had reached an answer and were done thinking about it (lines 8 and 9). The whole discussion shown in Table A.2 lasted just sixty seconds.

Table A.2. Dialogue between two participants, Jenn and Victoria, which occurred about four minutes into the interview, just after they had mixed together the baking soda and acids.

Line	Speaker	Dialogue
1	Jenn	Hmm. Okay, so it's bubbly. So that's probably…
2	Victoria	Carbon dioxide? Oxygen?
3	Jenn	Yeah, it's one of those. Some sort of gas <laughter> that makes it, you know, it sounds fun. <laughter> I’m like so like…ADD. Um, yeah, well, okay, so like with yeast, that's what, some type of microbe. Right? And what, protist? I should know this, I’m in microbio.
4	Victoria	You’re in micro, I dunno.
5	Jenn	I think it's protist, so that means that it excretes something. So then it probably has like, releases carbon dioxide, and so it makes it grow. So then if that's a gas, then it would kind of do the same. Right?
6	Victoria	Self purifies, gets rid of… the gas.
7	Jenn	Yeah. Reacts ‘til it can, can't, and then it makes it rise. That's, that’s my…
8	Victoria	That’s our consensus. <laughter>
9	Jenn	That’s my… educated guess.

Although Jenn and Victoria indicated that they were finished answering, it is not clear if they had constructed an answer that they found satisfying. They seemed to be in a brainstorming frame initially, recalling bits of information from their microbiology and chemistry classes but never clearly connecting them in a causal manner to the phenomenon. It is possible that their reasoning made sense to them, but the more likely explanation is that they prioritized getting an answer that was close to correct (presumably as determined by the interviewer). This strategy of offering some ideas had likely brought them success in their past courses. In their current organic chemistry course, partial credit was awarded for invoking some correct ideas, even if not clearly connected. Thus, the students' behavior in this portion of the interview makes sense if they were approaching the interview from an oral examination frame.

The speed with which they arrived at an answer also makes sense in light of typical education experiences. In most college science classes, including the chemistry courses at this university, the material is covered at a rapid pace, and even if questions are posed to students during class that ask them to try to come up with an explanation, they are only given a short time to do so. Furthermore, these students received timed assessments that required them to be able to produce answers quickly. Even in their study time, students likely feel pressure to work quickly in order to accomplish all of the tasks they need to for all of their classes. Conversely, the students had likely not experienced many formal education situations in which they were expected to discuss and grapple with problems for lengthy periods of time.

Negotiating Frames

A person's frame is influenced by a number of internal and external factors, including the people around them. Because we interviewed students in pairs, we were able to identify instances where the two students seemed to initially approach the situation from different frames and through dialogue establish a shared frame. Consider the following excerpt from the interview with Denise and Mandy (Table A.3). This occurred soon after Denise and Mandy had finished mixing together vinegar and baking soda and lemon juice and baking soda. They had observed that the bubbles produced by mixing the lemon juice and baking soda lasted longer than the bubbles produced by mixing vinegar and baking soda.

Table A.3. Dialogue between Denise and Mandy, in which they initially seem to be in different frames but end in the same oral examination frame.

Line	Speaker	Dialogue
1	Denise	Are we supposed to tell you our conclusion or…?
2	Interviewer	Yeah, just tell me what you're thinking. You like one better than the other?
3	Mandy	Look, it's still going.
4	Denise	Yeah. It probably has a slower reaction. Uh, I don't know. I feel like maybe vinegar reacts faster and so that's why it's…
5	Mandy	I'm just mixing it up.
6	Denise	Okay. Um, well, I feel like we both agreed that this lemon juice is gonna work better because… um…
7	Mandy	Still got the, the bubbles going for it.
8	Denise	Yeah, and I feel like maybe the vinegar kind of like dissolves it right away whereas like the lemon juice, it like reacts with it, but doesn't like dissolve it as fast. Or like the baking soda, maybe? I don't know if it's like, um…
9	Mandy	I think the fact that there's still bubbles in this one, that it would make more sense than to just need, to just be adding like liquid into cinnamon rolls.
10	Denise	Do we have to like give you like some like chemistry background?
11	Interviewer	Oh no, no, you can just explain it whatever way, like makes sense to you.
12	Denise	Well, I mean, this could be since, I feel like this is more acidic, so it has more resonance structures.
13	Everyone	<laughter>
14	Denise	And like, maybe it's like more stable, so it's not as reactive.
15	Mandy	What's, what's baking soda. What is that real one's name?
16	Denise	[Laughing] I don't know.
17	Mandy	Oh my god.
18	Denise	Oh, wait. Is it…hmm?
19	Mandy	Just, is it carbon…?
20	Denise	Car–, carbonate? No. No, it's not.
21	Mandy	Isn't it sodium…?
22	Denise	Sodium carbonate?
23	Mandy	Bi… <laugh> I don't know.
24	Denise	I don't know. Well, anyways, um, yeah, that's our conclusion. <laugh>

Throughout this excerpt, we argue that based on the questions she asked and the vocabulary she used, Denise was adopting an oral examination frame. In line 1, she sought to establish how she and Mandy should frame their response, asking, "Are we supposed to tell you our conclusion…?" Her assumption, based on this question, was that she and Mandy were expected to produce an answer for the interviewer. The language "supposed to" suggests recognition of an authority figure (*i.e.*, the interviewer) and indicates that Denise likely saw the interview situation as akin to a classroom situation in which teachers tell students what to do. Furthermore, the phrase "tell you our conclusion" conveys a goal of obtaining some sort of final answer. This is also consistent with classroom norms. Typically, when a teacher asks a question, either in class or on an assessment, the student is expected to give the correct answer with a short time constraint, as discussed previously.

Although the interviewer tried to destabilize Denise's oral examination frame with the response, "just tell me what you're thinking," (line 2), Denise's follow up question in line 10 indicates that Denise persisted in an oral examination frame. She asked, "Do we have to like give you like some chemistry background?" It seems she did not feel that their response so far had met the appropriate criteria for an answer. She did not see her everyday knowledge as appropriate to this situation but rather felt the need to give specialized knowledge. Furthermore, her initial explanation (line 8) used chemistry terms like "dissolve" and "react" but in a way that did not seem to make much sense even to her. She also, unprompted, invoked resonance structures (line 12), which are specialized drawings for communicating molecular structure and are often expected on assessments. Throughout this portion of the interview, it was not clear what Denise was saying or how her ideas would connect to the phenomenon. Therefore, it is unlikely that these ideas helped Denise explain, to herself or to others, why lemon juice produced longer lasting bubbles than vinegar.

In contrast, Mandy did not seem to be in an oral examination frame at the beginning of this excerpt. In line 7, she noticed the bubbles had lasted in the lemon juice cup, and in line 9, she backed up Denise's assertion that lemon juice was the better choice, saying, "it would make more sense." This implies that their decision and reasoning needed to make sense to her rather than, for example, use scientific terms from class. Furthermore, she referenced the cinnamon rolls in line 9, demonstrating that her thinking was connected to the scenario presented to them. In contrast, the context was entirely absent from Denise's reasoning, or at least the reasoning she voiced aloud.

However, Mandy switched to an oral examination frame partway through this excerpt (line 15), possibly in response to cues from Dense and despite the attempts of the interviewer to destabilize this frame. In line 15, she asked Denise for the "real name" of baking soda, by which she meant its chemical name. She and Denise then try to recall the chemical name (sodium bicarbonate) but were not quite able to remember. They ultimately decided to let it go, and Denise wrapped up their conversation, saying "Well, anyways, um, yeah, that's our conclusion" (line 24). Mandy did not add anything else and tacitly accepted that they have given their answer. Although they started this portion of the interview in different frames, they both ended in an oral examination frame, and it was not until later in the interview that both students demonstrated evidence of a sensemaking frame.

Sensemaking Frame

Although in the previous example the oral examination frame was stabilized over the sensemaking frame, there were moments where students spent longer periods in a different frame. Denise and Mandy, who initially settled into an oral examination frame, later transitioned into a sensemaking frame. In the excerpt shown in Table A.4., they two of them discussed the role of heat. Denise suggested that heat might affect the rise, drawing on her knowledge that heat generally speeds up reactions (lines 1 and 3). Mandy agreed and articulated a relationship between number of bubbles and dough rising (line 4). The simple mechanism of air bubbles causing dough to rise put forward by Mandy prompted Denise to identify a gap in the explanation. She asked, "don't you put the yeast in before you bake it?" (line 5). Mandy responded with her own question, "you soak it in water first, don’t you?” (line 6). They then debated the temporal sequence of baking, drawing on their own experiences, in order to determine if heat plays a role in the rising process.

Table A.4. Dialogue between Denise and Mandy demonstrating a transition into a sensemaking frame, marked by vexing questions.

Line	Speaker	Dialogue
1	Denise	I feel like the heat, like from the oven might kind of affect it too.
2	Mandy	The flavor? Yeah…
3	Denise	Like make it, or maybe like make it react faster, so it won't be as like, you know? Like compared to yeast, I feel like it might have a different effect. ‘Cause heat like speeds up reactions. <laughter> So… I don't know, I feel like I would just use the lemon juice just for flavoring. I don't know if it would really help as much for rising, for making it rise. It could for a little bit.
4	Mandy	I mean, yeah, [inaudible] a lot of bubbles and bubbles mean gas and gas means air can rise.
5	Denise	Wait, well actually I feel like you, like don't you put the yeast in before you bake it? To make, you like you–
6	Mandy	You soak it in water first, don't you? The yeast in water first and then you put it in?
7	Denise	‘Cause I know like my sister, she like makes cinnamon rolls and she like, lets it rise first before baking. So maybe it would rise first and then like baking it wouldn’t really affect it as much.
8	Mandy	Cause isn't it like when you, yeah, ‘cause yeast will rise, and then when you, when you bake stuff with yeast in it, doesn't that, that does another process of something.
9	Denise	Of what?
10	Mandy	I don't know, does it make it rise more, does it make it stop rising or something? Has to rise slightly more, I guess.
11	Denise	Uh… I don't know.
12	Mandy	But baking soda anyway makes stuff rise. So…
13	Denise	I guess, yeah.
14	Mandy	That’s why [inaudible] put it in muffins. That and like baking powder. I don't really know the difference though.
15	Denise	Interesting.

Although the tone is hard to convey in a transcript, the questions in lines 5 and 6 seem to be examples of the vexing questions described by Odden and Russ. In this portion of the interview, the vexing questions marked a transition into sensemaking. Denise realized that the mechanism summarized by Mandy did not make sense with the steps involved in baking, and she drew attention to this gap with her question. Mandy replied with a question of her own, demonstrating her willingness to join Denise in sensemaking. The following discussion of when rising occurs is consistent with the *identifying setup conditions* component of mechanistic reasoning identified by Russ et al. (2008). Drawing upon work by Machamer et al. (2000), they described setup conditions as “the spatial and temporal organization of entities that begin the regular changes of the mechanism that produce the phenomenon” (p. 512). Although Mandy and Denise do not flesh out a complete mechanism for dough rising, the beginning of mechanistic reasoning coupled with their initial desire to address their gap in knowledge, is consistent with a sensemaking frame.

Lack of Mechanistic Reasoning at the Molecular Level

Although some students spent time in a sensemaking frame, there was relatively little mechanistic reasoning at the molecular level. Rather, as students engaged in thinking about the phenomenon, they seemed more inclined to brainstorm other ingredients with which they could experiment. For example, Inaya drew on her baking experience to suggest using egg whites (Table A.5). She knew that whipped egg whites resulted in a fluffier cake and thought it might work for cinnamon roll batter as well (line 1). When asked how the egg whites make cakes fluffier (line 2), Inaya was unsure (line 3) but was able to describe in detail the texture and appearance of the whipped eggs (lines 5 and 7). Saanvi drew on her knowledge from cooking shows to reason that whipping egg whites adds air to them (line 4) but did not offer any further mechanistic insight.

Table A.5. Dialogue between Inaya and Saanvi showing engagement with the phenomenon but in the form of brainstorming ideas to try rather than mechanistic reasoning.

Line	Speaker	Dialogue
1	Inaya	This is more of like a baking perspective. I don't know how well it would work. But if you take egg whites and whip 'em up, usually that gives you like a super fluffy consistency, and kind of like mix that into your batter. I don't know if it would still be like... cinnamon roll batter 'cause it would be super hard to like roll out and like move and stuff. But I do know that like whipped egg whites give you that super fluffy consistency, and it makes your cakes a thousand times better.
2	Interviewer	Um... And, and what does the, do you know what the like whipping the egg whites does to them to make them fluffier?
3	Inaya	I'm not sure. Go–
4	Saanvi	In the cooking shows, they always say, "Adds air to it." So I guess it adds air.
5	Inaya	It goes from like this gooey consistency to like, almost like meringue, I wanna say. 'Cause it, it–
6	Saanvi	Like a whip cream sort of texture.
7	Inaya	Even lighter than that. It has like a lot of air in it. Like if you have like one egg white, it literally goes up to like this big. And you just like fold it in, and it, it'll be really good.

There are several possible explanations for the lack of mechanistic reasoning. Partially, this is likely due to the line of questioning from the interviewer. To trigger further discussion, she often asked about other ingredients or invited students to suggest ones they thought might work as yeast substitutes. This likely pushed students into seeking an empirical solution rather than an explanative one. The nature of the phenomenon likely impacted the reasoning elicited as well. Most students seemed satisfied with using bubbles to explain why dough rises; they did not need to invoke interactions at the molecular scale. Given how intellectually demanding and time-consuming mechanistic reasoning can be, it makes sense that students would employ other strategies, such as trying things out, or that they would choose to be content with an explanation that did not drill down to the molecular level.

Limitations

This study was undertaken as an exploration into sensemaking in the context of chemistry. As such, only a small number of students were interviewed. We do not claim that the themes we have observed are generalizable to the larger population of students enrolled in organic chemistry at this institution or other institutions. Furthermore, there are likely other factors that we did not uncover due to the small sample size.

In addition, we only interviewed students in the context of a single phenomenon, that of using baking ingredients as yeast substitutes. The phenomenon is likely an important aspect of the sensemaking process; it plays a role in bounding the explanations students might come up with and the extent to which they are interested in figuring out what is happening. It would not be trivial to select phenomena that interest students and are explainable (at least to some extent) with the chemistry content covered by a particular chemistry course. More work would be needed to understand which phenomena support sensemaking and why.

Finally, the interview setting is considerably different from a classroom setting, and we do not know how much of what we observed in the interviews would transfer to the classroom. These interviews were conducted by someone outside of the course, and none of the work was graded or presented as preparation for graded assessments. The student pairs we interviewed were friends, and no other students were present. It is unknown how students will be affected by being in a larger group, many of whom they do not know well.

Future Directions

We designed an exploratory study to probe students' sensemaking in the context of chemistry-related phenomenon, modeled on the example from physics education (Odden & Russ, 2019b). Initial analysis suggests that students initially approached the interview with the goal of getting to an answer quickly but transitioned into sensemaking later in the interview. Future analysis will characterize the factors that facilitated these frame shifts. This will allow us to better understand the mechanisms by which sensemaking frames can be initiated and sustained. However, we do not know the extent to which the findings from our study will transfer to a formal education setting or what other phenomena would be effective foci. Thus, there remains much work in regard to curriculum development and chemical education research before we know the feasibility and the outcomes of implementing sensemaking opportunities into formal education.

It may be fruitful to compare the findings of this project with the findings of a related project in which students were interviewed about their experiences in an organic chemistry course. Both projects targeted students' epistemic cognition but in very different contexts. While the project described here, which I will call Sensemaking Tipping, focused on a situation that could be encountered outside of the classroom, the other project, which I will call Epistemological Messaging, focused on formal education experiences. The interview protocol for Epistemological Messaging was designed to elicit evidence of students' epistemic aims, their criteria for evaluating knowledge, and the processes they saw as effective and ineffective in reaching their epistemic aims. The analysis, to which I contributed, focused on how the course shaped their understanding of knowledge and knowledge construction. A major finding was that the assessments and accompanying answer keys served to constrain what counted as acceptable knowledge products. For example, students perceived the need to use certain terminology in their explanations in order to receive credit on assessments, regardless of whether they themselves deemed these words useful for expressing their understanding.

Future work will make use of the analytic techniques described in these two projects to further probe how students' epistemic cognition is influenced by their chemistry courses. We are especially interested in exploring discussion classes, in which students are expected to work together in small groups, assisted by a teaching assistant. These learning environments have two major advantages for studying epistemic cognition. One, they typically involve more dialogue and interactions among students than lecture classes, which makes it easier to infer students' thinking. Second, they offer flexibility in experimenting with different interventions designed to affect epistemic outcomes. Because multiple sections run in parallel, quasi-experimental study designs are feasible. For example, one could experiment with different phenomena relating to acid-base chemistry to see which are associated with more students engaging in particular frames or

References

Berland, L. K., & Hammer, D. (2012). Framing for scientific argumentation. *Journal of Research in Science Teaching*, *49*(1), 68–94. https://doi.org/10.1002/tea.20446

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.

Collins, A., & Ferguson, W. (1993). Epistemic forms and epistemic games: Structures and strategies to guide inquiry. *Educational Psychologist*, *28*(1), 25–42. https://doi.org/10.1207/s15326985ep2801_3

Glaser, B. G. & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine.

Hammer, D., & Elby, A. (2002). On the form of a personal epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 169–190). Lawrence Erlbaum Associates, Inc.

Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 89–119).

Haraldsrud, A., & Odden, T. O. B. (2024). Using feedback loops from computational simulations as resources for sensemaking: A case study from physical chemistry. *Chemistry Education Research and Practice*. https://doi.org/10.1039/D4RP00017J

Hunter, K. H., Rodriguez, J.-M. G., & Becker, N. M. (2021). Making sense of sensemaking: Using the sensemaking epistemic game to investigate student discourse during a collaborative gas law activity. *Chemistry Education Research and Practice*, *22*(2), 328–346. https://doi.org/10.1039/D0RP00290A

Jiménez-Aleixandre, M. P., Bugallo Rodríguez, A., & Duschl, R. A. (2000). “Doing the lesson” or “doing science”: Argument in high school genetics. *Science Education*, *84*(6), 757–792. https://doi.org/10.1002/1098-237X(200011)84:6<757::AID-SCE5>3.0.CO;2-F

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about Mechanisms. *Philosophy of Science*, *67*(1), 1–25.

MacLachlan, G. L., & Reid, I. (1994). *Framing and interpretation*. Melbourne University Press.

Lemke, J. L. (1990). *Talking science: Language, learning, and values.* Ablex Publishing Corporation.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas* (p. 13165). National Academies Press. https://doi.org/10.17226/13165

Odden, T. O. B., & Russ, R. S. (2019a). Defining sensemaking: Bringing clarity to a fragmented theoretical construct. *Science Education*, *103*(1), 187–205. https://doi.org/10.1002/sce.21452

Odden, T. O. B., & Russ, R. S. (2019b). Vexing questions that sustain sensemaking. *International Journal of Science Education*, *41*(8), 1052–1070. https://doi.org/10.1080/09500693.2019.1589655

Odden, T. Ole. B., & Russ, R. S. (2018). Sensemaking epistemic game: A model of student sensemaking processes in introductory physics. *Physical Review Physics Education Research*, *14*(2), 020122. https://doi.org/10.1103/PhysRevPhysEducRes.14.020122

Rosenberg, S., Hammer, D., & Phelan, J. (2006). Multiple epistemological coherences in an eighth-grade discussion of the rock cycle. *The Journal of the Learning Sciences*, *15*(2), 261–292.

Russ, R. S., Lee, V. R., & Sherin, B. L. (2012). Framing in cognitive clinical interviews about intuitive science knowledge: Dynamic student understandings of the discourse interaction. *Science Education*, *96*(4), 573–599. https://doi.org/10.1002/sce.21014

Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education*, *92*(3), 499–525. https://doi.org/10.1002/sce.20264

Sevian, H., & Couture, S. (2018). Epistemic games in substance characterization. *Chemistry Education Research and Practice*, *19*(4), 1029–1054. https://doi.org/10.1039/C8RP00047F

Tannen, D. (1993). *Framing in discourse*. Oxford University Press.

Supporting Information

Tables A.S1 through A.S6 provide transcripts of the interviews. All information identifying particular people or the institution at which these interviews took place has been removed. Pseudonyms were given to the participants to protect their identities.

Table A.S1. Transcript of interview with Ashley and Kristin.

Speaker	Timestamp	Dialogue
Interviewer:	00:13	Alright. So.
	00:17	What I wanna talk about today, uh, is baking. Um, do either of you like to bake? Do a lot of baking?
Ashley:	00:26	Um, just a little bit, um, make cookies around the holidays. <laughter>
Interviewer:	00:30	Yeah.
Kristin:	00:31	Like to bake, but don't do it enough.
Interviewer:	00:33	Yeah. Are you guys still in dorms right now? Or are you–?
Kristin:	00:37	We're in an apartment. We've, we've made a few things, but not the, not the best space for that nowadays.
Interviewer:	00:43	Yeah. Probably not a lot of time either.
Ashley & Kristin:	00:45	Yeah.
Interviewer:	00:46	Um, yeah. Uh, do you guys, are you into any of the like baking shows? Great British Bakeoff…
Ashley:	00:53	Not really.
Kristin:	00:54	Yeah. I don't know, we used to watch a few, um, or just like, I don't know, we used to watch like Cake Boss a long time ago. Um, and a few of them, but not recently. <laughter>
Interviewer:	01:06	No, that's fair. I feel like I wasn't really into baking either or any of the baking shows, and then someone got me started on Great British Bake Off, and I was like, "Oh, they're actually like nice. 'Cause all the other like cooking, baking shows, it was just like, I got so tense. Um, but yeah, I, I'm more into the eating side of things, personally. I just keep trying to get my sisters to be like, "Hey, you wanna bake this?" <laugh>
	1:31	Um, so, um… One of the things I learned through watching Bake Off, uh, is the importance of yeast in, in, some baking things. Um, so any, any idea what yeast does in like bread or…?
Ashley:	01:51	Well, it makes it rise.
Interviewer:	01:52	Makes it rise.
Ashley:	01:53	Yeah.
Kristin:	01:54	Yeah. <laughter>
Interviewer:	01:56	Yeah. Um, yeah, definitely. Um, you ever baked with yeast before?
Ashley:	02:05	Um, yeah, we've made like pretzels before.

Speaker	Timestamp	Dialogue
Interviewer:	02:08	Ooh, pretzels, nice. Um… so… uh, you're definitely right about the rise because, uh, one time I tried to use my sister's cinnamon roll recipe that called for yeast, and I put in the yeast, and it turns out the yeast had gone bad. Uh, 'cause when I pulled them out of the oven, they were just like the flattest. Um, I ate, I ate a couple because I was like, they're covered in cinnamon and sugar, they can't be that bad, but the texture, not, not great. Um, so, got rid of that yeast. Um, but it turns out, I was, I was looking on the internet as well, and if you don't have yeast or all your yeast has gone bad, um, you can actually use a one-to-one mixture of vinegar and baking soda, um, dissolved in milk, um… according to the internet. Um, so, uh, do you guys think that would, that would work? Um, and I, I, I did bring in some vinegar and baking soda, um, if you wanted to play around with mixing them together. Um, but any ideas of that, if I should trust the internet here?
Kristin:	03:20	Um…
Everyone:	03:21	<laughter>
Ashley:	03:24	Yeah. I don't know. Prob— I feel like there's stuff with like baking soda that you, like when you're little, you make like volcanoes explode or whatever. So I feel like it kind of like bubbles, like yeast bubbles when mixed with vinegar. So maybe that would work too.
Kristin:	03:42	Yeah. I know like different things can have like similar effects, but it also… It's… I don't know, it seems like it could have the same effect, but… with the liquid it's kind of like a different… I don't know, it's more, it's like liquid and yeast is not liquid, so I don't know if that would impact it… differently.
Interviewer:	04:09	Yeah. That's a good question. Um… You talked about, about making bubbles. Um, should, should we try putting in in, to see, see? Don't be shy.
Ashley & Kristin:	04:25	<laughter> [Ashley and Kristin reach for the cups, vinegar, and baking soda.]
Interviewer:	04:26	Um, I meant to bring in clear cups, but… And my broken tupperware <laughter> [Ashley struggles to open the tupperware containing the baking soda.]
	04:37	There we go. Um, so yeah, if you want, feel free to dump a little bit together. [Ashley carefully shakes some baking soda into the cup.]
	04:48	Probably should have brought a spoon or something. You guys are my first interview FYI so. [A little bit of baking soda is spilled.]
	04:58	Don't worry, I came prepared [gesturing to paper towels]. [They pour the vinegar into the cup with baking soda.]
Kristin:	05:04	Definitely a lot of bubbles.
Interviewer:	05:06	<laugh> Alright. Uh, So what do you think, would the bubbles help in the baking? What do you think's going on with it?

Speaker	Timestamp	Dialogue
Ashley:	05:21	Yeah, I think so, because I, I think that's how yeast works. Like it kind of… I don't know. Like I took culinary in high school and they're like, oh, it like eats the like glucose and then it like burps out bubbles. <laughter> Like, that's what they said. Um, and then like the bubbles or what, let it rise because there's just, I don't know, more air pockets and then it gets taller. So I feel like that's similar.
Kristin:	05:55	Yeah. I was just wondering… if, I don't know, if it… would make too many bubbles, like in this where like, if the gas gets released, then it would end up being denser again because it wouldn't be trapped in there, like making it more airy, but I don't know. It…
Everyone:	06:23	<laughter> [Kristin swirls the contents of the cup around again.]
Interviewer:	06:24	The answers in the bubbles.
Kristin:	06:26	Yeah. <laughter>
	06:28	I don't know. I did a science day thing, um, with one like group, and we did an experiment with like hydrogen peroxide and yeast. And it was, we needed like soap to help like trap the air bubbles that, like the yeast was catalyzing the reaction of the hydrogen peroxide, like splitting into water and gas. Um… So… I don't know, it, I guess… I don't know if yeast… causes the reaction or just like speeds it up. Um… So I don't know how that compares to what this is doing too. BUT it seems like, if this makes bubbles and it could be in the baking stuff, that that could be trapped and still make it rise.
Interviewer:	07:30	That's an interesting idea that the trapping, the bubbles, um, ‘cause it does say to like have these together in milk.
Kristin:	07:37	Oh yeah.
Interviewer:	07:39	Uh, so, and I, I honestly, I've been wondering about, “Well, what does milk do there?” Um, maybe you're, you're onto something.
Kristin:	07:49	Maybe so. I don’t know.
Everyone:	07:50	<laughter>
Interviewer:	07:52	So, um, if I were to give you like a loaf of bread made this way, um, with baking soda and vinegar instead of yeast, um, do you think you'd, you'd be willing to try it? Would you eat it? Would you be worried about it? Would you be like, “Oh yeah, this should be good”?
Ashley:	08:11	Yeah, I would try it. I feel like vinegar though has like a pretty strong taste, so I don't know how that would affect the flavor of the bread. Um…
Kristin:	08:21	Yeah. I mean, it's still edible. So—
Everyone:	08:24	<laughter>
Kristin:	08:27	Worth a try.
Interviewer:	08:29	Wouldn’t poison you, you're saying?
Kristin:	08:31	Yeah.
Interviewer:	08:31	Yeah. Might not, might not taste as good.
Kristin:	08:34	Yeah. Yeah.
Everyone:	08:38	<laughter>

Speaker	Timestamp	Dialogue
Interviewer:	08:40	Um… Alright. Uh, Do you think, um… something else would work? Uh, like if we didn't have vinegar maybe, or maybe we were worried it wasn't gonna, um, taste as good. Um, what if we had something like, uh, lemon, lemon juice? You see that kind of like, a bit of that added to a lot of baking stuff. Um, do you think that would, that would work instead? The lemon juice with baking soda?
Kristin:	09:11	I feel like it might because… I don't know, it says, well, the vinegar is acidic. And lemon juice–
Ashley:	09:19	–is also acidic–
Kristin:	09:20	–I've been told, is acidic.
Everyone:	09:22	<laughter>
Kristin:	09:23	So it could have–
Ashley & Kristin:	09:24	–similar properties.
Kristin:	09:26	Similar reaction. I don't know about the ratio… um, because… it might be more or less acidic, but if you want to keep it the same, um…I don't know, if you want it to react the same, it might have to change that up a little bit or experiment with it, but… I feel like it could work. <laugh>
Interviewer:	09:51	You, you'd try that bread too?
Kristin:	09:53	Yes, yes.
Interviewer:	09:54	Um, do you think it would, would do the same thing where it would make all the bubbles?
Ashley:	10:01	I would guess so. I don't know. Since they're both acids, they're both acidic, I feel like it would have like the same reaction pathway or whatever. I don't know.
Interviewer:	10:12	Yeah. Just, just might need to change the amounts a little bit.
Ashley:	10:16	Yeah.
Interviewer:	10:18	Alright. Um, so vinegar is a clear liquid. Um… If we tried a different clear liquid, um, maybe you have some vodka in your, um, in your kitchen and maybe, maybe not vinegar. Um, 'cause this is college, um. Not assuming anything about you guys specifically, maybe, maybe your roommate. Uh, do you think that would, that would work, if instead of vinegar and baking soda we had some, some vodka or… some other alcohol.
Kristin:	10:55	Yeah, vodka's an alcohol, so it probably is more basic. Um, if it has alcohols, have OH, they probably want to give away the H rather, the hydrogen, rather than take on another one.
Ashley:	11:12	Yeah.
Kristin:	11:12	So I feel like it wouldn't… Assuming that this is like an acid and that's like a base, an acid-base reaction or something, then I think it would be a very small reaction if there was any visible reaction, if they're both more basic.
Ashley:	11:35	Yeah, but I don't know. I don't really know what, what, how much the difference is, but I know that like glucose can either, it could be like lactic, er, it can go through like alcohol fermentation or…
Kristin:	11:50	Lactic acid fermentation.

Speaker	Timestamp	Dialogue
Ashley:	11:52	Yeah. Um… So I don't know if, since that's what like yeast does… to make, um, like the bubbles. I don't know, if like, since it's the alcohol… if that's like also the product, maybe that's like too far down the line in the process since you wanted to make the bubbles and not just like add the bubbles or whatever, you know?
Kristin:	12:21	Yeah.
Everyone:	12:24	<laughter>
Interviewer:	12:26	That's, that’s okay. We don't need to know the answer. Um, what do you think, like, is in the bubbles?
Kristin:	12:34	Um… some sort of gas. I don’t know if it would be… oxygen or… like… hydrogen gas or… <laugh>
Ashley:	12:49	Or like carbon dioxide?
Kristin:	12:51	That's [inaudible].
Interviewer:	12:52	Something like that, yeah. One of those, one of those gasses. Alright. Um, do you have any other ideas for, for yeast substitutes? I've been just throwing out a couple random ones that comes to mind. Do you, you think of anything else that, that might work or something you've, you've heard about?
Kristin:	13:15	Um…
Ashley:	13:20	I don't know. I'm just trying to think back to culinary and like freshman year of high school, like all the different leavening agents.
Interviewer:	13:26	You didn't think you were gonna use THAT knowledge when you walked in, did you?
Ashley:	13:31	No <laughing> Um, I kind of think that maybe like eggs, could be work, work as leavening agent too? I don't know what about them… would help, but, um…
Kristin:	13:47	I should have taken culinary <laughter>. Um… Yeah, do you know anything about the difference between baking powder and baking soda? <laugh>
Ashley	13:59	Yeah, I kind of feel like they both were leavening agents, but I don't, I know that you, you can't use 'em interchangeably. I don't know what, yeah, I don't really know.
Interviewer:	14:10	That one always gets me. I have to like triple-check the recipe, is it baking powder or baking soda?
Kristin:	14:18	Yeah. Um…
Interviewer:	14:22	I should have brought some in maybe to, and we can compare what happens when you add vinegar.
Kristin:	14:27	Next time.
Interviewer:	14:29	<laughter>
Kristin:	14:30	Um… I guess… I don’t know, if you think of like… carbonated things? They already have, or like–
Ashley:	14:47	They have bubbles in them too.
Kristin:	14:48	I don’t know if that could be used. Yeah. That releases gasses. If that could be trapped as well. It seems very strange, but…
Interviewer:	15:02	But, but if that's the secret of like trapped gas in there, that, that makes a lot of sense.
Ashley:	15:07	Isn’t that a thing, like soda bread or something? I feel like I've heard of that.
Kristin:	15:11	There's something called soda bread. I don’t know how it’s made.
Everyone:	15:14	<laughter>

Speaker	Timestamp	Dialogue
Interviewer:	15:16	I can't say I've made it either. Um, but it sounds like it probably involves soda.
Ashley:	15:23	Maybe it’s just baking soda, I don't know.
Interviewer:	15:24	Yeah, oh, that's… I don't know. Yeah. Um… Yeah. Uh, you talked, you mentioned a little bit about maybe this is like an acid-base thing, so you just need something, something with an H, um, and, and, and some sort of base to go with it. Um… So… you think of… other, other acids. Um, if you had something like, I feel like… in, in class, the go-to example is like HCl or something. Um, would you, would you wanna try that or eat anything made with that?
Kristin:	16:14	Um… <laugh> I feel like that's probably too strong… of a base? Uh, er, acid. Um, I don't know. I feel like… it could be… dangerous to eat. <laughter> Um, I don't know, what do you think?
Ashley:	16:39	Yeah, I don't really know. Um…And I don't know, like, I don't know if all like the acids that you could use would make like a gas as a product or if it would be, I don't, like a liquid or solid. You don't want that to be forming either.
Interviewer:	16:59	Right. Probably, probably wouldn't taste good.
Ashley:	17:02	Yeah.
Everyone:	17:02	<laughter>
Kristin:	17:04	Um… Yeah… We… We did some reactions yesterday that made carbon dioxide. <laugh>
Interviewer:	17:19	Oh yeah? Are you guys in… [OChem lab], [OChem II]?
Ashley:	17:23	OChem. [II]. Yeah.
Interviewer:	17:25	Obviously [OChem II]. Um…
Kristin:	17:28	Um… <laughter>
Interviewer:	17:33	That's, that's okay if you don't have any other ideas. Um, uh, I have to laugh. Um, this is just a total aside, but I did a practice interview on some of the, the undergrads in our group. Uh, and they were like, “Yeah, let's try HCl. Yeah, I’d totally eat that.” And I'm sitting there going, “Uh, that's kind of strong, that might burn a little bit.” And you're both like, “Mm, that could be dangerous.” <laughter> Like the boys were just like, “Yes, let's eat it!” Um, alright. Um… Well, uh, that is, that is honestly all the questions I have for you. Um… Anything else you want to add at all? Um, about, about what's going on here? Uh… in our imaginary kitchen.
Kristin:	18:31	Um…hm. <laughter>
Interviewer:	18:36	It was very open ended. You're like, “Uh…”
Kristin:	18:39	Um, trying to think of something. Um… I don’t know.
Interviewer:	18:54	<laugh> There, there are only so many things that you're allowed to eat. Um… Well, uh, if not, I'm gonna, um, turn off the recording equipment here.

Table A.S2. Transcript of interview with Brad and Connor.

Speaker	Timestamp	Dialogue
Interviewer:	00:08	Alright. So the topic I wanted to talk about today, um, is baking. Do either of you like to bake at all?
Brad:	00:18	I bake a little bit.
Interviewer:	00:19	A little bit. Yeah.
Connor:	00:20	I mean I have. Not very often.
Interviewer:	00:24	It's probably kinda hard when you're an undergrad and got all those busy classes and probably not terribly large kitchens.
Connor:	00:30	No.
Interviewer:	00:32	Um, is there anything you like in particular like to bake?
Brad:	00:36	Um… Pecan pie. Otherwise, I guess there's baking [inaudible]. I don't deal much with a cake. Otherwise…
Interviewer:	00:49	Mostly, mostly stick with the pies?
Brad:	00:51	I mean, I do cookies sometimes, I guess, so.
Interviewer:	00:55	Yeah, um… Do you, either of you watch any of the like baking shows I feel like were really popular, especially over COVID?
Brad:	01:02	No.
Connor:	01:03	I don't.
Interviewer:	01:03	Uh, that's where like most of my baking knowledge comes from, but mostly I'm on the side of like, convincing my sisters to bake something so that I can eat it. Um… <laughter> Definitely like the eating better. Um… But I did have a, a recent experience where I, I tried to make my sister's homemade cinnamon roll recipe, um… And that involves yeast. Um, so you guys familiar with what yeast does with baking? Yeah. So I found out the hard way that my yeast had gone bad, and I came out with these very like flat, dense cinnamon rolls. It was very disappointing 'cause this was like a overnight process of um, letting them rise. Um, so I was very disappointed, and I really wanted the cinnamon rolls, so I was trying to look up, um, yeast substitutes, and I found a couple of options. So one is vinegar and baking soda and then another is lemon juice and baking soda. And so my question for you guys is: Do you think either of these would work? Why do you think they work? Um, and which one do you think would be best? Um, and I, I did bring in, um, some cups and stuff here if you wanna play around with mixing them, um, if that would help you and see what happens.
Brad:	02:26	From, well, just from starting… I think you said that's vinegar, right?
Interviewer:	02:30	Yeah, here's some… vinegar. Feel free to… I'll put these closer.
Brad:	02:35	I don't know if you like [inaudible] the experience, experiments in high school where vinegar and baking soda, which is…
Connor:	02:40	Yeah. It's basically the volcano.
Brad:	02:42	Yeah. So I feel like that would definitely help it rise a lot because there's all like the gas or whatever's generating.
Connor:	02:48	Yeah. Assuming it's CO_2 and that's what yeast does.
Brad:	02:51	Makes things rise, yeah. Lemon juice, I have no idea.
Connor:	02:55	I mean, I guess they're both acids, like that's citric acid and that's gonna be acetic acid so…
Brad:	03:01	True.
Connor:	03:02	…they might do the same thing. Just different strengths.

Speaker	Timestamp	Dialogue
Brad:	03:04	Should we mix a little bit?
Connor:	03:05	Yeah, let's do it.
Interviewer:	03:09	You guys are the first ones who like stopped to think about it first. The others were just like, “Yeah, let's pour it together.”
Connor:	03:16	I’ll put some baking soda in the bottom of a couple glasses.
Brad:	03:18	Um, we’re gonna have to like somehow too kind of be accurate with like, is this the measurements? Is this all we have, just the spoon?
Interviewer:	03:27	I'm sorry. Yes. I, uh…
Brad:	03:29	I'll put like…
Interviewer:	03:30	I did not think to bring in a bunch of measuring implements. Maybe I should in the, in future ones.
Brad:	03:36	I don’t wanna blow this place up but yeah.
Interviewer:	03:40	Yeah. Hence, hence the paper towels.
Connor:	03:44	I don't know how…
Brad:	03:45	I think that should be–
Connor:	03:46	…aggressive this reaction's gonna be.
Interviewer:	03:50	I'll let you know if you put in enough to like hit the ceiling. <laughter> I'll, I'll cut you off there.
Connor:	03:56	I mean, if we did that with–
Brad:	03:58	Just put the vinegar in this <laughter>
Connor:	04:05	Alright, should we just…?
Brad:	04:06	Kind of eyeball it. [Pours in the vinegar]
	04:12	So yeah. It's kinda like…
Connor:	04:13	I mean, yeah, it bubbles,
Brad:	04:15	…what you’d expect.
Interviewer:	04:16	Surprise!
Brad:	04:20	Should we see what the lemon juice does?
Connor:	04:22	Yeah. Might as well. It's kinda…
Interviewer:	04:37	It's got that weird squeeze top.
Connor:	04:38	Yeah. …I mean they both do… something.
Brad:	04:44	This looks like it was a little more… I mean, it might have been the way just like this came up, but it definitely seemed like it came up slower and it's staying there.
Connor:	04:53	Yeah. Which I feel like would be better for a substitute of yeast ‘cause it’s…
Brad:	04:56	For baking cause it's like…
Connor:	04:59	…kind of aggressive.
Brad:	05:01	…longer term. And I think yeast too is… It takes longer versus like just regular baking soda or whatever.
Connor:	05:08	Yeah, ‘cause it takes like, you let stuff rise like over a couple hours and that was over… like five minute, five seconds.
Brad:	05:19	‘Cause you use yeast in bread, right? And like you let the bread dough rise before you bake it?
Interviewer:	05:24	That's what I've seen. I've never tried bread myself, but yeah.
Brad:	05:29	So this does like a–
Connor:	05:30	I mean, it’s just, it's staying a lot better too.
Brad:	05:33	Yeah. So honestly I feel like… the lemon juice with this might actually be better. I don’t know what you're thinking.

Speaker	Timestamp	Dialogue
Connor:	05:41	Yeah, I think it would be too, but I also, you know, might have to be baking something that's lemon.
Interviewer:	05:48	Yeah, how do you think it would taste with either of these?
Connor:	05:51	Don't, don't plan on trying ‘em.
Brad:	05:53	We'll just like assume we're going for looks and not tastes.
Connor:	05:56	Yeah. Probably okay. Gonna get anything more outta this? No, not really.
Brad:	06:08	So it's just like more to the question… versus like what we think?
Interviewer:	06:13	Um… I, I have some other questions on, on kind of related stuff. Um… Uh, like if you have any other ideas of what might also kind of produce the same effect? Um… What else might be a good yeast substitute?
Connor:	06:34	I'm trying to remember what baking soda is.
Brad:	06:38	See, my mom would know other substitutes, but…
Connor:	06:44	‘Cause like these are, I know are both acids, so probably an acid with baking soda would work.
Brad:	06:50	Yeah. Is there, are we supposed to like think about things beyond just baking soda?
Interviewer:	06:55	Uh, you should, you can, sure, yeah.
Brad:	06:57	Because…
Connor:	06:59	CO_2.
Brad:	07:02	Pure CO_2. I could be way off in saying this, but doesn't like egg yolk or something help a rise?
Connor:	07:10	Probably.
Brad:	07:11	Honestly, I don't know. See, it's funny ‘cause I did a baking experiment in chemistry in high school. Like it was the end of the year project, and I had to make, I made cupcakes with like yeast, baking soda, and there was something else. I don't remember what it was. But either way, I forgot to take ‘em out of my backpack over this summer, so the next week school year, I took them out, and they were just moldy. <laughter>
Interviewer:	07:34	Oh man.
Connor:	07:35	I was like, I came back from winter break and my roommates had left a bunch of stuff like on the counters. Of course I was the first one back. I really don't know what would, other than these, like I don't know what else is like slightly acidic.
Brad:	07:56	Um… Vinegar, lemon juice.
Interviewer:	08:05	I know it's kind of weird to try and picture like what's in your kitchen right now.
Brad:	08:09	Especially too ‘cause like right now, what?
Connor:	08:11	I mean, I guess like limes and all that like citrus probably has like the same stuff.
Brad:	08:15	Yeah. I feel like, yeah lime would be similar. Yeah, but right now my like refrigerator’s pretty empty as far as like baking supplies.
Connor:	08:23	Yeah. I have flour and sugar and that's about it.
Brad:	08:27	I know there's apple cider vinegar. Except that's probably the same things.
Connor:	08:32	Yeah, I guess.
Brad:	08:33	[Reading vinegar bottle] Distilled white vinegar.
Connor:	08:34	I just don't know what makes it different things.
Brad:	08:38	Like they behave similar.
Connor:	08:40	[Reading vinegar bottle] Distilled with water to 5% acid strength.

Speaker	Timestamp	Dialogue
Brad:	08:44	Does this say acid strength?
Connor:	08:48	I have no idea.
Brad:	08:55	Well, there’s baking powder.
Connor:	08:57	I don't know what the difference between baking powder and baking soda is. I just know that they're not interchangeable… in recipes.
Brad:	09:03	Yeah.
Interviewer:	09:05	I have to look it up every time.
Connor:	09:08	I know that like… you can clean with baking soda.
Brad:	09:13	True. I think baking powder would, do you think it'd behave the same? As baking soda? If you were to like mix it with these two again.
Connor:	09:23	I got, I got, I can go back to my apartment and try. I’ll let you know.
Brad:	09:29	So I said putting those with baking powder would be something to try.
Connor:	09:36	Yeah, I'm trying to think of something that, ‘cause these obviously made CO_2. I'm pretty sure that's the gas that gets created. And so like what other things have we learned about that would create CO_2? …I don't know… at all.
Brad:	09:56	There’s the reaction we just talked about like two weeks ago.
Connor:	09:58	Yeah. I don’t–
Brad:	09:59	I don’t remember what the reactions were called.
Connor:	10:01	I don't think we got that in [inaudible] kitchen. <laughter>
Interviewer:	10:03	I'm pretty sure that'd be illegal. Um, can you think of anything that might not like make CO_2 when you combine but already like have CO_2, um, that might work?
Connor:	10:16	Soda.
Brad:	10:18	True. There’s soda… Like any like fizzy drink, kinda like similar, similar to soda.
Connor:	10:27	Yeah. Club soda. Any of it really would have the CO_2 you need. I mean, could you use beer? Beer has yeast.
Brad:	10:36	True. Actually that is in… I don't know how it like works, but they do put like obviously beer in like… when you deep fry stuff like beer-battered fish, stuff like that.
Connor:	10:47	Yeah. And I've heard of like beer bread.
Brad:	10:49	And like the coating on that then kind of does like get real fluffy.
Connor:	10:53	It does. I know I've heard of like recipes that like use beer as like… I, I don't know if it's in like place of yeast in like making bread, but I've seen it. Maybe that would work ‘cause it's got the CO_2 in it. It might have some active yeast. Maybe not.
Brad:	11:13	I think that would be reasonable to try at least. Um…
Interviewer:	11:20	Do you think any other alcohols would work or just, just something like beer? Um… Do you think other alcohols would be like, maybe with the baking soda, do you think those would be like acidic enough to, to give you the bubbles? Um…
Connor:	11:40	I mean it's ethanol.
Brad:	11:44	Would that work?
Connor:	11:46	It might be a little acidic. I don't know what the pH, er, pKa of ethanol is.
Brad:	11:52	I don't either.
Connor:	11:54	I think it's around… water.
Brad:	11:57	I know it's like kind of somewhere like, like not super strong, but it's not like the weakest.

Speaker	Timestamp	Dialogue
Connor:	12:04	So I mean, maybe? On its own, probably not ‘cause it's… I think they get rid of all of, any yeast that they use.
Brad:	12:13	Yeah.
Interviewer:	12:18	I should have brought in just like water. Obviously, I'm not allowed to bring in alcohol for this, but like just plain water. Um… To mix if not, if those are similar. Um… Uh, one thing it also mentioned online is that in the case of the vinegar baking soda combo, they recommend having it with milk. Um, like a one-to-one ratio of milk and vinegar. Um, any idea what the role of milk is playing there? …Like how is, how does that help?
Brad:	12:55	Is it like before it, before it heats up?
Interviewer:	12:56	Any id–, I don't know the answer.
Interviewer:	12:58	Uh, yeah.
Connor:	13:00	When you put it in.
Interviewer:	13:02	Yeah.
Brad:	13:02	All I know is that when you like have milk heated up, obviously like it always wants to run over. But I guess before it would be heated up… What is in milk?
Connor:	13:14	I don’t know. The only thing I know that is in milk is lactose.
Brad:	13:18	I was gonna say, like more like…
Connor:	13:20	‘Cause you'd have some sugars in there.
Brad:	13:24	You have some sugars. Would you, would fats have anything to do with it?
Connor:	13:28	I don't think so.
Brad:	13:30	I don't know how that would behave with that.
Connor:	13:33	Do you remember what kind of milk they said?
Interviewer:	13:35	<laugh> I do not.
Connor:	13:37	Skim?
Brad:	13:39	[Inaudible] do low-fat sometimes [inaudible] difference.
Connor:	13:42	Yeah. I, I don't really know what else is in milk. I know like calcium, but I don't see calcium doing anything really.
Brad:	13:48	I feel like that wouldn’t either.
Interviewer:	13:51	Someone in another interview suggested that maybe the milk helps the bubbles stay longer?
Connor:	14:00	Maybe. ‘Cause whenever you have like a glass of milk and you pour it, there's always the bubbles on the side that just kind of stay there forever.
Brad:	14:07	I don't know what could be in milk that does that though.
Connor:	14:10	I have no clue. Maybe just, maybe the fats.
Brad:	14:14	A little more… Help with like a little more structure. Lipids. Um…
Connor:	14:23	Protein’s in milk. I feel like there's protein in milk.
Brad:	14:27	Yeah, there is. Proteins too could have something to do with it. …I feel like, yeah, those maybe actually like protein lipids or something like that.
Connor:	14:40	I mean, if you're already baking with–
Brad:	14:41	Basically help lo–, or help the structure.
Connor:	14:44	Yeah. I mean, if you're already baking with milk, it's not like you're adding like another thing.
Brad:	14:56	Yeah, I don't know what else would do it.
Connor:	14:57	Yeah.
Brad:	14:58	That's weird. <laughter>

Speaker	Timestamp	Dialogue
Interviewer:	14:59	Yeah. Um, and like I said, I genuinely don't know the answer to that. I was trying to figure it out. Um… Any, any other ideas of yeast substitutes or things you would wanna try, or any, any other comments about what might be going on here?
	15:28	It's okay if the answer is no. I just, no, I don't wanna cut you off if you're still rolling with ideas.
Connor:	15:29	No, no, I was just kind of… No, I've–
Brad:	15:33	I feel like everything… I've kind of thrown things out there.
Interviewer:	15:38	Alright. Um, well that, that's it, that's the interview.

Table A.S3. Transcript of interview with Emily and Lisa.

Speaker	Timestamp	Dialogue
Interviewer:	00:55	So do either of you guys like to bake at all? Do much baking? Yeah?
Emily:	01:01	I'm a box girl. I can do it from a box, but I can't do it from scratch.
Lisa:	01:04	Yeah, but like the boxes, like honestly you can make like better stuff a lot of the time.
Emily:	01:07	Yeah. Yeah.
Lisa:	01:09	I'm in food science, so.
Interviewer:	01:10	Oh, you're food science. Oh, this is gonna be right up your alley.
Lisa:	01:13	And like if I don't know this I'm gonna be like—
Emily:	01:16	She's going to be kinda carrying us.
Lisa:	01:16	I have to, I have to flex my one semester of food science classes right now.
Interviewer:	01:21	Oh, that's great. Uh, what do you, what do you like to bake?
Lisa:	01:25	Um… Well I kind of like, like, like bread just because it's like, so like, putzy. <laughter> Like unnecessary to make your own bread. Um, but I'll bake anything basically, um… Not cornbread.
Interviewer:	01:42	Ah. 'Cause you don't like the taste or is it annoying to make?
Lisa:	01:46	No, I like cornbread alright. But I just, my mom would make, uh, this like cornbread with like almond slices and like cranberries on top and I just–
Interviewer:	01:54	Oh!
Lisa:	01:54	I know, it sounds delightful. I don't like it. <laughter>
Interviewer:	01:57	Never heard of that. Um, you like cakes, muffin out of the box?
Emily:	02:04	Yeah, yeah. I kinda like muffins a lot. I used to make those a lot growing up. Um, but I feel like now it's transitioned to just like quick cookies for like people when you're just like, we need to comfort you.
Interviewer:	02:14	Yeah, right. You guys probably don't have a ton of time with all your classes. Um, do you watch any of the, the baking shows on Netflix? Like Great British Bake Off?
Lisa:	02:27	I'm more of a Food Network person, but my favorite show is Good Eats. Which I think is Cooking Channel technically.
Interviewer:	02:33	I haven't seen that one.
Lisa:	02:35	It's got Alton Brown. It's like from the eighties. It's so great.
Interviewer:	02:37	<laugh> Nice. Well, I, um, I am not a baking expert myself. Um… And actually the, the most advanced thing I tried was cinnamon rolls. Um, and it turns out my yeast was bad.
Emily:	02:57	Oh, no.

Speaker	Timestamp	Dialogue
Interviewer:	02:59	Um, and so they came out of the oven, they didn't rise, and I was like, “Hmm, this might be a problem.” But I put 'em in the oven anyway. Um, they were not great. Like I, like I tried 'em ‘cause I was like, “Oh there's cinnamon and sugar on here, like how bad can it be?” Um, but they, without a rise and without that fluffiness. They were just not as good. Um, so I was very disappointed. I wanted to make some more. Um, and so I was looking up yeast substitutes since apparently mine was not working. Um, and I came across, you can mix vinegar and baking soda or like lemon juice and baking soda. Um… And so my question for you guys is, do you think both of these would work? One of these would work? And which one do you think would be better? Um, and I, I brought lemon juice, vinegar, and baking soda in, and some cups, if you wanted to put 'em together, um, see what happens or anything.
Lisa:	04:02	So we can like, before we answer, we can do this first?
Interviewer:	04:05	You, you can do some chemistry.
Lisa:	04:06	Okay.
Emily:	04:06	I mean, I've done baking soda and vinegar together before to like clean stuff.
Lisa:	04:11	So I think what's gonna dictate this is which one's more concentrated in acid.
Emily:	04:16	Mm, yeah.
Lisa:	04:17	‘Cause then you're gonna be fluffier.
Emily:	04:18	Mm-hmm. It's diluted. And this is a hundred percent from concentrate.
Lisa:	04:23	So interesting thing about bottled lemon juice, they actually measure the pH of that. So if you make like mayonnaise, you should use that. Not fresh lemon juice because fresh lemon juice can like vary in its pH. Learned that from Good Eats, not from my classes.
Interviewer:	04:35	<laugh> Right. You don’t learn useful things in class.
Lisa:	04:37	Yes.
Emily:	04:37	Lemon juice concentrate.
Lisa:	04:41	So…
Emily:	04:43	Look at that. Sorry.
Lisa:	04:43	No, you like, go ahead. I have an idea, but I just wanna hear, but like, I don't wanna like dominate the conversation so like.
Emily:	04:48	Oh, no, you go for it. I feel like [indistinct].
Lisa:	04:49	Okay. I was feeling like what we could do would be like…I mean, we can't really measure out if it's gonna be the same amount, but if we measure like the same amount of baking soda in each of these, and then the same amount of that and just saw which one reacted more, we could say that one worked better. Maybe.
Emily:	05:03	Yeah. I like that idea.
Lisa:	05:06	Okay. So these caps look like relatively the same size, so I’ll use that for measuring that. So I guess just baking soda… a spoonful?
Emily:	05:13	Yeah, we could just use a spoon, yeah.
Interviewer:	05:17	I did not bring measuring implements, I'm sorry.
Lisa:	05:20	It's all good. I apologize if it overflows.
Interviewer:	05:23	Oh, you're good. I brought in paper towels.
Emily:	05:25	Yeah.
Lisa:	05:27	Does this look like a nice level—?

Speaker	Timestamp	Dialogue
Emily:	05:29	Ooh, these are not the same. Oh wait, hang on, hm, ’cause they break here, I feel like that's less.
Lisa:	05:35	We can use the same cap for both of them then.
Emily:	05:37	Can we cross contaminate?
Interviewer:	05:38	Yeah, that's fine.
Lisa:	05:40	Is… I forget, is the acid in lemon juice the same? It's not acetic acid, right?
Emily:	05:46	It’s citric acid.
Lisa:	05:47	It’s citric acid?
Emily:	05:49	Is that a thing?
Lisa:	05:50	It is a thing. But I feel like I heard that citric acid, like what you get in like candies is like not from citrus fruit. It's like from… Like, uh…
Emily:	05:59	It doesn't say.
Lisa:	06:02	It's like from something else. I think it's like fermented or something. Okay, that's like almost the same.
Emily:	06:08	Okay. Yeah. This size or this size? I feel like this size.
Lisa:	06:12	Yeah. Also I feel like lemon juice and vinegar is gonna be less gross than vinegar and lemon juice.
Emily:	06:19	Yeah.
Lisa:	06:20	Okay. Wait… um…
Interviewer:	06:22	What do you need?
Lisa	06:23	I’m like, I'm like, if we can't do this simultaneously, like we're just gonna have to like memorize how high it gets. Well, we'll look at the—
Interviewer:	06:29	I got some paper here. If you need to use [indistinct]
Lisa:	06:38	We're scientists, okay. And so it gets to about here.
Emily:	06:44	Yeah.
Lisa:	06:44	Okay.
Emily:	06:46	Can I use this piece?
Interviewer:	06:47	Yeah.
Emily:	06:48	Okay. [Wiping off cap]
	06:58	Sorry.
Interviewer:	07:01	It's okay. Controlling your variables.
Emily:	07:03	Mm-hmm.
Lisa:	07:06	I mean, I technically, I feel like we should be, like we should use like water in one of them, but we're not gonna do that.
Interviewer:	07:13	I only have coffee in my water bottle.
Lisa:	07:14	That kind of be cool actually. Would that…?
Interviewer:	07:18	Not giving up my coffee.
Lisa:	07:18	I don't think that would really do a lot.
Emily:	07:21	It’s Monday. We don't need to give up.
Interviewer:	07:24	Well maybe for science.
Lisa:	07:27	So this one seems more bubbly.
Emily:	07:29	I was gonna- Mm, interesting. Cause I feel like this goes so like it just happens. This is like more… sustainable…
Lisa:	07:37	Interesting.
Emily:	07:38	…In a sense. I don't really know why… if I'm being, I mean, I kind of… I don't have a legit reason.

Speaker	Timestamp	Dialogue
Lisa:	07:45	So I've heard with like beer, if your beer has more protein, it's going to be like have more of a head, you know, like the foamy part. But like, I don't think there's a lot of protein in lemon juice. So I don't know…
Emily:	07:58	Yeah, where are you getting the [indistinct]
Lisa:	07:59	I don't know where you'd be getting that. I think this looks like a nicer cinnamon roll though.
Emily:	08:04	Yeah, it's fluffier. And I feel like a nice little like lemon flavor to the, I don't know if that'd like bake out, but like, that'd be kind of interesting. I had an orange cinnamon roll before. That's kind of random, but it was really good actually.
Lisa:	08:17	I believe it.
Interviewer:	08:18	Ooh, I bet, actually.
Emily:	08:23	And like, yeah, this one's lasting longer. I mean, it's already down here. Obviously, this one… It does have a more like beer consistency.
Lisa:	08:34	Yeah. <laughter>
Emily:	08:36	Like, with the–
Lisa:	08:37	I guess like lemon juice would have more sugar…?
Emily:	08:40	Yeah.
Lisa:	08:41	Can I look at the nutritional values? Oh, actually wait, does this part matter though? Like… Oh, they don't give you the nutritional value of vinegar, I guess there's none. Okay.
Interviewer:	08:52	Huh?
Lisa:	08:53	Maybe that's the kind where–
Interviewer:	08:53	I bought this one in a store. This isn't from like a lab.
Lisa:	08:56	I was gonna say maybe that's for like cleaning technically.
Interviewer:	08:59	Maybe.
Lisa:	09:01	[Looking at lemon juice label] Well, this doesn't have like any nutritional value either.
Emily:	09:04	This says it only has 5% acid strength. I wonder what this–
Lisa:	09:08	What do you think 5%, do you think that means like, like molar?
Emily:	09:13	I don't know. Well it's, maybe it's down to its 5%. Pfft, I dunno. [Reading label] Great use in salads.
Interviewer:	09:26	Glad they included that bit.
Lisa:	09:31	Okay. So should we say the lemon ones?
Emily:	09:34	Uh, yeah.
Lisa:	09:35	Okay. We think the lemon one would be better.
Interviewer:	09:36	The lemon one? Nice. Alright.
Lisa:	09:37	Did you bring us like the ones made from it so we can see?
Interviewer:	09:40	Oh, you know, I should have done that.
Lisa:	09:42	Okay. Um that's okay.
Interviewer:	09:45	That, that is a great idea. Um, I wasn't sure, I honestly, I debated bringing in like baked stuff 'cause I thought about how cruel is it to ask you about baking and then not have anything? Um, but then like when I was originally planning this and stuff, there was still all the mask mandates, so I was like, "Oh, I'm probably not allowed to." Um, you'll have to, just have to go home and try it, I guess.
Emily:	10:08	Yeah. Like you could bake them.
Interviewer:	10:10	You can use the gift card, money to–
Lisa:	10:12	In the dorm?
Interviewer:	10:13	–buy the stuff and then yeah. Oh, are you still in the dorms?

Speaker	Timestamp	Dialogue
Lisa:	10:17	It's a choice.
Interviewer:	10:19	Puts, makes it a little trickier to be in the kitchen, but um. So, um, just to go back to your, your point about the lemon juice one lasting longer. What impact do you think that would have on like bread in the oven?
Emily:	10:37	Well, I don't really know like the ins and outs of this, but I know that like when you make stuff, you're supposed to like let it rise and like sit there for a long time to rise. And I feel like that makes it, like, it helps it like be fluffier and stuff like that. So if we just had this, if we had the vinegar just rise for like a hot second, then it would come down. It probably wouldn't be as fluffy, it'd be a little bit more dense. I dunno if that's…
Lisa:	11:05	So like, I guess the thing that I'm not totally sure with is that I, you know, most of the time, like, you know, you have, you let your dough rise and then like if for cinnamon rolls, you knead it, roll it out. Cut, cut, cut. And then let them rise again. I don't think either of these would actually let you do that second rise. I think it'd all get reacted the first time. So like, but if we were just like, we had two doughs and we just let them rise…? I don't know. <laughter>
Interviewer:	11:33	I don't know the answers to these questions either so…
Emily:	11:40	Does the first rising have any impact on like the second rise? Like if there wasn't a second rising, like, are you still gonna be able to, if it was like a good first rise?
Lisa:	11:49	Like if you were, if we were just making like bread, if we didn't have cinnamon rolls, if we just like had bread and just like put it in the pan, just like, poof.
Emily:	11:56	Yeah.
Lisa:	11:57	I honestly, I feel like they would honestly maybe work the same, like in regards to like, if they would still be bubbly at the end of it.
Emily:	12:04	Yeah.
Lisa:	12:07	Just 'cause it's like, if we just like put them in the oven, it's like, would the faster one almost be better just because like… 'Cause it's also like hardening while it's in the oven. So like if this one takes longer. I don't think it really took longer. I think it just stayed longer.
Emily:	12:21	Yeah. I think they, they happened like the same general time.
Lisa:	12:29	Mm-hmm.
Emily:	12:33	[Swirling cups] You can kind of get 'em going again. Not obviously to the same extreme, but… The lemon one's a little bit like worse at that.
Lisa:	12:41	If you like cinnamon rolls, but you don't like yeast, you can make them with kind of like a biscuit-y dough. It's very good. I've had them too. Makes them, you can make 'em a lot faster too then.
Emily:	12:51	This one's better at that. Going, like going again.
Lisa:	12:58	So this one's in, um… We can push this one's equilibrium more. We're scientists.
Emily:	13:04	Oooh, that was good.
Lisa:	13:06	Um, I think in conclusion more research is required.
Emily:	13:12	Ooh wait, you can mess this one a little bit more. I don't know, maybe I'm making things up now.
Lisa:	13:15	This one, like still, this one actually like smells like nice. It smells like, it honestly kind of smells like, like powdered lemonade.
Emily:	13:21	That's why I'd rather have this.

Speaker	Timestamp	Dialogue
Lisa:	13:23	This one just, this one just kind of smells like water.
Emily:	13:27	What kind of water are you drinking?
Interviewer:	13:28	It doesn't smell like pickles at all?
Lisa:	13:30	No, ‘cause like, I mean, I feel like, I feel like all… ‘Cause this is just like acid and vinegar, right? Er sorry, acid and water. So like, I guess it's just making like salt, right?
Emily:	13:42	I was just smelling this, I guess, when I was smelling the very vinegar-y smell.
Interviewer:	13:45	The unreacted, yeah. Um… I don't know if this changes anything, but actually with the vinegar one, they recommend putting it with a little bit of milk. Any idea what that does? I'm genuinely curious.
Lisa:	14:00	Uh… you would make buttermilk. It would curdle the milk, but they would both curdle milk, so…
Interviewer:	14:13	And, and that's… good?
Lisa:	14:17	It depends if you want buttermilk. <laughter>
Interviewer:	14:23	Alright. Huh. Um… Do you guys know of any other yeast substitutes, or do you have any ideas of other things that might work or things you'd want to try?
Lisa:	14:40	Okay…
Interviewer:	14:41	Emily’s like, we've got two, why would we need more options? <laughter>
Emily:	14:42	Yeah. I’ll stick with these two.
Lisa:	14:44	So if you wanted to over-engineer the thing, I think there is something you can do. I've seen it like on, like I think it was, uh… I think it was Chopped where they like took their dough and they like injected air into it, and it was like, it's so over, like it's, like you would never do it at home, but you could get that fluffiness if you just like straight up injected air into it instead. I do not recommend that, but like…
Interviewer:	15:09	Just, just regular air?
Lisa:	15:10	Yeah. Like what, like I think it was like, you wouldn't be able, like you would probably just knead it all out. Like it wouldn't rise, but if you just like… Yeah.
Interviewer:	15:19	Yeah.
Lisa:	15:20	Like, I mean like, ‘cause that's basically what this is doing, right? And then you'd bake it right away. It wouldn't, I don't think it would taste very good. I think it would be chewy or something like that.
Interviewer:	15:29	Huh.
Lisa:	15:30	I don't know.
Interviewer:	15:32	What do you think's in the bubbles, um, for both of these?
Emily:	15:38	Isn't it like carbon dioxide or something?
Lisa:	15:45	CO_2, H_2O, and a salt? Well, the water and salt are not in there though. I'm just thinking. I think this is like [General Chemistry I] when we learned that reaction.
Emily:	15:59	I think I got that wrong on an exam once, so it's kind of scarred me ever since.
Interviewer:	16:03	Oh, I’m, I’m sorry to bring that up.
Emily:	16:04	What was in the bubble.

Speaker	Timestamp	Dialogue
Lisa:	16:09	Again, if you didn't want to have like the airy but you were okay with like biscuit-y ones and you think, if you have like a bunch of, a little, like, you know, when you make like a biscuit dough or a pie crust, you have all the little pieces of butter in it. And then when you bake it, they kind of like, the water in them, like evaporate away, so you would get like the air pockets that so you got croissants and stuff like that. So like if you were okay with playing around with the texture of what your finished product is, you could get air pockets without it being CO_2.
Interviewer:	16:37	You add butter instead.
Lisa:	16:39	Yeah.
Interviewer:	16:39	More butter always sounds good to me.
Lisa:	16:40	Butter, lots of little pieces. And then just like… But it would be, it'd be more of a croissant or a biscuit.
Emily:	16:49	A croissant cinnamon roll would be really interesting.
Lisa:	16:50	They're called, called morning rolls.
Interviewer:	16:53	Oh, they exist!
Emily:	16:54	There you go.
Interviewer:	16:57	I'm gonna like, forget this interview, just like tell me everything about cinnamon rolls. What do I need to try? Um… So, uh, one thing that I can never keep straight, uh, is the difference between baking soda and baking powder. Do you think baking powder would work here? Any, any ideas?
Lisa:	17:20	You’re smart. Okay. Wait, are you, are you trying to give us a hint? Do you know the answer to this?
Interviewer:	17:27	No, I, I genuinely don't. Um, and, and like I said, not after a right answer. I know you probably like, that’s probably like first day of food science class, but I, I don’t know.
Emily:	17:33	You definitely probably know. I think the powder wouldn’t work.
Lisa:	17:38	So. Okay. So powder, I think it would work, but you wouldn't add acid to it because baking powder is actually, um, I don't think it's like, I dunno if you'd say it's like pH neutral, but it can react with itself because it's baking soda plus other stuff. Um… So that's why it's like, I think if like, um, if you, so that's why you can't substitute the two, ‘cause if you had something that called for baking powder and you just added baking soda to it, it wouldn't have, there wouldn't be anything acidic in like the batter to react with it to make the bubbles. But if you… But yeah. Am I making sense?
Interviewer:	18:18	You're, you're saying baking powder already has an acid and a base?
Lisa:	18:20	Baking powder has baking soda in it. Yeah.
Interviewer:	18:21	But it, but it doesn't react just sitting on your shelf?
Lisa:	18:28	Uh, it's powdered. I think what I've seen is that if you wanna substitute it, you use baking soda and like… I wanna say like some sort of starch, but that doesn't make sense… Man. I’ve only had one semester!
Interviewer:	18:52	Oh, yeah, it's okay. Like I said, you probably know way more about this than I do. Uh…
Lisa:	18:56	No, but baking powder has baking soda in it. I know that.
Interviewer:	19:04	All right. Um… So one other suggestion, um… I don't know, have you guys ever had beer bread? I love beer bread. Um…
Lisa:	19:22	I have a very unfortunate story about that. The research doesn’t, it doesn’t pertain to research.

Speaker	Timestamp	Dialogue
Interviewer:	19:29	After the camera's off. Um… So, so for that you, you add beer to the batter, um… To the, I, I, I go your route where I buy all the powdered stuff together. I recently learned that that's like super expensive compared to like just adding the flour and salt yourself. Uh, uh, but for that, I think the, the beer is used to get, get some of that rise? Correct me if I'm wrong. Um, so do you think any like alcohol adding, adding to your batter would, would work? Um, beer or, or say like vodka or something? Do, do you think those would do the same thing, make the bubbles?
Lisa:	20:20	Is it like just straight or is it like with baking, like with–?
Interviewer:	20:24	We, we can do with baking soda.
Emily:	20:29	Um, I mean, I feel like everyone talks about how like alcohol is like super acidic or like—
Lisa:	20:38	Is it?
Emily:	20:39	Not like super acidic, but like…
Lisa:	20:41	I thought it was like, I thought it was like water.
Emily:	20:44	Oh. I don't know. I thought I had some… acidic properties. I don't know if that's necessarily a hundred percent sure.
Lisa:	20:58	Wait, no, keep going, keep going. Sorry. I'm just like, I'm doubting myself.
Emily:	21:01	Maybe I just am…
Lisa:	21:03	It's, it's protic.
Emily:	21:05	Yeah. I'm, I'm 21, so I don't want this to sound weird, but like, I know that after I've drank, my stomach has been like really funky after, and that's like, everyone's like, oh, it just feels like acidy and kind of like gurgly and stuff like that. And not even like a, like after like two beers or something like that, so I feel like I've always heard people say like, it's kind of like acid-y. But maybe that's just beer. 'Cause I…
Lisa:	21:31	I think that might be the carbonation in the beer though. Like, 'cause like, do you feel that way after like…? Like I, you can drink with your parents in this state. I like, I never noticed that…? And I've had like, like I guess I've had a mojito and that's carbonated, so like, I don't know.
Emily:	21:57	But is that from… Is the carb–, I don't know. I'm making stuff up. I don't know that much about alcohol.
Lisa:	22:08	Well, you did just recently turn 21 so I would expect that.
Interviewer:	22:11	Um, for obvious reasons I wasn't allowed to bring any in to, to do a third experiment.
Lisa:	22:21	I feel like, um, I think the reason beer bread works is because of the, um, just the carbonation that's already in there. So I don't know if like soda or like sparkling water would work necessarily or a kombucha, but I don't think it's actually the alcohol. But I could be wrong 'cause I have not tried this.
Interviewer:	22:46	Fair.
Emily:	22:46	Yeah, I don't really know.
Interviewer:	22:52	Alright. Uh, any other ideas or thoughts related to, to this?
Emily:	23:02	I'm gonna look up if alcohol's acidic or not.
Lisa:	23:05	Okay. I feel like, oh! Well… Okay, 'cause I'm, I'm think, I'm trying to think now because um… Isn't it like, uh, you took micro, right?
Emily:	23:20	Yeah.

Speaker	Timestamp	Dialogue
Lisa:	23:21	Okay. So like isn't it like when you have like, um, lactose fermenters, they make lactic acid and the CO_2, H_2O, salt. So I think, I think you're right. I don't think alcohol itself, like just like ethanol is acidic, but I think like beer–
Emily:	23:37	It's fermented.
Lisa:	23:38	Or yeah. Like I think that would be, I forgot about that. So I think, I think you're right with that like [inaudible] like not pure bleach-your-insides alcohol would be.
Emily:	23:47	Right. Well, and not that this is like, actually never mind, that's a different thing. I don't know. I think it maybe you'd have to like specifically have like the fermented one.
Interviewer:	24:04	Gotcha.
Emily:	24:06	That makes sense. I feel like.
Lisa:	24:10	Yeah.
Emily:	24:13	Cause isn't that…? Mm, no.
Lisa:	24:18	Mm. Would… do you think we had yogurt mixed with something would work?
Emily:	24:24	I've heard of that as substitutes, but I don't really know why. Pfft.
Lisa:	24:28	'Cause yogurt is acidic 'cause of the lactic acid.
Emily:	24:32	Yeah.
Lisa:	24:34	Right. So…
Emily:	24:35	But do you think it would get as much of a rise or is it just like a…?
Lisa:	24:37	I don't, I don't know, I just, this still might make like a brick, but like, it might be a little less dense maybe. Um…
	24:59	Um, do you have several days to make this bread?
Interviewer:	25:04	Uh, hypothetically, sure.
Lisa:	25:06	Okay. 'Cause isn't it another thing too, you can do like wild fermentation. So you could just let your dough sit out by the window sill for a while until wild yeast infects it. Is that a thing you could do?
Interviewer:	25:16	That's a thing?
Lisa:	25:17	I think so. It's not really recommended… anymore.
Interviewer:	25:19	Wild yeast? Is it just like in the air?
Emily:	25:20	That sounds kind of… dangerous.
Lisa:	25:21	But it's like, it's kind of, like it's kind of like sauerkraut, you know? You don't like add stuff to sauerkraut, you just let… the stuff in the environment get into the sauerkraut.
Emily:	25:29	I don't think I could be a food science major. I think I would never want to eat again.
Lisa:	25:33	Um, I think like if you had several days, but that would technically just be using yeast but not the prepared kind. Also it's cold outside. I don't know how much yeast is floating around.
Interviewer:	25:44	It probably wouldn't be great right now in Wisconsin, but maybe, maybe further south? Wild yeast, hmm…
Emily:	25:52	I wonder if like a cranberry juice would be like…
Lisa:	26:03	So sour.
Emily:	26:04	I know, but I wonder if it would like, there's like a range of like juice concentrates that you would like… 'Cause I feel like, I think of like lemons and limes and or, and like oranges is super acidic versus like–
Lisa:	26:18	Oh yeah, lime juice would probably work.

Speaker	Timestamp	Dialogue
Emily:	26:20	Yeah. I don't really think of like… grape juice as necessarily like as acidic as this, but maybe it is.
Lisa:	26:29	I feel like cranberry, like, because like, you know, when you always get bottled cranberry juice, there's always a bunch of sugar added to it. So, so it is very acidic, so I feel like cranberry juice could work, but I feel like, is it like… I feel like, I think that would work? It's sour. It's very, it's very sour. Once in my food science class, we had to taste stuff, and my professor added like a bunch of citric acid to apple juice but looked like apple juice. It was a very sad way to pop in the morning and taste like that.
Interviewer:	27:03	I'd probably have trust issues with that professor after that.
Lisa:	27:06	Yeah, it took, to be fair, he was like, "Rank the sourness," but like… Tt was not great. <laughter>
Interviewer:	27:16	Yeah, bet that didn't taste real great with the coffee. I don't know if you're a morning coffee drinker.
Emily:	27:25	I wonder if coffee would work. Is coffee acidic?
Lisa:	27:28	I think coffee's alkaline.
Emily:	27:30	Oh… I feel like I've always heard about it like breaking down your teeth enamel.
Lisa:	27:40	I think it just stains your teeth. I think soda breaks down your teeth enamel. Or I'm not sure though. Now I'm doubting myself again. I'm just like, cause that coffee's bitter, right, though, and like bitter stuff is acidic? Er, ach, not acidic, basic, right?
Emily:	28:01	Yeah.
Lisa:	28:01	Like there's not really a lot of basic foods. Ludivisk is.
Emily:	28:06	Ew.
Lisa:	28:06	Yeah. There's also this like fermented soy bean thingy that's like Japanese. It's very bitter. Except you look at it, it looks like it's going be kind of like sweet. It's not. I've never tried it. It really smells apparently.
	28:28	Yeah, I'm not sure. Sorry, what was the question?
Interviewer:	28:30	Oh no, this is great. Um, I just wondered if you guys had other ideas and you did. Um…
Emily:	28:40	Yeah.
Interviewer:	28:42	So bottom line, um, you would eat bread made with either of these? If you were baking it, you'd go with the lemon juice. Do I have that right?
Lisa:	28:55	Probably although I'm more likely have vinegar on hand than lemon juice because I believe in like fresh lemons, not bottled lemon juice, so…
Emily:	29:06	Yeah.
Interviewer:	29:08	Okay. All right. Um… Anything else to add before I turn things off? I don't wanna cut you off if there's more you want to say.
Emily:	29:20	I think I'm good.
Lisa:	29:24	I think I'm good too.
Interviewer:	29:25	Lisa's like, "I have so many ideas."
Lisa:	29:28	Um, yeah. I feel like it's just kind of hard to think like what you normally have in your pantry and then like, without like seeing it and going through and eliminating what you think would work.
Interviewer:	29:37	Fair.

Speaker	Timestamp	Dialogue
Lisa:	29:39	I feel, well… I mean like buttermilk is acidic, so like maybe you could try to do something with that, with like that and like baking soda again. Yeah, I don't really have any other ideas.
Interviewer:	29:57	Sounds, sounds good.

Table A.S4. Transcript of interview with Inaya and Saanvi.

Speaker	Timestamp	Dialogue
Interviewer:	00:04	Okay. So like I mentioned in the emails, uh, the thing I want to talk about today is baking. Uh, do either of you like to bake, have much experience with it?
Inaya:	00:13	I do.
Interviewer:	00:14	You do? Anything in particular you like to bake?
Inaya:	00:18	I like cakes and cupcakes mostly... Yeah.
Interviewer:	00:23	Um... Do either of you watch any of the, the baking shows on Netflix or elsewhere?
Saanvi:	00:31	Yeah, sometimes. I'm like blanking on the name, but there's one Netflix show with beginning bakers. Do you know what I'm talking about?
Interviewer:	00:42	Oh! I, I think I do.
Saanvi:	00:45	Okay. Well that's one of my favorites.
Inaya:	00:48	I used to watch Cake Boss but not much anymore.
Interviewer:	00:52	Nice. Um... Well, I'm not that much of a baker, I'm mostly like out-of-the-box kind of thing. Um, just dump in an egg and stuff. But um, my sister had this really great cinnamon roll recipe, and I thought I would try it and make some homemade cinnamon rolls. Um... And I found out my yeast had gone bad, um. So you guys know what yeast does? [Nodding] Yeah, so these cinnamon rolls, um, came out so flat. It was so sad. I tried eating them anyway 'cause I was like, "They're, they're covered in cinnamon sugar, like how bad can it be?" Um, but ugh, no, that texture was just way off. It was, it was pretty sad. Um, so I, I looked up, um, some like yeast substitutes on the internet, and um, some of the options that they suggested are like vinegar and baking soda or lemon juice and baking soda. Um, and so my question for you guys is, um, first of all, do you think these would work? Um, and if so, which one do you think would be better? Which one would you want to use if you were making bread or cinnamon rolls or whatever you wanted that involved yeast?
Saanvi:	02:13	I don't know about you, but I feel like I would go with maybe the vinegar and baking soda combination just 'cause like, you know, middle school science projects, with the volcanos with baking soda and vinegar. Umm, I guess–
Interviewer:	02:29	I did bring these in in case you do want to mix them together and have a little fun with it.
Saanvi:	02:34	Yeah [crosstalk]
Inaya:	02:37	I don't know about the lemon juice because it's going to give it like that real like citric taste, and it's just not gonna, I don't know.
Saanvi:	02:45	For cinnamon rolls?
Interviewer:	02:46	Or, or something else.
Saanvi:	02:46	I mean, unless you do like a citrus-y, like I don't know, like a lemon-flavored cake or something, then like...

Speaker	Timestamp	Dialogue
Inaya:	02:53	That could work. But then it wouldn’t be like, mm, I feel like it would be like... not sour, what’s the word?
Saanvi:	03:00	Tangy?
Inaya:	03:01	No, like... I’ll tell you when I think of the word.
Saanvi:	03:05	‘Cause I feel like if we’re thinking about flavors, vinegar might taste a little weird. So I don’t know. Can we choose just baking soda? Is that an option?
Interviewer:	03:18	Sure, yeah.
Saanvi:	03:19	‘Cause, okay, ‘cause I know in the, okay, I always get confused between baking soda and baking powder, but I know in the past, um, when I was trying to make like pancakes, and I wanted them to be fluffier so I added like a little bit of baking powder, and it did the trick. So I don’t know if that’s an option but...
Interviewer:	03:43	That’s certainly an option.
Saanvi:	03:44	Do you wanna...?
Inaya:	03:45	I want to see how much it goes with the lemon ‘cause I’ve done it with vinegar but not lemon.
Saanvi:	03:54	Do you think that’s good?
Inaya:	03:55	You should put more.
Saanvi:	03:56	More?
Inaya:	03:56	Yeah.
Saanvi:	03:57	I don’t want it to overflow.
Inaya:	03:59	It won’t be that bad, I don’t think. It’s just a little bit.
Interviewer:	04:03	I’ll stop you if you put in enough that it’s gonna like hit the ceiling. <laughter>
Saanvi:	04:11	Not very scientific measurements for accurate comparison.
Inaya:	04:18	That’s it? Okay, I’ll put more drops in. Okay, so it is fizzy. That’s fizzing up a lot, maybe more than the vinegar.
Saanvi:	04:32	I think that one stays bubblier for a longer time. So maybe lemon might be a better option just ‘cause like... you know...
Inaya:	04:47	It’s like solid.
Saanvi:	04:47	...you want it to stay bubbly. Wow, this smells terrible. <laughter>
Inaya:	04:57	Yeah, you’re right, I think you should go with lemon.
Interviewer:	05:02	Because the bubbles last longer?
Saanvi:	05:04	Bubbles last longer, yeah.
Inaya:	05:06	It might give it a more fluffy... I don’t know, like look, I guess.
Saanvi:	05:11	I wonder, and like, if you stir it around like when you’re mixing a batter or something, it’s not, it seems like, least likely to, yeah, pop the bubbles or whatever.
Inaya:	05:26	Put this [inaudible]. What the heck? <laughter>
Saanvi:	05:27	So I guess, just be very quick and fast with your process of shoving it in the oven.
Interviewer:	5:34	Before the bubbles go away? Um... Well you kind of already got into this, um, but I was going to ask if you guys had any other good ideas for yeast substitutes. You mentioned like just the baking soda. Um, anything else that you think would help give it that rise?
Saanvi:	05:56	Um, you could, I don’t know, try like beer or um, carbonated water or something. I don’t know if you add water, do you add water too when you bake a cake? No, okay.

Speaker	Timestamp	Dialogue
Inaya:	06:08	No. But I see where you're coming from with the carbonated water, that might like.
Saanvi:	06:14	Yeah. 'Cause you know with like the tempura batter, they use like beer or soda or something like that to give it the fluffiness. But I know that's more of a crispy texture versus [inaudible].
Inaya:	06:29	I've seen videos where people, um, make like three-ingredient cakes with like soda, and that also kind of makes it fluffy. I've never tried it. I don't know if it works or if it's like a fake video. But I think like the bubbles, the carbonated stuff maybe will give it that rise too.
Saanvi:	06:46	Yeah, I've seen like the Sprite pie. I don't know if you've heard of it. But they just, it's like a store-bought pie crust and then you dump in a can of Sprite and then like cake, box cake mix that you just sprinkle in, you don't even stir it, and then like slices of butter on top and bake it.
Interviewer:	07:04	And that, that makes a pie?
Saanvi:	07:06	It makes a pie apparently.
Interviewer:	07:07	Oh. I've never heard of this. I'm, I'm very curious now.
Inaya:	07:13	I've heard it tastes a lot better than it looks. Like it looks absolutely disgusting but yeah. <laughter>
Saanvi:	07:26	Yeast is like a living organism, right? So there's no way you can like revive it?
Inaya:	07:33	[Inaudible]
Interviewer:	07:36	I did not try to do CPR on my bad yeast. I gave up on it. <laughter> Um... Uh, so... Uh, you mentioned, you mentioned beer because it's carbonated. Do you think, um, some other like alcohol would work? Like a wine or like a vodka or something? Or do you think it's that–?
Saanvi:	08:00	I was thinking beer because they use yeast to make it.
Inaya:	08:04	Beer is fermented yeast.
Saanvi:	08:04	It's fermented, yeah so. Maybe you could try kombucha. I don't know if that would work as well. Um... This is totally out of the box, but what if you put sauerkraut or kimchi? <laughter>
Inaya:	08:18	Oh my god.
Interviewer:	08:19	Oh, 'cause those are both fermented? Would you [inaudible]
Inaya:	08:25	[inaudible] cinnamon roll.
Saanvi	08:29	'Cause there are living organisms in that that would eat the sugar in your cake, right?
Inaya:	08:37	True.
Interviewer:	08:39	In, in principle that, that sounds like a good idea. I think I would also be a little hesitant to try it. Um, so one thing I, I didn't mention, with the vinegar, the internet – um, the all-knowing internet – recommended you add it with a bit of milk. Any idea what the milk does?
Saanvi:	09:08	Is it 'cause the milk has lactose which is a sugar, oh no wait, no, that makes no sense.
Inaya:	09:15	Doesn't it curdle in the vinegar? Like–
Saanvi:	09:18	That's what I was thinking. Like wouldn't you have a cheese?
Inaya:	09:18	Right? 'Cause that's what happens with lemon.
Interviewer:	09:22	With, with lemon in milk?
Inaya:	09:23	Yeah, 'cause, I don't know, you get like–

Speaker	Timestamp	Dialogue
Saanvi:	09:25	Like you know paneer, like the cheese? Like it's usually milk, well I guess you boil it first, bring it down to room temp, and then you add vinegar or lemon juice which curdles it. I don't know if it would give your cake a chunky texture, but I don't know why they'd recommend that if that was the case. <laughter>
Interviewer:	09:49	Yeah, I... I was a little hesitant. Um, I don't understand, why?
Inaya:	09:56	Did you try it?
Interviewer:	09:57	Um... no. I instead, because this was like a, it was a fairly lengthy process to roll out these cinnamon rolls, and I just really wanted cinnamon rolls, so instead I just went to the store and bought some, um, I'll admit.
Inaya:	10:10	That's fine.
Saanvi:	10:11	Were they like the Pillsbury rolls ready to bake ones?
Interviewer:	10:15	Yes, and, and those are good too. They, they might not be as good, but like... I just, I was craving them. So yeah, I haven't had a chance to actually experiment and try. Um... and in another interview, they were like, "Wait, are you going to pull out cinnamon rolls made with each?" And I was like, "That would have been a great idea. Wish I thought about that, but sorry, no, I don't have any for you to taste." <laughter> Yeah, uh... So... um, any idea what's in the bubbles?
Inaya:	10:55	[pointing at cups] Oh, you mean like, in the– [crosstalk]
Interviewer:	10:57	Yeah, sorry, that you, you made when you put these together.
Saanvi:	10:58	Carbon dioxide?
Inaya:	11:01	That's correct, yeah.
Saanvi:	11:02	Okay. Thank you. <laughter>
Inaya:	11:05	No, I remember doing a science fair project on this. They like blew up balloons with the, um, air that came out and it [inaudible]. Um, yeah I think it is carbon dioxide, the bubbles.
Interviewer:	11:18	Nice, so I, I just ask 'cause you talked about some of the, the yeast or other organisms, stuff that can produce the gas, um, and then here, you don't have any, um, any organisms. Well, there's probably some, but not involved in this, um, so I was curious what you thought was going on there. Um... Any, any other ideas of what you could maybe mix together to, to make those CO_2 bubbles? If not, that's totally fine.
Saanvi:	11:56	I know in our OChem class, we've talked about a couple of reactions that have CO_2 as a byproduct, but I can't remember the exact reactants. But I also don't know if that's safe to consume so.
Interviewer:	12:09	Probably not. <laughter> Most of the time that answer's no.
Inaya:	12:15	This is more of like a baking perspective. I don't know how well it would work. But if you take egg whites and whip 'em up, usually that gives you like a super fluffy consistency, and kind of like mix that into your batter. I don't know if it would still be like... cinnamon roll batter 'cause it would be super hard to like roll out and like move and stuff. But I do know that like whipped egg whites give you that super fluffy consistency, and it makes your cakes a thousand times better.
Interviewer:	12:45	Um... And, and what does the, do you know what the like whipping the egg whites does to them to make them fluffier?
Inaya:	12:55	I'm not sure. Go–
Saanvi:	12:56	In the cooking shows, they always say, "Adds air to it." So I guess it adds air.

Speaker	Timestamp	Dialogue
Inaya:	13:02	It goes from like this gooey consistency to like, almost like meringue, I wanna say. 'Cause it, it–
Saanvi:	13:08	Like a whip cream sort of texture.
Inaya:	13:10	Even lighter than that. It has like a lot of air in it. Like if you have like one egg white, it literally goes up to like this big. And you just like fold it in, and it, it'll be really good.
Interviewer:	13:20	I might have to try that. Um, getting data for my research and for my personal life. Um... <laughter> Yeah, interesting, just getting, getting more of that air in there. Any... any other thoughts related to this? Um, other things you'd want to try? Or... That's very open-ended so.
Saanvi:	13:52	I feel like this is very much a reach, but you know, you can make like a, a pseudo-cinnamon roll with a tortilla and put cinnamon and frosting, roll it up, cut it up. And I don't, I believe tortillas don't require yeast. I think it's just like flour and water and something like that, I don't know.
Interviewer:	14:15	Sounds reasonable.
Saanvi:	14:15	So I guess you could make like homemade tortillas and then it would [inaudible] that process. Yeah.
Interviewer:	14:21	Yeah. Yeah. Put a little twist.
Saanvi:	14:26	I think the texture would be totally off but.
Interviewer:	14:29	But if you're not expecting super fluffy, maybe that's, that's tasty. Cinnamon-roll like. Or, what'd you call it, pseudo?
Saanvi:	14:39	Pseudo, yeah. <laughter> I don't think I have any more ideas.
Interviewer:	14:48	Fair enough. Well those are, those are actually all my questions.

Table A.S5. Transcript of interview with Mandy and Denise.

Speaker	Timestamp	Dialogue
Interviewer:	00:27	Alright. So, uh… the topic I wanna talk about today, uh, is baking. Do either of you like to bake?
Mandy & Denise:	00:37	Yeah.
Interviewer:	00:38	Yeah. What do you guys like to bake?
Denise:	00:40	I like baking brownies.
Interviewer:	00:41	Ooh.
Mandy:	00:43	Uh, I like basically making anything like cake, cookies, muffins, y'know.
Interviewer:	00:49	All that good, good baked good stuff. Um, do you, do you also like watching any of the, the baking shows?
Mandy:	00:58	I really like, um, the Great British Baking Show. That's my favorite TV show.
Denise:	01:04	I've watched like one episode of that. <laughter>
Interviewer:	01:08	I got really into that one a couple years ago, and now me and my friend are like, "Oh, it's out! We gotta watch this!" Um… And I'm just like, "Ugh, I wish there were people in my life who baked that much." I, I, I don't mind baking, but I like the eating better. Um…
Denise:	01:24	Yeah, me too. <laughter>
Mandy:	01:26	I like baking so I can eat afterwards.

Speaker	Timestamp	Dialogue
Interviewer:	01:28	Yeah. Gotta have that reward. That's why, you know, people are like, “Chemistry is just like baking,” and I'm like, “No, you can't eat your chemistry.” Um, most of the time. Um… So like I said, I, I am not that great a baker. Um, a couple weeks ago, I tried to make some cinnamon rolls, um, from scratch because I was feeling ambitious and my sisters had done it and I was like, “Well, if they can do it, I can do it.” Um, and it turned out my yeast had gone bad. Um, so, you guys know what yeast does?
Denise:	02:07	Yeah. To make it rise, I think.
Interviewer:	02:09	Yeah, so these were like just very flat, very dense cinnamon rolls. And I tried to eat them anyways because I was like, “There’s cinnamon and sugar, like how bad can they be?” And I was like after, “Eh, no.” It was just not the same without that texture. Um, so I was looking up since apparently my yeast was bad. Um, I was looking up some yeast substitutes, um, and the internet, um, said you could use like, uh, vinegar and baking soda or lemon juice and baking soda. So I'm wondering if you guys think either of those would work, and if they both work, which one do you think would be better? I, I, I brought in this stuff in case you wanna play around with combining them and and seeing what happens, so feel free, figure out which. <laughter> What's gonna rescue the cinnamon rolls?
Denise:	03:01	I feel like maybe vinegar because it like bubbles.
Mandy:	03:05	Yeah, ‘cause I’ve seen the stuff of it with it bubbling up.
Denise:	03:06	Yeah. And…
Mandy:	03:08	Like volcano stuff with that, making it bubble over [inaudible].
Denise:	03:14	Yeah. Okay. We could do one with lemon too. Okay. You wanna put the same amount for each one?
Mandy:	03:28	Sure.
Denise:	03:29	Just do like…?
Mandy:	03:31	How much should we do? Like a spoonful?
Denise:	03:34	Maybe like two.
Mandy:	03:36	Two spoonfuls?
Denise:	03:36	Or three.
Mandy:	03:37	Okay. Should I start with two? I should do another one.
Denise:	03:47	Do another one.
Mandy:	03:48	Yeah, let’s do another one.
Denise:	03:54	Okay. Okay, you can see what this one’s at. So it's kind of even. Um, you wanna dump the same amount?
Mandy:	04:18	How much do you wanna put in? I feel like this one.
Denise:	04:24	Um… What is, wait let me read this. [Reading bottle] ‘Cause that's acidic. What is this?
Mandy:	04:38	That’s acidic too. But is that more acidic than this?
Denise:	04:41	I think that's more acidic.
Mandy:	04:43	You think that’s more acidic?
Denise:	04:43	Yeah.
Mandy:	04:43	What is in lemon juice?
Denise:	04:45	This is diluted with water.
Mandy:	04:46	That one’s diluted?
Denise:	04:47	What is this?
Mandy:	04:48	Just lemon juice.

Speaker	Timestamp	Dialogue
Denise:	04:48	It's just concentrate. A hundred percent.
Mandy:	04:54	It’s got citric acid.
Denise:	04:55	Okay. We can do like… Um, you wanna do this one, Right? We can do like…
Mandy:	05:08	We could just do it until it starts making stuff happen.
Denise:	05:11	We could just pour until that line or something. Like pour it until it fills this line, or do you think that’s too much?
Mandy:	05:18	I mean I feel like that would be a lot, but something would happen.
Denise:	05:22	Let's just put it up to this line so that they're both even for amounts.
Mandy:	05:24	Okay, okay.
Denise:	05:25	Okay.
Mandy:	05:31	Oh, I can't even tell how much I have.
Denise:	05:33	I can't either.
Mandy:	05:36	I just stopped pouring.
Denise:	05:39	Oh, interesting.
Mandy:	05:40	Look at that…
Denise:	05:42	I didn't even put that much in it.
Mandy:	05:44	Oh, I didn’t either. Oh look, I’ve got a really good amount in there.
Denise:	05:49	Are we supposed to tell you our conclusion or…?
Interviewer:	05:52	Yeah, just tell me what you're thinking. You like one better than the other?
Denise:	05:54	Um, well I feel like the lemon juice probably would work better. I feel like the lemon juice probably–
Mandy:	06:03	Look, it's still going.
Denise:	06:04	Yeah. It probably has a slower reaction. Uh, I don't know. I feel like maybe vinegar reacts faster and so that’s why it’s…
Mandy:	06:19	I’m just mixing it up.
Denise:	06:21	Okay. Um, well, I feel like we both agreed that this lemon juice is gonna work better because… um…
Mandy:	06:33	Still got the, the bubbles going for it.
Denise:	06:37	Yeah, and I feel like maybe the vinegar kind of like dissolves it right away whereas like the lemon juice, it like reacts with it, but doesn't like dissolve it as fast. Or like the baking soda, maybe? I don't know if it's like, um…
Mandy:	07:05	I think the fact that there's still bubbles in this one, that it would make more sense than to just need, to just be adding like liquid into cinnamon rolls.
Denise:	07:15	Do we have to like give you like some like chemistry background?
Interviewer:	07:18	Oh no, no, you can just explain it whatever way, like makes sense to you.
Denise:	07:25	Well, I mean, this could be since, I feel like this is more acidic, so it has more resonance structures.
Everyone:	07:32	<laughter>
Denise:	07:34	And like, maybe it's like more stable, so it's not as reactive.
Mandy:	07:38	What's, what’s baking soda. What is that real one's name?
Denise:	07:42	[Laughing] I don't know.
Mandy:	07:43	Oh my god.
Denise:	07:43	Oh, wait. Is it…hmm?
Mandy:	07:47	Just, is it carbon…?
Denise:	07:49	Car–, carbonate? No. No, it’s not.

Speaker	Timestamp	Dialogue
Mandy:	07:50	Isn't it sodium…?
Denise:	07:52	Sodium carbonate?
Mandy:	07:55	Bi… <laugh> I don’t know.
Denise:	07:57	I don't know. Well, anyways, um, yeah, that's our conclusion. <laugh>
Interviewer:	08:03	Alright. So the lemon juice had more bubbles. Um, it's gonna make better bread. Um, and you think that might be ‘cause it's more acidic.
Denise:	08:11	Yeah, probably.
Mandy:	08:13	Mm-hmm.
Denise:	08:13	It's like, wait, if it's more acidic, is it more reactive or no?
Mandy:	08:16	<laughter> Um… the lemon juice?
Denise:	08:22	I forgot. Well, I feel like if it has resonance, it's gonna be more stable. Right? Pfft.
Mandy:	08:30	Yeah.
Denise:	08:31	Okay. <laughter> Um, yeah, that's our, I guess, our conclusion.
Interviewer:	08:40	Alright. Do you think you would be willing to try like bread made with either of these or, or just the lemon juice one?
Mandy:	08:48	I’d try with both. Maybe start with this one, but I, the lemon might make it taste funky is my only thing.
Denise:	08:54	I like lemon flavored things, so I feel like I would do the lemon.
Interviewer:	08:58	You're like, “Added bonus!”
Mandy:	09:00	But if you’re trying, if you’re trying to make like cinnamon rolls with like lemon juice…
Denise:	09:03	I feel like the heat, like from the oven might kind of affect it too.
Mandy:	09:07	The flavor? Yeah…
Denise:	09:07	Like make it, or maybe like make it react faster, so it won't be as like, you know? Like compared to yeast, I feel like it might have a different effect. ‘Cause heat like speeds up reactions. <laughter> So… I don't know, I feel like I would just use the lemon juice just for flavoring. I don't know if it would really help as much for rising, for making it rise. It could for a little bit.
Mandy:	09:34	I mean, yeah, [inaudible] a lot of bubbles and bubbles mean gas and gas means air can rise.
Denise:	09:39	Wait, well actually I feel like you, like don't you put the yeast in before you bake it? To make, you like you–
Mandy:	09:46	You soak it in water first, don't you? The yeast in water first and then you put it in?
Denise:	09:49	‘Cause I know like my sister, she like makes cinnamon rolls and she like, lets it rise first before baking. So maybe it would rise first and then like baking it wouldn’t really affect it as much.
Mandy:	10:01	Cause isn't it like when you, yeah, ‘cause yeast will rise, and then when you, when you bake stuff with yeast in it, doesn't that, that does another process of something.
Denise:	10:12	Of what?
Mandy:	10:12	I don't know, does it make it rise more, does it make it stop rising or something? Has to rise slightly more, I guess.
Denise:	10:19	Uh… I don't know.
Mandy:	10:20	But baking soda anyway makes stuff rise. So…
Denise:	10:24	I guess, yeah.
Mandy:	10:27	That’s why [inaudible] put it in muffins. That and like baking powder. I don't really know the difference though.

Speaker	Timestamp	Dialogue
Denise:	10:32	Interesting.
Interviewer:	10:32	Oh, I was just gonna ask you that actually. Um, yeah, ‘cause, ‘cause these weren't the only, um, options. These are the ones I remember from the internet. Um, but I was wondering if you guys had any other ideas of what might work. Um… And, and I was gonna ask about baking powder too ‘cause also white powder.
Mandy:	10:52	Yeah, I don't, I don't know the difference. I know that some things just require baking soda, and some require both. Some just require baking powder.
Denise:	10:59	Mm-hmm. I don't really know the difference either. I guess I don't bake that much. Like I usually just get like boxed stuff and make it. But like sometimes I make stuff from scratch, but um, usually I look up a recipe.
Interviewer:	11:13	Yeah, and you’re just like, “Okay, baking powder. Let's…” I do the same thing.
Denise:	11:15	I feel like if you like, like I now feel like whip it, like whip the batter or something with like a fork, it like, like creates more air or something in it. So it like…
Mandy:	11:25	Gets fluffy. Yeah.
Denise:	11:26	Yeah. I guess that might be a little different than like rising, I don't know.
Mandy:	11:31	I know if you over-beat stuff though, it'll make it like tough.
Denise:	11:34	Oh really? I didn't know that.
Mandy:	11:35	I’ve done that to muffins before. They were not good.
Denise:	11:39	Interesting.
Interviewer:	11:43	Anything else, like in your kitchen, that you would wanna try that you think could maybe like, make, make the bubbles to give it a good rise? Any.. Say you didn't have vinegar or lemon juice. Or if you didn't have baking soda…
Denise:	12:03	Um… Maybe like carbonated water? <laughter> I don't know.
Mandy:	12:09	Mm, yeah, we could do carbonated water. Oh yeah, soda.
Denise:	12:11	I mean there's, there's like gas in there.
Mandy:	12:14	Don’t people put like… soda in things?
Denise:	12:17	I don’t know.
Mandy:	12:20	I feel like I’ve heard that before. But if you put sprite in something.
Denise:	12:20	Yeah. Like carbonated water or like soda probably could do something. I mean these like produce gas, so I feel like that if you use these that you might as well just try the soda.
Interviewer:	12:33	Yeah. I’ve, I've never baked soda bread <laughter> um, but that's a thing, so… I, I don't know if that's a name from like baking soda or actual soda but you might be on to something.
Mandy:	12:41	Yeah, that’s where I heard of it, yeah.
Denise:	12:45	Or you, like can’t you use beer or something?
Mandy:	12:48	Yeah, my friend’s made beer bread before.
Denise:	12:50	Yeah. Maybe can use beer ‘cause also that's like fermented too so it might have like yeast. I don't know. Or like, I mean, I know it probably has like gas or something.
Mandy:	13:04	Yeah. Beer's a good one.
Interviewer:	13:07	Any idea what that gas is? Out of curiosity?
Denise:	13:10	Like, um, probably… I don't know, carbon dioxide?
Mandy:	13:15	Yeah, that’s what I was thinking, carbon dioxide, I dunno.

Speaker	Timestamp	Dialogue
Interviewer:	13:19	Yeah. Sounds reasonable.
Denise:	13:20	We're like both in microbiology too. And we should, we should know this. We were both in microbiology.
Interviewer:	13:27	You guys probably know more than me about that stuff. I never took the biology side of things, so… Um, so yeah, I honestly don't know how yeast works really. Um… And that's why I feel like, um, I don't know about you guys, but people like expect me to be able to, 'cause I'm a chemist. They're like, "Oh, can you explain how this works?" And I'm like, "Um… nope. We don't use any ingredients, uh, like this in the lab. Like it's all stuff you're not allowed to buy. Um, so, um, just saying I, I'm no expert in this either. Um.
Mandy:	14:07	I learned one thing from watching the Great British Baking Show is that when you're adding like your yeast and you're like making bread or something and you add like salt, you should add them on like different sides of the bowl or something. Because if you like just pour them on top of each other, it'll like, it'll either like start the yeast, like activating it or whatever, like earlier than you want or something like that.
Denise:	14:28	Hmm. Interesting. I didn't know that.
Mandy:	14:32	No, yeah, I learned that.
Denise:	14:32	I guess you learn a lot from watching TV. <laughter>
Interviewer:	14:37	Yeah, well. I never used to like watching that stuff 'cause I'd always get so hungry, and then I figured out if, as long as I made something beforehand, then it was fun. Um… still jealous but. Um… So you mentioned uh, beer maybe, um. And maybe 'cause there's some yeast in there, maybe it's like kind of fermented. Um, do you think any like other sources of, any other alcohols would work? Like, I don't know, like a wine or like a vodka or something like that. Do you think…? Those get used sometimes in baking stuff.
Denise:	15:14	I think maybe wine. Wine's also fermented. I don't know about vodka. I don't know <laugh> if that's fermented but… probably like wine and beer.
Interviewer:	15:30	Yeah.
Denise:	15:32	Or like…
Interviewer:	15:33	You'd eat that bread?
Denise:	15:33	Don't they like produce like lactose or something? Er…
Mandy:	15:37	What produces lactose?
Denise:	15:39	Like when you ferment things, doesn't it produce like lactate or something?
Mandy:	15:43	Yeah, you can... I know one of the things is like – we just, we literally just talked about this in class too – it's like if you, when you ferment one of the products, it's gonna be like butyric acid or something or like, um…
Denise:	15:55	Well that be putting like…
Mandy:	15:57	Lactate. Yeah.
Denise:	15:58	Well like also I feel like, um, like the fermented stuff, it like, the bacteria inside, like eat the yeast, I think, as like, don't they eat the yeast? Like you add–
Mandy:	16:09	They eat the products of the yeast. Fermented, they eat the fermentation products. Are you talking about that?
Denise:	16:14	I don't know what I'm talking about anymore.

Speaker	Timestamp	Dialogue
Interviewer:	16:16	<laughter> That's okay.
Denise:	16:19	Um…
Interviewer:	16:21	This is a space to try ideas out.
Denise:	16:23	Also I feel like vodka isn't as like bubbly compared to like beer.
Mandy:	16:29	Yeah. I feel like beer would be the best. I don't really know what would happen with wine. I mean… It might still be effective, but I don't know if it'd be still as effective as beer.
Denise:	16:41	I think wine might produce a little gas still though.
Mandy:	16:46	What even is vodka made of?
Denise:	16:47	I dunno. I, I don't know. <laughter> Ethanol?
Interviewer:	16:52	Yeah. I, I think it's mostly like pure ethanol.
Denise:	16:55	Yeah. Yeah. So probably, yeah, I think probably just like wine besides beer.
Interviewer:	17:01	For obvious reasons, I wasn't allowed to bring those in <laughter> for the demo unfortunately, but uh, um… just have to think about it. Um… any… Oh, I almost forgot to ask. Um, so one of the things that they talk about with specifically the vinegar and the baking soda is having it with a little bit of milk. Any idea what the milk does? I genuinely don't know.
Denise:	17:29	Well, I know for like a microbio lab, if you have, if you drink like acid, you have to drink like milk, or milk of magnesia. So it probably like, um, neutralizes acidity.
Mandy:	17:43	Another, when you're like trying to make like substitute, um, buttermilk, you can mix vinegar and milk… I think. And then, then, yeah, that's like a substitute for buttermilk. I've done that before.
Interviewer:	17:58	Yeah?
Mandy:	17:59	I think there might be something else too, but I don't remember.
Denise:	18:01	Like buttermilk, is it like more like… chunky or…like thicker?
Mandy:	18:06	Pretty much just like thicker. Yeah. It's thicker milk.
Denise:	18:08	Yeah. Maybe, maybe makes it more like precipitate or something. <laughter> I dunno. That'd be kind of cool though if it did make it like precipitate or something.
Interviewer:	18:19	Yeah. Or I was wondering, um, someone actually in a, a different interview wondered if it kind of helped keep the bubbles longer?
Denise:	18:29	Ohh… Could be. I think that'd be interesting to know though.
Mandy:	18:38	Because when you mix like, like milk and vinegar on its own, I don't think anything much happens, but I don't even know like…
Denise:	18:45	Like traps…
Mandy:	18:45	…adding, adding, baking soda, what that would do.
Denise:	18:51	Would it like trap the like gas in there or…?
Mandy:	18:54	Or do you think it would make it like, like acting as like a thickening agent, I guess? The milk.
Denise:	18:59	Thickening agent? <laughter>
Mandy:	19:00	Yeah, like make it thicker and then the bubbles would be thicker, like these bubbles are kind of thick.
Denise:	19:06	Mm, yeah. Like they kind of like… stick more to each other.
Mandy:	19:10	Yeah, and they stick around, and these are kind of just dissipating right away.
Denise:	19:12	Mm-hmm. Yeah. It could, it could happen.
Interviewer:	19:18	Um… Any, any other ideas of what's going on here or anything else that you would want to try?

Speaker	Timestamp	Dialogue
Denise:	19:31	Hmm. Well, I feel like… if you were trying to use something… you'd probably want some kind of acid with some kind of base <laughter> so that it'd be like reactive. Um… But also I feel like if you, if it's just like a super strong acid and like a base and it'd probably be like this and like, react really fast… and not like stay as long. So maybe if you tried like something that's like a weak base and weak acid or something like that…
Interviewer:	20:06	Yeah.
Mandy:	20:08	Yeah.
Interviewer:	20:08	That, that sounds reasonable. And then if there's any leftover, it doesn't like scald your insides.
Denise:	20:16	Yeah.
Interviewer:	20:17	Awesome. Um, well that's honestly all the questions I have for you, so… last chance to, to say anything else, otherwise we can, we can wrap it up here.
Denise:	20:34	Any other thoughts, Mandy?
Mandy:	20:36	I don't think I have any other ones. Do you?
Denise:	20:39	Uh… nope. <laughter>
Interviewer:	20:42	Alright.

Table A.S6. Transcript of interview with Victoria and Jenn.

Speaker	Timestamp	Dialogue
Interviewer:	00:07	Alright. Um… So like I said, we're gonna talk about, um, some baking. Do either of you like to bake?
Jenn:	00:15	I do.
Interviewer:	00:16	You do?
Jenn:	00:17	Yeah.
Victoria:	00:18	I've baked like once in my life. So limited knowledge. <laugh>
Interviewer:	00:22	So you do the baking and you do the eating?
Victoria:	00:24	<laughter> Yeah.
Jenn:	00:25	Nice friendship, yeah. <laughter>
Interviewer:	00:27	What do you like to bake?
Jenn:	00:29	Um… Okay, so before, I'm celiac, so before I found out I was celiac, I did a lot of like muffins and like cookies and… Yeah, that's pretty much it. A lot of like, you know, whatever, whatever's in the pantry I would make so…
Interviewer:	00:46	Yeah. Um, it's probably a little harder to do a lot of baking now. Are you guys in apartments, or are you still in the dorm?
Jenn:	00:54	Yeah, I'm in a house.
Victoria:	00:55	I'm in an apartment.
Interviewer:	00:56	Yeah. Okay. You at least have kitchens.
Victoria:	00:59	Yes.
Interviewer:	01:00	You guys watch any of the baking shows on TV? I feel like a ton of people [inaudible]
Jenn:	01:05	Oh, the, the Great British–
Victoria:	01:06	I have. In the past.
Jenn:	01:07	What one? I watch the Great British Baking Show.
Interviewer:	01:10	That's my favorite.
Jenn:	01:12	Yeah, that one's good. I don't know what else. That's pretty much it. It's all I know.

Speaker	Timestamp	Dialogue
Interviewer:	01:17	Yeah. I mean… They don't get better than that, right, so like…?
Jenn:	01:21	Yeah. I don't have to watch any other ones.
Interviewer:	01:22	That's right. I feel like, um, I'm, I'm not really much of a baker or cooker or anything despite being a chemist. Um… Which confuses my family so much, but I feel like, you know, I watch Bake Off and then I'm like, “Oh, you didn't prove it long enough.” And I'm like, “I don't know what I'm saying.” <laughter> Because when I tried to bake something recently, I tried to make my sister’s cinnamon roll recipe, and I was like, “Cinnamon rolls, you know? I, I can do this.” Um, turns out my yeast had gone bad. They came out so flat. <laughter> Um, it was very disappointing. Um… And I, I, I tried eating a couple ‘cause I was like, they're covered in cinnamon sugar, right? But I was just, after a couple, I was like, “Oh, this texture's terrible.” Um, so I don't know if you have done any baking with yeast, um, but uh, turns out it's kind of important. Um, but I, I did read on the internet, um, that if you don't have yeast or if it's gone bad, you can use vinegar and baking soda as a substitute. Um, you're nodding, have you heard that before?
Victoria:	02:27	I've heard of multiple substitutes ‘cause I bought baking powder instead of soda or vice versa. And that was one of them… I think.
Interviewer:	02:36	Ah, nice. Um, so you've, you've heard of it. Have you tried it?
Victoria:	02:41	Um… once, maybe?
Interviewer:	02:42	I know you said you're not super into baking.
Victoria:	02:44	I’ve made muffins like three times. <laughter> One of them, yes.
Interviewer:	02:49	Um… fair enough. Um… Well I, I did bring in, um, some baking soda and some vinegar. Um… But I was wondering if you guys had any idea how this might work as a yeast substitute. And if it helps to… What happened to my spoon? There we go. Um, and if it helps to, to mix them together, um, see what happens…?
Jenn:	03:21	Sure, why not? <laughter>
Interviewer:	03:22	Right. It's, it’s not chemistry unless we're putting stuff together, right?
Victoria:	03:26	Do it.
Jenn:	03:26	I feel very like… this table is [crosstalk]
Jenn:	03:36	Okay. [inaudible] How much should I put in there? Hey, it says Ziplock. Sorry. <laughter> Anyway.
Victoria:	03:44	Should I pour it?
Jenn:	03:47	Watch it like explode. Okay. Actually well back up.
Victoria:	03:52	Can’t see.
Jenn:	03:54	Hmm. Okay, so it's bubbly. So that's probably…
Victoria:	03:59	Carbon dioxide? Oxygen?
Jenn:	04:00	Yeah, it's one of those. Some sort of gas <laughter> that makes it, you know, it sounds fun. <laughter> I’m like so like…ADD. Um, yeah, well, okay, so like with yeast, that's what, some type of microbe. Right? And what, protist? I should know this, I’m in microbio.
Victoria:	04:24	You’re in micro, I dunno.
Jenn:	04:26	I think it's protist, so that means that it excretes something. So then it probably has like, releases carbon dioxide, and so it makes it grow. So then if that's a gas, then it would kind of do the same. Right?
Victoria:	04:39	Self purifies, gets rid of… the gas.
Jenn:	04:42	Yeah. Reacts ‘til it can, can't, and then it makes it rise. That's, that’s my…

Speaker	Timestamp	Dialogue
Victoria:	4:49	That's our consensus. <laughter>
Jenn:	4:50	That's my… educated guess.
Interviewer:	04:53	Makes total sense to me.
Jenn:	04:54	Yeah? Okay, cool. <laughter>
Interviewer:	04:58	I don't know the biology side, so you could tell me yeast had like…
Jenn:	05:02	Well.
Interviewer:	05:03	…legs or something, I'd be like, "Okay." <laughter> I don't know.
Jenn:	05:07	Don't take my word for any of it.
Interviewer:	05:12	Um, alright. So you think, uh, yeast makes gas bubbles, this made bubbles, so that's, and that's how the rise works.
Jenn:	05:19	Yeah. Some sort of expansion… in that sense. Okay. <laughter>
Interviewer:	05:27	You can totally play with stuff. Um… I didn't bring any super dangerous chemicals so.
Jenn:	05:33	Darn.
Victoria:	05:32	So would this be like… acidic?
Jenn:	05:34	Mm. [inaudible] What is baking soda?
Victoria:	05:40	I don't know what baking soda is to be honest.
Jenn:	05:41	If this is, if this probably is what, like an acid-base reaction? Of like…
Victoria:	05:46	I…I would guess.
Jenn:	05:48	That would make sense. And then it…
Victoria:	05:50	'Cause this would be acidic… perhaps.
Jenn:	05:53	And then it releases CO_2. What was the reaction we just learned about that? Decarboxylation. That's the one. Right?
Victoria:	06:01	Oh! Yeah.
Jenn:	06:02	Is that a thing? Is that a thing?
Victoria:	06:03	That's a thing.
Interviewer:	06:04	That's a thing.
Jenn:	06:04	Okay, okay. That <laugh> I think that, right? Cause then you'd have–
Victoria:	06:09	That carboxylic acid, yeah.
Jenn:	06:11	Cause what's, what's the chemical formula of vinegar?
Victoria:	06:16	Genuinely could not tell you. <laughter>
Jenn:	06:19	Isn't that like acetate or something? Is that not right?
Victoria:	06:22	I have no, I don't know.
Jenn:	06:23	Hmm. Acetic acid? No. That's not right either. Is it?
Victoria:	06:27	I don't know. I could not make an educated guess, to be honest.
Jenn:	06:32	Um… Yeah. That's, so if that's, if that's a carboxylic acid on the end, you can have… like a decarboxylation or something. Wait, don't you, shoot, do you need water and H_3O^+ though. I think. This is why…
Victoria:	06:52	<laughter> This is why the exam did not go well.
Jenn:	06:54	This is why it did not go well.
Interviewer:	06:57	Okay. I would guess–
Jenn:	06:57	'Cause there's no water in that, it's just, right? <laughter> Ingredients.
Interviewer:	07:03	Check it out.
Jenn:	07:05	Cincinnati. Okay. Diluted water.
Victoria:	07:10	Diluted water?
Jenn:	07:11	With water, sorry. Diluted water. <laughter>
Victoria:	07:13	Diluted with water. I was like, "What is…" Diluted with what?
Jenn:	07:16	To 5% acid strength. It's diluted–
Victoria:	07:18	Oh, it is acid!
Jenn:	07:18	It's diluted with water to 5. So yeah.

Speaker	Timestamp	Dialogue
Victoria:	07:21	It is an acid-base. Maybe.
Jenn:	07:22	Well that makes sense.
Victoria:	07:24	Look at that.
Jenn:	07:24	Yeah. Right? Yeah, yeah.
Victoria:	07:26	Hmm.
Jenn:	07:26	Hmm. <laughter> Hm. Look at that. <laughter>
Interviewer:	07:30	And so if the, the vinegar is the acid, you're saying that's the… base?
Jenn:	07:33	Yes, so this is, this is the base, then. And then… from my limited… carbon dioxide knowledge, decarboxylation of… If this is a…
Victoria:	07:46	If it's a carboxylic acid.
Jenn:	07:46	If that's a carboxylic acid. 'Cause it makes bubbles. Yeah?
Victoria:	07:51	Bubbles out of solution.
Jenn:	07:52	Yeah. And that would, yeah.
Interviewer:	07:56	Alright. So… So you're saying if I were to make, say a loaf of bread, if I, if I learned how to make a loaf of bread, which I would have to do first, um, and then made it with baking soda and vinegar instead of yeast, would you, would you eat it? Assuming I like followed all the, y'know, hygiene protocols.
Jenn:	08:18	Well, I mean, I wouldn't. <laughter> Because I can't.
Interviewer:	08:21	Oh, right, right, sorry!
Jenn:	08:23	But, but if I could, yes.
Interviewer:	08:25	Like it would be edible?
Jenn:	08:26	Yes.
Victoria:	08:27	Yes.
Jenn:	08:28	If I wasn't diseased. Yeah. Sorry. <laughter>
Interviewer:	08:32	<sigh> I'm a terrible person for laughing at that, sorry.
Jenn:	08:34	No, it's funny. I think it, I make the joke all the time, it's so funny. 'Cause, why not? It's funny.
Interviewer:	08:40	'Cause people don't know how to react to it.
Jenn:	08:41	Yeah, yeah, exactly.
Interviewer:	08:43	You, you might as well have fun with it.
Jenn:	08:44	It's just, it's celiac. It's not anything major. <laughter>
Interviewer:	08:48	So, uh, vinegar, if that's an acid would… Say, if we swap that with like lemon juice or something, do you think that would work just fine too?
Jenn:	08:58	If it has a carboxylic acid.
Victoria:	08:58	Yes. Lemon juice is, lemon juice is acidic.
Jenn:	09:03	Yeah, but if it has, doesn't it need to have, if we say that, if our claim is that it's a carboxylic acid and then it releases CO_2 by decarboxylation, then it would need to have, lemon juice would need to have a carboxylic acid as well. Or yeah. So if that is true, then yes. I don't know if it would taste great if it's lemon-y. But I guess lemon bread, that's a thing.
Victoria:	09:24	That's a thing. Right?
Jenn:	09:24	Okay. Right?
Victoria:	09:26	Because it's definitely like an acid-base reaction. I don't know if it's decarboxylation, so I don't know if it needs to be a carboxylic acid.
Jenn:	09:33	Definitely acid-base.
Victoria:	09:33	But lemon juice is definitely acidic. So… it would work.
Jenn:	09:39	Yeah.

Speaker	Timestamp	Dialogue
Interviewer:	09:41	Alright. Um… What about something like, I'm just like thinking of some of the liquids in the kitchen, um… And like something like vodka, I don't know, we're on a college campus, there's probably alcohol somewhere around here. Do you, do you think that would work instead? It's a clear liquid.
Jenn:	10:00	It's a clear liquid, but is it, what even, if, if it's an alcohol…
Victoria:	10:04	It's an alcohol, it has the… OH group. So… I feel like… no?
Jenn:	10:13	Well it's like beer bread.
Victoria:	10:16	Beer bread?
Jenn:	10:17	You know how they put beer in? Have you not had beer bread?
Interviewer:	10:20	Oh, you're missing out. [crosstalk] Sorry, I have to on interject on that one.
Victoria:	10:24	I've never even heard of that, no.
Jenn:	10:26	Tastefully Simple beer bread? With like little seasoning packet-?
Victoria:	10:32	Mm-mm (negative).
Jenn:	10:32	Um, we have it at every Christmas and every family gathering. I'm very sad I can't partake in it. I'm just like making this a total personality trait. Um.
Victoria:	10:43	Do they make it with…?
Jenn:	10:45	You put a can of beer, it's like a mix, it's like a mix and you put a can of beer in it, and then you, y'know, cook it, but… I mean I'm sure that, that has like, doesn't beer have yeast in it? Is that a thing?
Victoria:	10:56	Oh, wait, yeah. It does. So then yeah, probably would.
Jenn:	11:01	Yeah. That is not right though, that's totally, what was the question?
Interviewer:	11:05	Oh, that's fine. We could, we could try beer too.
Jenn:	11:08	Vodka.
Victoria:	11:08	Could we use vodka?
Jenn:	11:10	Um, I don't know. 'Cause it's like it's,
Victoria:	11:15	It's got an OH. <laughter>
Jenn:	11:16	Well, yeah. Yeah.
Victoria:	11:19	Um…
Jenn.	11:25	So that and that… in there.
Victoria:	11:29	No. And the baking soda.
Jenn:	11:31	And the baking soda?
Victoria:	11:32	Yeah. We're replacing this.
Jenn:	11:33	Then probably not. 'Cause that's base in base.
Victoria:	11:37	Yeah. I would agree with that. Yeah.
Jenn:	11:43	Right?
Victoria:	11:44	I think so.
Jenn:	11:45	Yeah, 'cause how else? Why, I feel like it wouldn't react. If this is–
Victoria:	11:49	And if we're saying that's… the base. 'Cause we know this–
Jenn:	11:52	Well if it's a base, we don't even know if it's–
Victoria:	11:54	I think–
Jenn:	11:55	I feel like it has to be.
Victoria:	11:56	I feel like it would be. That would make sense.
Jenn:	11:58	Yeah. Okay.
Victoria:	11:59	So then no, it would not work.
Jenn:	12:00	It would not.
Interviewer:	12:01	Alright. I, I'm sorry, I wasn't allowed to bring in vodka to try that out. <laughter>

Speaker	Timestamp	Dialogue
Jenn:	12:05	No, it's okay.
Interviewer:	12:06	I think there are rules against that at least. Um…
Jenn:	12:08	Science. It's fine.
Interviewer:	12:10	Yeah. It's for science. <laughter>
Jenn:	12:11	Yeah.
Victoria:	12:12	It's for science.
Interviewer:	12:14	Um… You mentioned a little bit, uh, bleh. You mentioned baking powder. That's baking soda. Um… Do you think if we swap those two out? Um… So the vinegar with the baking powder?
Victoria:	12:32	I don't know the difference between baking soda and baking powder–
Jenn:	12:34	Me neither.
Interviewer:	12:36	I can never remember.
Victoria:	12:36	
Jenn:	12:42	Well then it was probably, if, if… If this reacts, then that's probably–
Victoria:	12:43	I think… What is this? Soda?
Jenn:	12:45	This is soda.
Victoria:	12:46	I think soda's… Which is the one that comes in the orange box?
Jenn:	12:51	Soda. It's soda.
Victoria:	12:52	It's soda? I think… that's what you… Shoot, I don't know. One's used for baking more and one's used for cleaning. <laughter>
Jenn:	13:00	Baking soda's used for cleaning.
Victoria:	13:02	Okay. So then maybe…
Jenn:	13:04	Baking powder is, is used for baking?
Victoria:	13:08	I know they can't be used interchangeably. 'Cause one you have to mix with the other things
Jenn:	13:12	And sometimes you use both.
Victoria:	13:14	And sometimes you use both.
Jenn:	13:15	Like the Nestle, Nestle? Nestle? Nestle cookies? The back of the bag? The chocolate chips? It has both.
Victoria:	13:23	'Cause I think baking soda, you have to mix with the acidic thing.
Jenn:	13:28	Yeah.
Victoria:	13:30	Baking powder you don't? …'Cause like as a substitute. 'Cause I think I bought baking pow–, er, soda, and I needed baking powder, so–
Jenn:	13:40	And so you had to–
Victoria:	13:41	I was supposed to mix it with something and then I gave up cause I didn't have the thing to mix it with. It was like buttermilk or something.
Jenn:	13:46	You know you could just make it… with the lemon juice, right?
Victoria:	13:50	I didn't have, I don't have lemon juice.
Jenn:	13:54	I don't think we do either. Anyway.
Victoria:	13:57	So I think, no.
Jenn:	14:00	Yeah. I second that.
Victoria:	14:01	Or yeah, I don't think so.
Interviewer:	14:05	Do you guys have any other ideas of good substitutes? Um, anything else you've maybe seen across the internet?
Victoria:	14:13	Buttermilk.
Interviewer:	14:13	Things you'd wanna try?
Victoria:	14:15	Buttermilk. I've heard.
Interviewer:	14:15	Buttermilk? Just, just buttermilk?
Victoria:	14:18	With the…
Interviewer:	14:19	With the?

Speaker	Timestamp	Dialogue
Jenn:	14:20	With, with this?
Victoria:	14:21	This, yes.
Jenn:	14:22	Okay. <laughter>
Victoria:	14:25	Or, I've heard milk that goes bad is the same thing. Which–
Jenn:	14:28	As buttermilk? Do not do that.
Victoria:	14:30	I, no, I wouldn't. I would not.
Jenn:	14:33	It's just like–
Interviewer:	14:34	But it's like chunky!
Jenn:	14:35	It's, yeah. That's–
Victoria:	14:37	That's what the internet told me. And I did not try it.
Jenn:	14:39	This is why we don't believe everything on the internet.
Victoria:	14:40	Exact–, I'm just throwing out ideas that hypothetically…
Interviewer:	14:44	No, it's great.
Victoria:	14:45	…could work chemically. Perhaps. Not to consume, maybe.
Jenn:	14:47	Yeah, I, I gotcha. What… What in the spoiled milk would make it a [inaudible] reactive with baking soda?
Victoria:	14:55	I don't know. I don't know. Maybe… I don't know. Maybe it oxidates or something.
Jenn:	15:06	Seems legit.
Victoria:	15:09	That's my best guess.
Jenn:	15:09	Um, I don't know what else. Um… Mmm… I don't know if water, water probably wouldn't, wouldn't do it.
Victoria:	15:21	I don't think so, I don't think water would be acidic enough.
Jenn:	15:23	Yeah, oh yeah. Battery acid. I'm just kidding.
Victoria:	15:27	Battery acid?! <laughter>
Interviewer:	15:30	You're the first to suggest that. <laughter>
Jenn:	15:32	Oh, well, y'know, first time for everything.
Victoria:	15:35	Geez!
Jenn:	15:36	No, okay, that was a joke. Um… I wonder if like vitamin C… 'Cause that's acid. It's–
Victoria:	15:45	Is it?
Jenn:	15:46	It's ascorbic acid. Yeah, because my roommate, my roommate was like, "Oh, I wanna have, like chewable, like, vitamin C." And she didn't get that. And she like, was like, "I'm gonna suck on it anyway." And she got like a chemical burn on her tongue 'cause it was acid. She's like, "My tongue really hurts." She's like, "This tastes like a warhead. Try it, it tastes good." And I spat it out. But then, she like held onto it and got a burn on her tongue. Yeah. So maybe vitamin C would work. <laughter> If you like, if it, if it's like liquid. 'Cause you obviously need a liquid 'cause then it would… Yeah? Like, um, the Emergen-C packets? You know?
Victoria:	16:24	Oh yeah…. Huh…. Yeah.
Interviewer:	16:30	Those are liquids?
Jenn:	16:31	You can like, you put 'em in water.
Victoria:	16:33	You add 'em to water to dissolve it.
Jenn:	16:35	Or like orange juice?
Victoria:	16:39	I don't know if it's acidic enough, but… it might be.
Jenn:	16:42	Oh, when you, when I, when I drink orange juice, my tongue gets raw. Maybe I should–. Or pineapple? Like pineapples?
Victoria:	16:48	Pineapple juice is more acidic.

Speaker	Timestamp	Dialogue
Jenn:	16:50	Yeah.
Interviewer:	16:51	Pineapple flavored bread.
Jenn:	16:53	That could be…
Victoria:	16:54	That sounds promising.
Jenn:	16:56	Could be interesting.
Interviewer:	16:57	I'd, I'd try it.
Jenn:	16:58	Yeah. Or like ch–, I don't think cherries are very acid–, well… I don't know. Sometimes, er, like blue–, er, like a sour blueberry? That's very acidic, it feels like. Maybe not. That's what it tastes like.
Victoria:	17:11	I'm just comparing it to lemon. Like if it's more acidic than a lemon. I don't know why that's my–
Jenn:	17:18	So a lime? Is lime more acidic? Yeah, it definitely is.
Victoria:	17:22	Is it?
Jenn:	17:22	Mm, maybe not.
Victoria:	17:23	I would say no.
Jenn:	17:24	Okay. Fine. <laughter>
Victoria:	17:28	I would say strong to medium acids.
Jenn:	17:32	Okay.
Victoria:	17:33	I don't know if a weak acid would work.
Jenn:	17:34	Probably not.
Victoria:	17:37	Or like how, I don't know where the line would be… between… like how weak it can be? But water definitely wouldn't work.
Jenn:	17:45	No. So water's pKa of…
Victoria:	17:47	And that's, what is that, pKa of 16?
Jenn:	17:49	Yeah, 15.7? 9?
Victoria:	17:49	15.9? 7? 9?
Jenn:	17:52	Is it nine or seven? <laughter>
Interviewer:	17:54	It's seven.
Victoria:	17:55	It's seven. Aw.
Interviewer:	17:55	It's sad that I know that.
Jenn:	17:57	Hey, I knew it too.
Victoria:	17:58	I knew it rounded to 16.
Jenn:	18:01	Umm…
Victoria:	18:03	More acidic than water though. I don't know how much more acidic than water though.
Jenn:	18:07	Like…
Victoria:	18:08	I feel like a weak acid wouldn't work, but I don't know why that's my gut. I have nothing to back that up.
Jenn:	18:15	Um… so… If it's, if that, what would have, what would that have to mean about the conjugate base? If it's a weak… If it's a strong acid, weak base? Right?
Victoria:	18:24	Mm-hmm. I don't know anything about baking sodas though.
Jenn:	18:31	Oh, I dunno, me either, me either, man. Um…
Victoria:	18:34	What is baking soda? I have no idea.
Jenn:	18:35	See like I'm gonna go home, and I'm gonna be like–
Victoria:	18:37	What is baking soda? I still don't remember.
Interviewer:	18:40	If you guys really wanna know, I can tell you what it is.
Jenn:	18:43	Yeah.
Victoria:	18:43	Sure. …Is it NaOH? Okay.

Speaker	Timestamp	Dialogue
Interviewer:	18:51	Um, and the vinegar is mostly… like you guys thought… a carboxylic acid.
Victoria:	18:55	Ooohhhhh…
Jenn:	18:59	Nice!
Victoria:	19:00	That is nice. That makes me feel a lot better.
Jenn:	19:03	Yeah. Okay. Oh, that makes sense. Yeah. That's just, that's a base. That's a big, that's a big base. Good base. Good work. Okay. <laughter> Good work. Okay. Um…
Victoria:	19:15	Look at us.
Interviewer:	19:16	Ego boost.
Jenn:	19:17	Yeah. Oh, oh yeah.
Victoria:	19:18	That much needed.
Jenn:	19:19	I needed anything after that exam. Yeah.
Victoria:	19:21	Much needed.
Jenn:	19:24	Yeah. Okay. That makes sense. In my head anyway. [inaudible] Cool.
Interviewer:	19:33	Alright. Um… Any, anything else you wanna add about this? Otherwise, I think…
Jenn:	19:42	No. I'm pretty, pretty confident in my… guess.
Victoria:	19:46	Yeah. I'm very happy to see that…
Jenn:	19:48	Yeah.
Victoria:	19:49	…carboxylic acid. Seeing that CO_2…
Jenn:	19:52	Mm-hmm. Yeah.
Victoria:	19:55	...bubble out of solution. Self purifying. Makes me feel… very good. <laughter>
Interviewer:	20:01	Purify? Are you also in [OChem lab] right now?
Victoria:	20:03	No, I'm not. Someone mentioned it in lecture. And then [my professor's]–
Jenn:	20:08	What? Pur–, purify? Yeah, I was like, "What are you talking about?"
Victoria:	20:10	Yeah, so like when you have like, when you have the gas, it'll bubble out of… your… liquids…
Jenn:	20:16	Oh, and then it makes it just–
Victoria:	20:18	…liquids, so then it drives it towards the products, and then so–
Jenn:	20:20	Le Chatelier's principle?
Victoria:	20:21	Mm-hmm. And so it'll react like all the way.
Jenn:	20:25	I definitely put Le Chatelier's principle on my exam when I didn't know what to put. <laughter> I was like, "Le Chatelier's principle!"
Victoria:	20:32	But so then it'll keep driving it, and it bubbles out, so you don't have to like purify it yourself.
Jenn:	20:36	That makes sense.
Victoria:	20:37	Does it naturally.
Jenn:	20:38	Yeah, yeah.
Victoria:	20:39	So bubbles are good… is what I learned. In the past week. That was very recent.
Jenn:	20:45	Wow. Good job.
Victoria:	20:47	Thanks.
Jenn:	20:48	You're welcome.
Interviewer:	20:49	Alright.

JOURNAL OF CHEMICAL EDUCATION

pubs.acs.org/jchemeduc

Article

Impact of Assessment Emphasis on Organic Chemistry Students' Explanations for an Alkene Addition Reaction

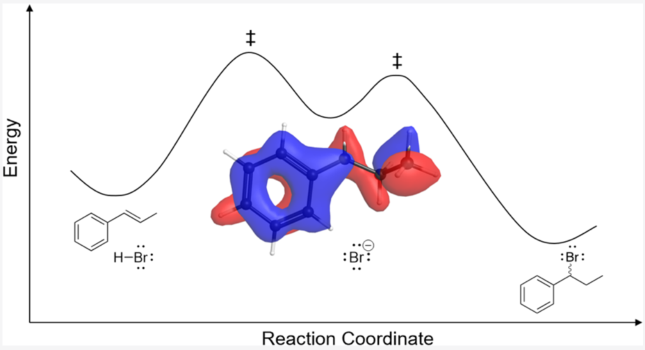
Kimberly S. DeGlopper, Cara E. Schwarz, Niall J. Ellias, and Ryan L. Stowe*

Cite This: *J. Chem. Educ.* 2022, 99, 1368−1382

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: To potentially engage students in "doing organic chemistry", organic chemistry courses should foreground weaving together structure- and energy-related ideas to construct causal accounts for phenomena. Here, we investigate whether enrolling in an organic chemistry course that places substantial emphasis (~50% of total points) on explaining phenomena on exams is associated with more productively justifying the outcome of a chemical process. This work occurred in the context of three learning environments that differed principally by assessment emphasis. The "explanation focused" course allotted 40−66% of points on exams to explaining phenomena while the other two enactments placed much less emphasis on connecting big ideas to how and why chemical processes occur (~0−25% of total points). Students enrolled in each course were given a prompt which asked them to draw mechanisms for a hydrobromination reaction and subsequently justify the regiochemical outcome of that reaction. We described student responses by noting the connections made between structure (of reactants, intermediates, transition states, or products) and energy. Most students described how charge or electron delocalization impacted the relative energies of the two possible intermediates or transition states. Other explanations invoked steric repulsion or differences in relative energy due to the degree of carbocation substitution. An examination of the association between learning environment enrollment and explanation code distribution revealed that students who enrolled in two semesters of an explanation-focused course were substantially *less* likely to leave out charge delocalization in their explanations while students who were never enrolled in the explanation-focused learning environment were substantially *more* likely to leave out charge delocalization. These findings suggest that changing what is assessed to better align with "doing organic chemistry" may be a promising avenue for reform.



KEYWORDS: *Second-Year Undergraduate, Chemical Education Research, Testing/Assessment, Addition Reactions, Mechanisms of Reactions, Reactive Intermediates*

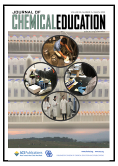
FEATURE: Chemical Education Research

■ INTRODUCTION

Organic chemists seek to understand and manipulate reactions with the goal of efficiently and selectively synthesizing molecules that possess properties suitable for particular functions (e.g., inhibition of disease-relevant macromolecules, impact resistance, conductivity). Toward that end, many organic chemists are engaged in designing new reaction systems, which involves predicting and rationalizing the possible outcomes of reactions based on knowledge of how chemical structure influences reactivity. Take, for example, the development of enantioselective organocatalysis, which recently won the Nobel Prize in Chemistry.[1] Enantioselectivity is achieved because the chiral catalyst alters the transition state structures leading to the two possible enantiomeric products and, in turn, the energy difference between the transition states. In some cases, the catalyst lowers the energy of one transition state relative to the other via noncovalent interactions while, in others, the catalyst raises the relative energy of one through increased steric repulsion. Constructing explanatory accounts of how altering system parameters might change reaction outcomes allows chemists to purposefully plan experiments directed at reaction optimization.[2] Stated succinctly, figuring out how a reaction is likely to proceed via connecting chemical structure (e.g., of intermediates, transition states, or products) and relative energy is a fundamental part of "doing chemistry".

There are compelling reasons to frame chemistry learning environments as opportunities to "do chemistry". Engaging

Received: October 14, 2021
Revised: January 12, 2022
Published: February 9, 2022



ACS Publications

https://doi.org/10.1021/acs.jchemed.1c01080

learners in creating, refining, and communicating knowledge about how and why the world works has the potential to focus the classroom community on "figuring out" aspects of their existence rather than "learning about" disaggregated skills and facts.[3] Emphasis on "figuring out" makes explicit that science activities and content knowledge are *always* related in the practice of science.[4] Specialized skills, such as drawing electron-pushing arrows, have no inherent meaning; one can learn the rules of "mechanism drawing" without ever considering donor−acceptor interactions that might occur in a system. Likewise, one may memorize the definition for "π-conjugation" or "asymmetric induction" without ever being able to use these ideas to articulate why something happened (and why other things did not). Skills and knowledge gain meaning when purposefully woven together with the aim of understanding how the world works or designing a solution to a pressing problem.

Unfortunately, it is fair to say that "doing science" is not a central focus of many organic chemistry courses. Indeed, it is common for the "correct" application of a skill or recall of a reaction product to be allotted far more points on assessments than the construction of causal accounts for phenomena.[5] This is troublesome since one may readily apply the "rules of the game" to enact a skill without understanding the chemical system related to that skill. Perhaps the best example of an explanatory tool that is often reduced to a skill is the electron-pushing formalism (EPF). Organic chemists make use of the EPF to reason about how the movement of electrons transforms a starting material into a product. However, several studies have demonstrated that students commonly "decorate" starting materials and intermediates with arrows rather than using curved arrow mechanisms as predictive tools.[6,7] When investigating how students reason about reaction mechanisms, Caspari et al. found that many students relied on teleological reasoning and used the favorability of a subsequent mechanistic step to justify proposing an earlier one, despite the molecule having no way of knowing that a subsequent step would be productive.[8] Even first-year graduate students often proposed mechanisms based on what would get them from the starting material representation to the product representation rather than based on arguments grounded in stabilization of charge or electrophilicity.[9] Potential energy surfaces constitute another potentially powerful model that may not hold meaning for students, as observed by Popova, Bretz, and Lamichhane et al.[10−12] Importantly, we argue that a tendency toward recall and decontextualized skill application is the result of inappropriate learning environment design, not a deficiency on the part of enrolled students. Students in the aforementioned studies were often quite adept at the performances emphasized and rewarded in their courses. Therefore, our focus in this study is not on documenting students' misconceptions but on examining how learning environments may be designed to support students in doing chemistry.

When we say a performance was "emphasized and rewarded" in this article, we mean that a task was assessed. Assessments serve two important roles in learning environments. First, they convey strong messages to students as to what is important in the course (and in the discipline in general).[13−18] Students who are frequently asked to explain why a phenomenon occurs on assessments will learn to do so, as Crandell et al. observed when comparing the explanations of students enrolled in a transformed course to those of students enrolled in a traditional organic chemistry course.[19] After a year enrolled in the transformed course Organic Chemistry, Life, the Universe and Everything (or OCLUE), which prioritizes reasoning with core ideas, students were more likely to provide causal explanations for an S_N2 reaction than students in the traditional organic chemistry courses.[20] The second role of assessments is to elicit evidence of what students know and can do. Ideally, evidence elicited by assessments should be used to inform instructors and to help students chart their learning priorities. The sorts of inferences that can be made from what students write or say are powerfully influenced by the structure of the prompt as well as the model of mind adopted by the instructor or researcher (more on that shortly). Responses to an assessment item that simply asks for a claim (such as "circle the most acidic molecule") provide no evidence of how students arrived at that claim. We argue that learning environments which foreground "doing chemistry" should consistently engage learners in constructing and critiquing causal accounts for phenomena.[21] This entails moving beyond asking solely for claims and toward expecting reasonable connections of big ideas (e.g., energy, donor−acceptor interactions) to phenomena.

Despite the influential role assessments play in signaling to students "what counts", we are aware of no scholarship exploring the impact of changing assessment emphasis (and little else) on student explanations for phenomena. Most assessment reform efforts in chemistry education have been coupled with curricular transformations. For example, the general chemistry course Chemistry, Life, the Universe and Everything (CLUE) integrates assessments focused on mechanisms underpinning phenomena and also reorganizes the curriculum around scaffolded sequences of big ideas.[22] Likewise, Chemical Thinking represents an overhaul of both curricular sequencing and assessment emphasis.[23] With regard to organic chemistry education, both OCLUE and Flynn's mechanisms before reactions courses have simultaneously reformed their curricula and assessments.[20,24] The context for this study is thus a bit unique: organic chemistry courses at a research-intensive Midwestern university follow a unified curricular sequence informed by a commercial textbook but (as we will see) emphasize different sorts of tasks on assessments. This presented us the opportunity to probe the relationship between assessment emphasis and students' justifications for agreeing or disagreeing with a given reaction outcome claim.

The Model of Mind Informing Our Study

Before describing the study in detail, it is important to clearly articulate the model of learning we are using to infer aspects of student cognition. The model of learning one adopts influences the conclusions and implications of a study by bounding the study designs that seem reasonable and the sorts of inferences one can draw from student response data.[25] To aid readers in understanding the constraints and affordances of our study design, we provide here a brief description of the model of learning we adopt and discuss in broad terms how it influenced our view on the roles of instructors and students.

Our study design and implementation were informed by the resources model of cognition put forth by Hammer and Elby, which draws on the knowledge-in-pieces framework from diSessa.[26,27] This model of mind assumes that knowledge exists in pieces that are activated in the moment for a specific purpose. These small-grain "resources", such as the primitive "more means more" (Ohm's p-prim), are not inherently "correct" or "incorrect" but may be more or less productive depending on context.[27] Connections between knowledge elements elicited by a given assessment or instructional scenario are not assumed to

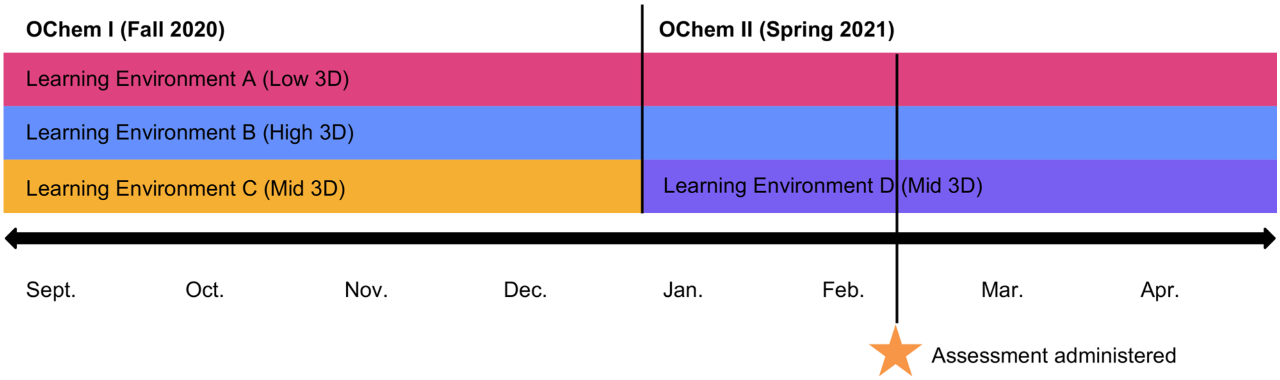


Figure 1. Timeline showing when each learning environment was operating during the Fall 2020−Spring 2021 academic year and when the assessment item used in this study was administered.

be stable across time and place (although some may be). In contrast, a theory−theory model of learning assumes stable, coherent knowledge structures which, if incorrect, are referred to in the literature as "misconceptions", "alternative conceptions", or "naïve theories".[28] Students possessing theorylike knowledge would be expected to offer consistent responses to questions that require use of the same naïve theory to answer. In fact, studies have shown that students' responses are often contradictory, depending on the contexts of the questions and the ways in which they are worded.[29,30] This inconsistency can be better accounted for by a resources model where activation of knowledge elements is dependent upon the specific context.

These two contrasting models of cognition result in very different roles for the instructor. An instructor with a resources model of cognition would interpret student responses as momentary coalescences of knowledge elements and seek to recognize, reward, and build upon productive resource use. Adopting a resources view of learning *requires* an instructor to attend to the substance of student thinking, even if the "incorrect" vocabulary is used.[31] By contrast, an instructor with a theory−theory model of cognition would interpret "incorrect" student responses as indicative of misconceptions and seek to replace these with the "correct" conceptions. Some scholars propose a process of rationally challenging and replacing student "misconceptions" in which students are shown the inadequacies of their naïve theory and offered a canonical theory as a more useful replacement.[32] Needless to say, there is no convincing evidence that a robust and useful understanding of fundamental chemistry ideas comes about through a series of rational "paradigm shifts". These models of cognition also have important implications for how we view students. From a resources perspective, students possess potentially productive knowledge elements that can be used to further their own learning. Theory−theory views align more with a deficit perspective on students, in which their prior knowledge is often problematic and needs to be rooted out rather than utilized.

We should note that "knowledge-in-pieces" and "theory−theory" models of mind represent two ends of a continuum rather than mutually exclusive theories of cognition.[33] Resources that are consistently activated together may approach theorylike stability. However, because we are not characterizing students' knowledge over time or across multiple prompts, we do not intend to make any claims concerning consistency or stability. A resources model of cognition is thus appropriate for our aim of identifying the productive connections students make in the moment when explaining a reaction outcome.

Study Goals

As part of a broader, ongoing effort to improve introductory organic chemistry courses at a large, research-intensive, Midwestern university, we aim to characterize the knowledge elements that students call to mind and connect when asked to reason about phenomena of interest to organic chemists. In this study, we asked students enrolled in three different learning environments (described below) to consider a benzylic alkene addition reaction, construct electron-pushing mechanisms depicting the formation of two possible products, and evaluate a claim regarding which would be the major product. Specifically, we sought to answer the following questions:

(1) What knowledge elements do students activate and connect when asked to explain the outcome of a benzylic alkene addition reaction?
(2) How do the electron-pushing mechanisms students draw relate to the explanations they provide?
(3) How does the intellectual work emphasized and rewarded on assessments given in each learning environment relate to the structure−energy connections invoked in student explanations?

Our focal phenomenon (addition of HBr to a benzylic alkene) was chosen due to the importance of π-conjugation in stabilizing the high-energy intermediate leading to the major product. The constructive overlap of a series of p-orbitals explains a wide variety of important chemical and biochemical phenomena (e.g., protein 3D structure, stability of druglike aromatic compounds) and should therefore be prominent among the powerful explanatory ideas students use for sense-making.

■ METHODS

Study Context

This study was conducted at a large, public, research-intensive university in the Midwest. The courses involved in this study were the on-sequence first and second semester of introductory organic chemistry (OChem I and OChem II) enacted during the Fall 2020−Spring 2021 academic year. These courses serve chemistry majors, biology majors, and chemical engineering majors; most students enrolled intend to pursue careers in the health fields. Both courses were administered entirely online due to the COVID-19 pandemic. As the data collection and analysis described here were undertaken primarily for the purposes of program improvement, IRB approval was not required. The

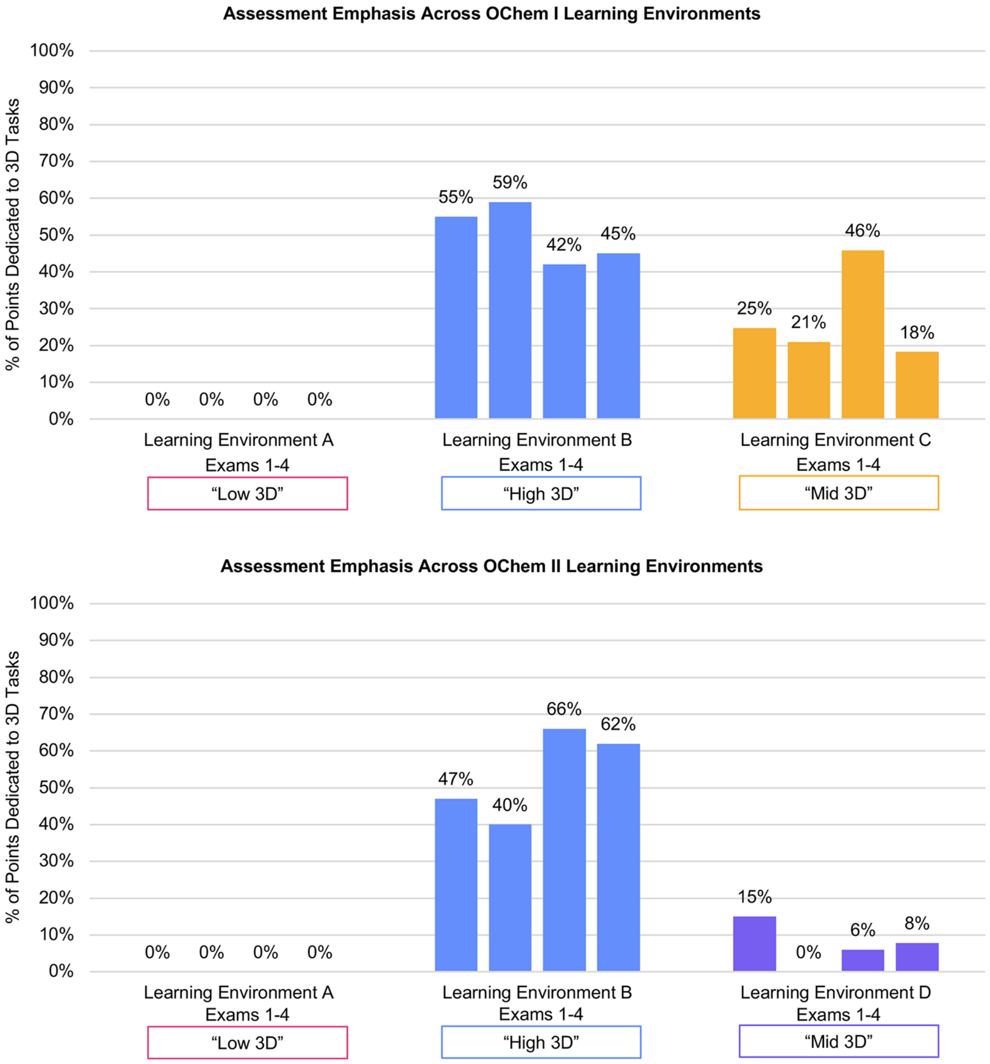


Figure 2. Percentage of points dedicated to 3D assessment items on exams given in each of the three learning environments during OChem I (top) and OChem II (bottom). Based on the data shown here, we designated Learning Environment A as "low 3D", Learning Environment B as "high 3D", and Learning Environments C and D as "mid 3D".

assessment prompts used as the outcome measure for this work were developed as part of another, IRB-approved study (ID 2020-0684). Accordingly, participant consent was obtained for the response-process interviews conducted during instrument development.

At this institution, organic instructors agree to use the same textbook and cover roughly the same set of content in their organic chemistry courses. However, each instructor has the freedom to operate independently and has full control over how they teach and assess students. Some instructors voluntarily choose to team up and write and administer common assessments. During the course of this study, four unique learning environments were enacted. We defined a learning environment to include all course sections that were taught by the same instructional team and had the same structure, assessment formats, and assessment emphases (see below for details on assessment emphases). Two learning environments (A and B) extended the full two-semester sequence. That is, Learning Environments A and B are characterized by one instructional team enacting both OChem I and OChem II with a consistent structure and assessment emphasis. Learning Environment C was an enactment of OChem I while Learning Environment D was an enactment of OChem II (Figure 1). Note that a student might not be enrolled in the same learning environment for both courses. For example, a student enrolled in Learning Environment A for OChem I could be enrolled in any of the three learning environments for OChem II. All learning environments proceeded through roughly the same sequence of topics for each course, and they all featured a mix of prerecorded lecture videos, synchronous problem-solving sessions with the instructor, and optional TA-led synchronous discussion sections. The sixth edition of Marc Loudon's *Organic Chemistry* was used as the textbook for all learning environments under study.[34]

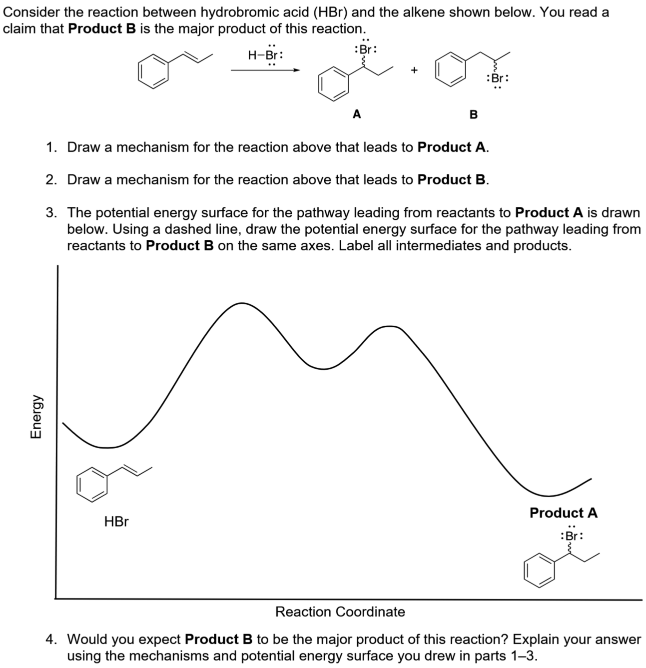
Since assessments convey strong messages to students regarding what sorts of performances are valued, the learning environments were characterized according to their assessment emphasis. The 3-Dimensional Learning Assessment Protocol (3D-LAP) was used to identify assessment items with the potential to elicit evidence of engagement with scientific practices, disciplinary core ideas, and crosscutting concepts (i.e., 3D learning).[35] Details of our 3D-LAP coding process may be found in the SI. We agree with modern science reform efforts that K−16 STEM learning environments should support the integration of activities scientists do (i.e., science practices) and content knowledge (i.e., core ideas).[4] Use of the 3D-LAP here is a concrete reflection of this commitment; LAP readouts help us capture the extent to which instructors emphasize and reward knowledge-in-use on the exams they give. Items which fulfill LAP criteria for potentially eliciting evidence of 3D performances require students to articulate reasoning underpinning claims made while non-3D items typically only require an answer in the form of a claim (e.g., product of a reaction, most acidic molecule) without the accompanying reasoning. It is important to emphasize that items denoted as 3D in this study have the *potential* to elicit evidence of knowledge-in-use. Meeting 3D-LAP criteria does not guarantee a given item *will* elicit evidence of engagement in a 3D performance. We refer the reader to the SI for examples of 3D and non-3D prompts as well as the complete set of exams administered during OChem I and OChem II in Learning Environment B. For more examples, please see our recent publication, which contains dozens of 3D prompts administered as homework and on exams.[36]

In Figure 2, the percentage of points dedicated to assessment items with the potential to elicit evidence of 3D learning is shown for each exam given in each of the learning environments. Exams 1−3 were given throughout the semester while Exam 4 was given as the final exam and was worth more points. For OChem I, Learning Environment B gave exams with 42−59% of the points dedicated to 3D items while Learning Environment C gave exams with 18−46% of the points dedicated to 3D items. None of the exams given in Learning Environment A contained a 3D assessment item. A similar trend was observed for OChem II. Learning Environment B gave exams with 40−66% of the points dedicated to 3D items and Learning Environment D exams with 0−15% of the points dedicated to 3D items, and none of the exams given in Learning Environment A contained a 3D assessment item. The substantial differences among learning environments prompted us to further investigate the link between assessment emphases and student-constructed explanations for phenomena. To aid readers, we will refer to Learning Environments A and B as low and high 3D learning environments, respectively, and Learning Environments C and D as mid 3D learning environments throughout the remainder of the paper.

Instrument

The assessment item used in this study was created using evidence-centered design (ECD).[37] Accordingly, a precise expectation for what organic-enrolled students should know and be able to do guided task design. The performance expectation that our task was built to assess was: "Construct and use an electron pushing mechanism and/or a reaction energy profile to evaluate the validity of claims as to the outcome of a chemical process." This performance expectation does not solely reflect what the authors deem valuable but reflects the consensus of all instructors who teach organic chemistry at the focal institution. Following unanimous approval of this performance expectation (along with several others), the authors designed an assessment to elicit evidence of student engagement in the specified performance (Figure 3). Response-process validity was established via cognitive interviews conducted with eight consenting students following IRB approval.[38] The wording was slightly altered in response to these interviews to clarify the intent of the prompt.

We chose to center the assessment item on a benzylic alkene addition reaction because, to explain its outcome, students need to grapple with how π-conjugation influences stability and how the relative energies of species present at various points in the reaction influence the process outcome. Attending to how charge delocalization impacts the stability of species in a reaction system is key to predicting and explaining the outcomes of a wide variety of reactions in organic chemistry. This prompt contained four parts. Parts 1 and 2 asked students to provide electron-pushing mechanisms that would account for the formation of Product A and Product B, respectively. In Part 3, students were given a potential energy surface with a curve depicting the formation of Product A. They were asked to add a second curve representing the formation of Product B. Responses to this part of the prompt were not analyzed in this study for two reasons. First, the different learning environments varied considerably in how much students were expected to construct and use potential energy surfaces whereas all of them emphasized electron-pushing mechanisms. Second, a cursory examination of responses suggested that we would not obtain much additional information beyond what the mechanisms and explanations provided. In the final part of the prompt, students were asked to use the mechanisms and potential energy surface they drew to state whether they agreed with the claim that Product B would be the major product and to explain their reasoning.



Consider the reaction between hydrobromic acid (HBr) and the alkene shown below. You read a claim that **Product B** is the major product of this reaction.

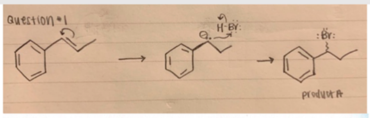
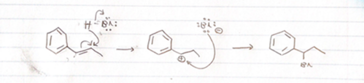
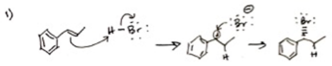
1. Draw a mechanism for the reaction above that leads to **Product A**.
2. Draw a mechanism for the reaction above that leads to **Product B**.
3. The potential energy surface for the pathway leading from reactants to **Product A** is drawn below. Using a dashed line, draw the potential energy surface for the pathway leading from reactants to **Product B** on the same axes. Label all intermediates and products.

4. Would you expect **Product B** to be the major product of this reaction? Explain your answer using the mechanisms and potential energy surface you drew in parts 1–3.

Figure 3. Assessment item used to elicit evidence of how students connect structure- and energy-related ideas to rationalize the outcome of a benzylic alkene addition reaction.

A complete, canonical answer to the prompt shown in Figure 3 is provided in the SI (Figure S4), but we include here an example of how one might productively connect knowledge of structure and energy to explain the outcome of the reaction. This reaction is kinetically controlled (irreversible), so the product that is formed most quickly (i.e., the reaction path with the lowest energy barrier to the rate-determining step) will predominate. The formation of a carbocation intermediate is the rate-determining step en route to both possible products. We can use the relative energies of the carbocations as a proxy for the relative energies of the transition states for the rate-determining step of each process (Hammond's postulate). As the starting

Table 1. Descriptions of Mechanism Codes and Examples of Students' Responses

Code	Description	Examples
1	Incorrect arrows and incorrect intermediates	
2	Correct arrows or correct intermediates	
3	Correct arrows and correct intermediates	

system is the same for both reaction paths, the path that proceeds through the higher-energy carbocation would be expected to have the higher energy barrier to the rate-determining step and thus be slower than the competing path. We can determine the relative energies of the intermediates by comparing their structures. The carbocation leading to Product A (Carbocation A) is stabilized by π-conjugation with the neighboring aromatic ring, resulting in delocalization of the positive charge. The carbocation leading to Product B (Carbocation B) is not in conjugation with the aromatic ring, so it is less stable (i.e., higher in energy). Since Carbocation A is lower in energy than Carbocation B, Transition State A should be lower in energy than Transition State B, which means that Product A should be the major product. The claim in the prompt is thus incorrect. In short, the relative energies of the first transition states determine the reaction outcome, and an examination of structural features that act to stabilize these transition states (e.g., π-conjugation) allows for a prediction of these relative energies.

Data Collection and Reduction

Permission to administer this assessment item was obtained from each professor. The assessment item was given in OChem II as a standalone homework assignment; a small amount of course credit was given to students for completing the assignment. This assessment was administered approximately halfway through the semester, directly after each learning environment had covered the chapter on the reactivity of benzylic systems. Students were given a week to complete the assignment. Student explanations from each learning environment were exported from the online learning management system into a spreadsheet and subsequently deidentified and randomized. 100 randomly selected student responses from each learning environment were compiled for the analyses described in the next section. Mechanism drawings were exported and renamed using the random ID that corresponded to the explanation.

Characterization of Student-Constructed Mechanisms

Each student submitted two mechanisms, one depicting the formation of Product A and one depicting the formation of Product B. These mechanisms were coded separately using a three-part coding scheme informed by prior work conducted by Grove et al., Crandell et al., and Houchlei et al. in their studies on students' mechanistic reasoning.[6,19,7] Responses with canonically correct arrows and intermediate structures were coded as "3", responses with canonically correct arrows or correct intermediate structures as "2", and responses with neither canonically correct arrows nor correct intermediate structures as "1" (Table 1). Note that we chose to ignore missing formal charges on bromide. All mechanisms were coded independently by the first three authors, and inter-rater agreement was calculated using Fleiss' Kappa.[39] A value of 0.90 was obtained, indicating high agreement.[40] Any discrepancies were resolved, resulting in consensus codes for the entire data set. For the chi-square tests, codes "1" and "2" were collapsed because of their low counts, resulting in a binary mechanism variable (i.e., incorrect and correct). We also elected to only use mechanism codes corresponding to the formation of Product A, since A is the major product, and 89% of responses earned the same code for both mechanism drawing prompts.

Description of Productive Structure-Energy Connections in Student Explanations

Our analysis of student-constructed explanations focused on describing how ideas related to molecular structure and energy were productively activated and connected in the context of the prompt we administered. This analytic focus was borne of the recognition that most phenomena of interest to organic chemists (and hopefully organic chemistry students) can be understood by connecting the structure of entities in a system (e.g., reactants, products, transition states, intermediates) to energetic changes that occur as these entities interact. For example, the energy barrier to produce a secondary nonconjugated carbocation from the benzylic alkene shown in Figure 3 is substantially higher than the corresponding barrier to produce a secondary benzylic carbocation due to the ability of the benzylic carbocation to delocalize charge via π-conjugation. The central importance of connecting structural to energetic accounts in reasoning about organic chemistry phenomena has underpinned past scholarship by Caspari et al. and Bodé et al. as well as contributions by Goodwin on the philosophy of organic chemistry.[41−44]

To describe the invocation of structure- and energy-related knowledge elements by students responding to our task, the first three authors read through the responses and noted patterns in the structural and energetic features discussed. Comparing and contrasting these patterns allowed the team to revise, combine, and collapse descriptive codes so that each code describes a distinct reasoning pattern. The consensus coding scheme that emerged from this dialogue can be found in Table 2. Explanations grounded in structural features other than π-conjugation were coded as "0" as were responses that lacked both structural and energetic components. Responses that described only π-conjugation or the energies of the transition states or intermediates were given a "1". Responses that described π-conjugation and related it to the energy of the

Table 2. Descriptions and Examples of Explanation Codes for RQ2 and RQ3

Code	Description	Examples
0	Does not describe π-conjugation or energy OR attributes energy differences to irrelevant structure feature (e.g., steric repulsion, carbocation substitution)	*I would expect Product B to be the major product in this reaction as it has less steric interaction with the attached benzene ring. Because I predicted less steric interaction, the energy threshold to reach Product B's intermediates and end product are lower than Product A's, so Product B takes less energy to create.*
1	Describes π-conjugation and does not connect to energy or connects to product energy OR describes energy of intermediate and/or transition state without relating to structure	*No, I would expect A to be the major product of the reaction because it involves a more stable carbocation intermediate that which is lower energy and easy to form. Product B has an additional intermediate due to a rearrangement which is higher energy and less stable.*
2	Describes π-conjugation and relates to intermediate energy	*Product B is not the major product because its higher in energy than Product A. A is more favored because the positive charge on the carbon can delocalize within the pi system of the ring which has stabilizing effects and lowers its energy.*
3	Describes π-conjugation and relates to transition state energy	*No, I would not expect Product B to be the major product of this reaction. The intermediate leading to the formation of product B is less stable as the positive charge is not delocalized into the pi system by being at a benzylic position. By Hammonds postulate, the transition state energy would then be higher and product B reaction would happen slower making product A the major product instead.*

intermediate were given a "2" while responses that described π-conjugation and related it to the transition state energy were given a "3". Both "2" and "3" responses represent a productive connection of structural and energetic ideas to explain the outcome of the focal reaction. Note that we conceive of these codes as descriptors representing the activation of knowledge elements in-the-moment rather than judgements of durable "understanding" or "misunderstanding".

Reliably bounding each of the bins described previously involved several rounds of joint coding of responses, discussion of inconsistencies in this coding, and revision of the coding scheme. One outcome of this discussion was the decision to treat descriptions of stability as synonymous with descriptions of energy (i.e., a more stable intermediate is a lower-energy intermediate.) This was done because many organic chemists use these terms interchangeably, including the organic instructors who taught the courses included in our study. We also decided that simply stating that a species was benzylic did not count as describing π-conjugation and that vague descriptions of energy, such as "energy pathway", were assumed to refer to the intermediate energy rather than transition state energy. Mentions of product energy were ignored as this is not relevant to determining the outcome of this reaction. Our iterative, reflective coding process gave us confidence that the codebook was sufficiently detailed to characterize the data. In the end, the entire data set was jointly coded by the first three authors, and consensus was reached on the code best describing each response. A more detailed description of the coding process is available in the SI.

Toward a Comprehensive Description of Structure- and Energy-Ideas Embedded in Student Explanations

After completing the coding process described previously, we recognized that many of the responses coded as "0" contained evidence of knowledge elements that are productive in different contexts. For example, the notion that steric repulsion can impact the energies of intermediates or transition states and influence the outcomes of reactions is a valuable resource for making sense of many organic chemistry phenomena. To gain a holistic sense of the structure- and energy-ideas that students thought useful for explaining benzylic hydrobromination (the aim of RQ 1), we realized that we would need a set of descriptors that are more nuanced than those found in Table 2. Thus, we revisited our coding scheme and endeavored to more thoroughly describe the features of intermediate, product, or transition state structure students connected to relative energy. Discussion of these structural elements led to a coding scheme consisting of four categories: electron/charge delocalization, steric repulsion, carbocation substitution, and other/none (Table 3). We then examined which molecular species students referred to when describing relative energy or stability, leading to the development of the five energy codes shown in Table 3. Using this expanded coding scheme, we were able to more precisely describe students' ideas and illustrate how they were connecting structural differences to energetic differences as part of predicting a reaction outcome. However, we acknowledge that, at the undergraduate level, constructing a correct explanation (i.e., one that aligns with scientific canon) is important. We thus maintained the original 4-bin coding scheme displayed in Table 2 for addressing RQ2 and RQ3.

Associations between learning environment enrollment and the distribution of codes describing student explanations were examined using a series of Pearson's chi-square tests. An

Table 3. Structure and Energy Codes, along with Examples, for RQ1

Structure Code	Example
Electron/ Charge Delocalization	*No, I would not expect it to be the major product. The carbocation intermediate for product B should be less stable because the positive charge cannot be delocalized. The carbocation intermediate for A is* ***resonance stabilized by the benzylic ring adjacent. The ability to delocalize positive charge is very stabilizing.*** *Since the carbocation intermediate for B might be less stable, by Hammond's postulate the transition state for B is higher, so A would form faster.*
Steric Repulsion	*Product A is the favored product because it has a lower energy potential than product B. This is because there is a* ***lower steric hindrance*** *with the Br addition in product A than in product B.*
Carbocation Substitution	*I would expect Product A to be the major product of the reaction because the* ***carbonation [sic] intermediate is tertiary, which is more stable than the secondary carbonation [sic] intermediate in Product B.*** *The intermediate of Product A has lower energy because it is more stable this will yield the major product.*
Other/None	*You would expect that B would be the major project as H−Br goes in from a backside attack to flip the bond so Br is facing downward. To get the other reaction would require a different resonance structure that is higher in energy*
	The reaction is kinetically controlled so because product B has a lower-energy intermediate, that intermediate will lead to the major product.

Energy Code	Example
Intermediate	*No, product A would be the major product because it goes through the faster path of the* ***more stable intermediate.*** *The positive charge on the intermediate can be delocalized to the ring, so* ***this intermediate is stabilized by resonance and therefore has a lower energy.***
Transition State (with or without intermediate)	*I would not expect product B to be the major product of this reaction because product A carbocation has pi conjugation with the ring which allows for distribution of the positive charge. This also allows product a to have* ***a lower transition state and energy of the intermediate*** *through pi conjugation.*
Intermediate and/or Transition State + Product	*Product A would be the major product of this reaction because the Br being closer to the ring allows for more stabilization of the carbocation intermediate. The ring is able to delocalize the positive change and is therefore a* ***more stable intermediate and product.***
Product	*I would not expect it to be the major product. I would expect product A to be the major product because there is opportunity for pi-conjugation between the carbon the Br is bonded to and the benzene ring.* ***This makes the molecule more stable then product B.***
None/Ambiguous	*I would not expect product B to be the major product since it would have a* ***higher energy pathway.*** *This is because the carbocation cannot delocalize with the pi conjugation from the ring as it would in product A.*

analogous approach was used to examine associations between the correctness of drawn mechanisms and the distribution of codes describing student explanations. All Pearson's chi-square tests were conducted using SPSS.[45] The output of each chi-square test included χ^2 and Cramer's V. Cohen's guidelines for effect size were used to interpret the value of Cramer's V.[46] For a contingency table containing three rows, values of 0.071, 0.212, and 0.345 for Cramer's V correspond to small, medium, and large effect sizes, respectively. The threshold of significance used for this study was $p \leq 0.01$. A post hoc analysis of each chi-square test which showed a significant association between variables was conducted in order to support inferences about the driver(s) of that significance. Standardized residuals for each cell were calculated by SPSS. Standardized residuals with positive values indicated more counts than expected by chance while residuals with negative values indicated fewer counts than expected by chance.[46] The magnitude of the standardized residual was compared to a critical value, which was 2.58 for a threshold of significance of $p \leq 0.01$.[47,48] Thus, cells with standardized residuals larger than 2.58 or smaller than −2.58 were considered drivers of the significant association.

■ FINDINGS

RQ1: What Knowledge Elements Did Students Activate and Connect When Asked to Explain the Outcome of a Benzylic Alkene Addition Reaction?

When justifying why one alkene addition product would predominate over the other, students commonly activated ideas related to charge/electron delocalization, steric repulsion, and carbocation substitution (Figure 4). Nearly two-thirds of the students attended to differences in charge or electron distribution. Most responses of this sort compared the delocalization of positive charge between the two potential carbocation intermediates. A typical response of this nature is shown here:

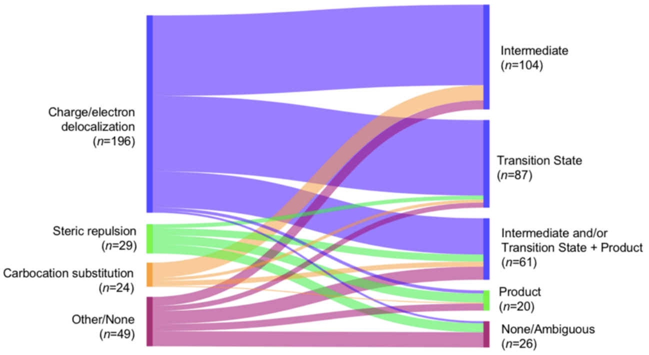


Figure 4. Sankey diagram depicting how students connected structural features (left) to the stabilities or relative energies of species (right) involved in the benzylic alkene hydrobromination reaction.

> *I would expect product A to be the major product because the [sic] proceeds down the lower energy pathway. Product A has a more stabilized intermediate because the positive charge is delocalized around the ring while in product B it cannot be delocalized because the positive charge is separated from the π conjugated ring by an sp3 hybridized carbon therefore raising the energy of the intermediate and favoring product A.*

A few students described how the bromine atom would donate or withdraw electron density. An activation of ideas related to substituent donation or withdrawal of negative charge may have occurred due to the utility of these ideas in explaining electrophilic aromatic substitution (EAS) reactions, which was the focus of the chapter prior to the chapter on reactions of benzylic systems. Most explanations of why one EAS product would be favored over other alternatives are grounded in how aryl substituents affect reactivity by altering the electron density of the ring.

The second most common factor cited in students' explanations was steric repulsion between the large bromine atom and the aryl ring. An example of an explanation utilizing steric repulsion is given here:

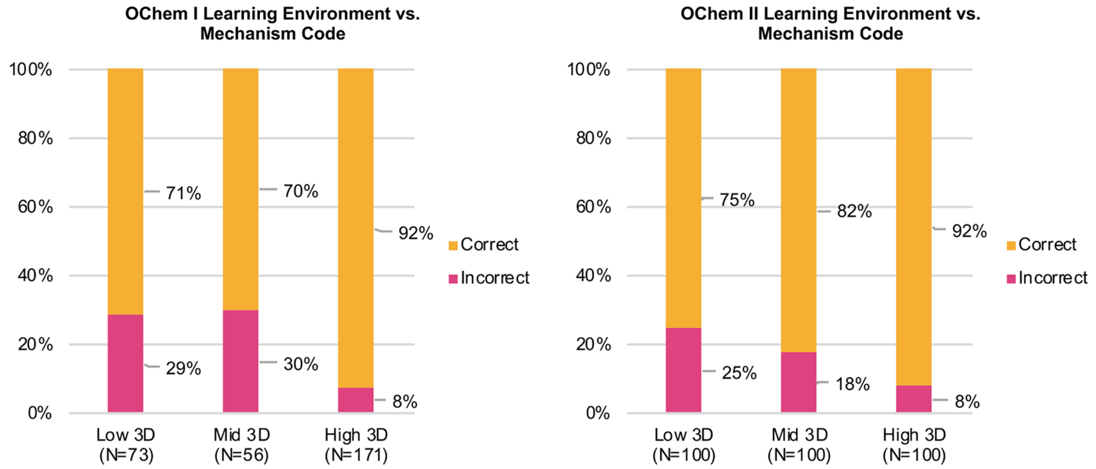


Figure 5. Distribution of mechanisms codes according to learning environment for OChem I (left) and OChem II (right). The vast majority of students in all learning environments drew the mechanism correctly.

> *I would expect Product B to be the major product mainly due to sterics. Br is a large atom and being closer to the aromatic ring may cause some steric repulsion, making the molecule more unstable and the intermediate higher energy. Therefore, the reaction make product B would include less sterics than A and thus, would have a lower-energy intermediate, form more easily, and have a more stable product.*

Steric repulsion is often a productive resource for rationalizing observations in organic chemistry, and at this point in the course, students had encountered it in the context of alkane conformations, other alkene addition reactions, alkyl halide reactions, and electrophilic aromatic substitution reactions. Indeed, the linkage between steric repulsion and the relative energy of intermediates described by the prior response is generally correct and very often useful! However, in reactions that proceed through carbocation intermediates, such as this one, steric repulsion has a minimal effect on the outcome because the nucleophile (i.e., bromide) reacts from above or below the planar reactive carbon.

The other structural feature invoked in several responses was carbocation substitution. In reactions that proceed through carbocation intermediates, the degree of substitution often dictates the regioselectivity of the reaction since increased substitution allows for increased hyperconjugation, which stabilizes the carbocation via charge delocalization. A tag of “carbocation substitution” denotes an invocation of intermediate substitution without linking this to electron delocalization (which was a separate tag). Accordingly, it is likely that many students whose responses were described by this code used “more substituted carbocations are more stable” as a heuristic.[49] Since both potential carbocation intermediates in this reaction are secondary, substitution is not a useful means of discriminating between the two reaction pathways referenced in our diagnostic prompt. Some students recognized this and concluded that the products would be formed in roughly equal amounts, as the following response illustrates: “I would expect both of the products to have an equal outcome because they are both secondary carbocations.” Other students believed that the carbocation leading to Product A was tertiary or “closer to being tertiary”, presumably because it was next to the larger aryl group.

To connect differences in structure to reaction outcome, most students included an energetic component to their explanations (Figure 4). Of the students who described charge/electron delocalization, most connected it to the stability of the intermediate, as shown in the first quote. In addition, a large number of students connected electron delocalization to the energy of the transition states, as follows:

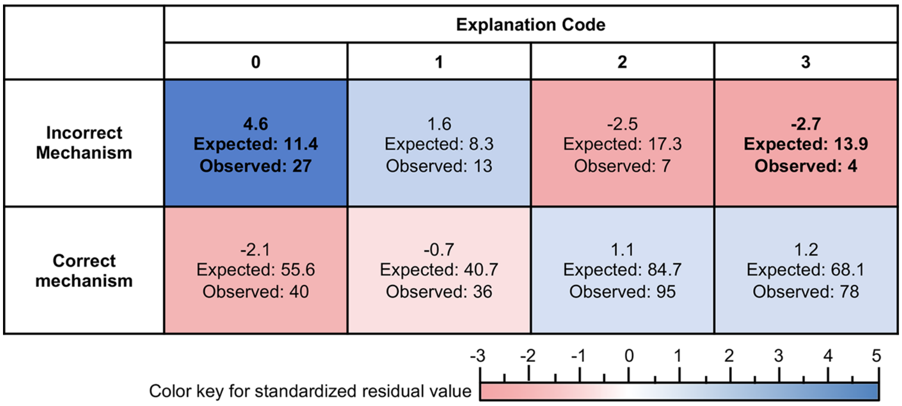
> *No, I would not expect it to be the major product. The carbocation intermediate for product B should be less stable because the positive charge cannot be delocalized. The carbocation intermediate for A is resonance stabilized by the benzylic ring adjacent. The ability to delocalize positive charge is very stabilizing. Since the carbocation intermediate for B might be less stable, by Hammond’s postulate the transition state for B is higher, so A would form faster.*

Only a handful of students used electron delocalization to justify a difference in product energies. More commonly, students connected electron delocalization to the intermediate and/or transition state along with product energy:

> *I would NOT expect product B to be the major product of this reaction. This is because when the carbocation intermediate is formed for the reaction that leads to product A, it is a more stable carbocation than in B. I determined this because the positive charge is located on the benzylic carbon in reaction A, which is very stabilizing due to it being in pi conjugation with the benzene ring. The positive charge is close enough to the ring that it can be delocalized in the ring which is very stabilizing. Therefore, since the carbocation is more stable in reaction A, this would mean that it would produce the more stable product.*

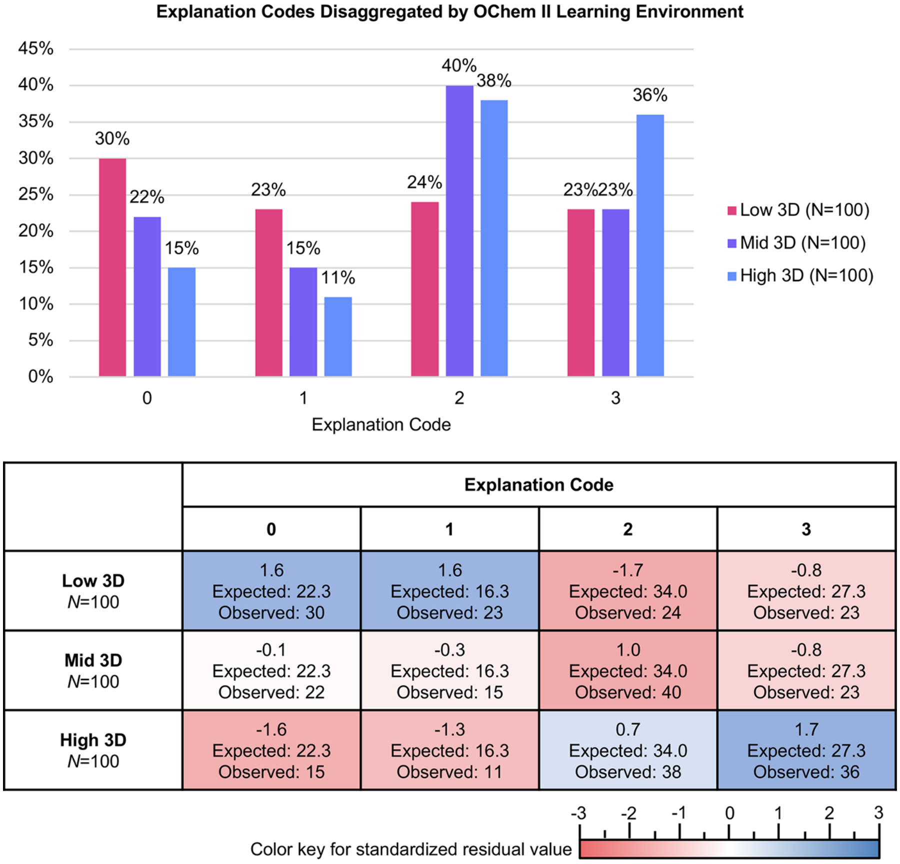
Finally, a few of the students who invoked electron delocalization either described energy in ambiguous terms (e.g., “lower-energy pathway”) or did not discuss energy or stability at all in their explanation. Attributing the outcome to differences in steric repulsion was mostly associated with descriptions of the relative energy difference between the two possible products. Unsurprisingly, carbocation substitution was mostly connected to the stability of the carbocation intermediate.

Overall, an analysis of explanations revealed that students possess useful ideas of how structure impacts the stability of a molecule and how relative energy at various points along competing reaction pathways dictates the outcome. From an instructional standpoint, these results suggest that it may be productive to prompt students to consider the relative impacts



	Explanation Code			
	0	1	2	3
Incorrect Mechanism	**4.6** **Expected: 11.4** **Observed: 27**	1.6 Expected: 8.3 Observed: 13	-2.5 Expected: 17.3 Observed: 7	**-2.7** **Expected: 13.9** **Observed: 4**
Correct mechanism	-2.1 Expected: 55.6 Observed: 40	-0.7 Expected: 40.7 Observed: 36	1.1 Expected: 84.7 Observed: 95	1.2 Expected: 68.1 Observed: 78

Figure 6. Contingency table for the χ^2 test examining the association between mechanism code and explanation code. Standardized residuals and expected and observed counts are reported in each cell. Standardized residuals greater than the critical value (±2.58) are in bold. Cells with positive standardized residuals are shaded blue while cells with negative standardized residuals are shaded red.



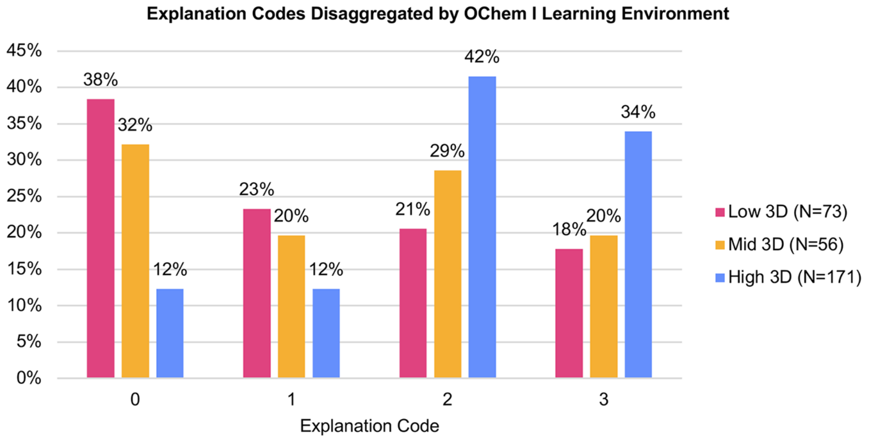
	Explanation Code			
	0	1	2	3
Low 3D *N*=100	1.6 Expected: 22.3 Observed: 30	1.6 Expected: 16.3 Observed: 23	-1.7 Expected: 34.0 Observed: 24	-0.8 Expected: 27.3 Observed: 23
Mid 3D *N*=100	-0.1 Expected: 22.3 Observed: 22	-0.3 Expected: 16.3 Observed: 15	1.0 Expected: 34.0 Observed: 40	-0.8 Expected: 27.3 Observed: 23
High 3D *N*=100	-1.6 Expected: 22.3 Observed: 15	-1.3 Expected: 16.3 Observed: 11	0.7 Expected: 34.0 Observed: 38	1.7 Expected: 27.3 Observed: 36

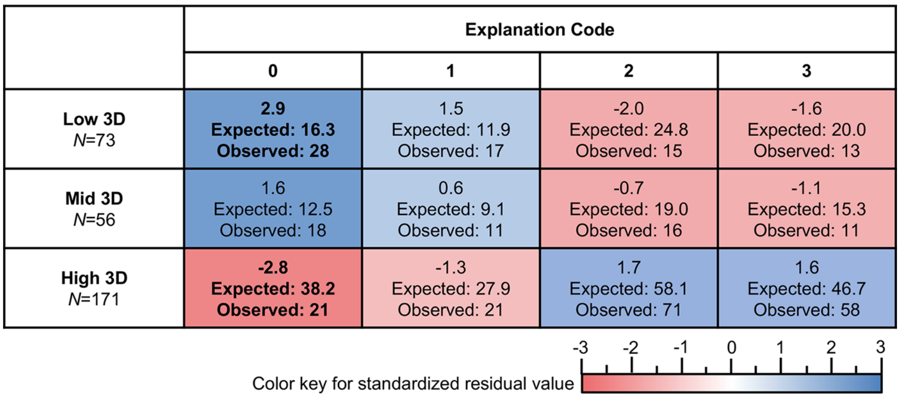
Figure 7. Distribution of explanation codes according to OChem II learning environment (top). Contingency table for the χ^2 test examining the association between OChem II learning environment and explanation code (bottom). Standardized residuals and expected and observed counts are reported in each cell. Standardized residuals greater than the critical value (±2.58) are in bold. Cells with positive standardized residuals are shaded blue while cells with negative standardized residuals are shaded red.

of various factors on stability and to consider if and how intermediate and/or transition state energy, product energy, and reaction outcome are interrelated. As demonstrated, this type of analysis provides more insight into how instruction can build on students' ideas than an analysis that identifies students' responses as merely "right" or "wrong".

RQ2: How Did the Electron-Pushing Mechanisms Students Drew Relate to the Explanations They Provided?

Most students across all learning environments drew mechanisms with correct intermediates and electron-pushing arrows (Figure 5). Slight differences were observed between the aggregate distributions for Mechanism A and Mechanism B. Students who earned different codes for A and B tended to leave off an arrow in one of the mechanisms. Two students altered the reaction conditions for Mechanism B by adding in light or peroxides to render it a radical reaction, which would favor the





	Explanation Code			
	0	1	2	3
Low 3D *N*=73	**2.9** **Expected: 16.3** **Observed: 28**	1.5 Expected: 11.9 Observed: 17	-2.0 Expected: 24.8 Observed: 15	-1.6 Expected: 20.0 Observed: 13
Mid 3D *N*=56	1.6 Expected: 12.5 Observed: 18	0.6 Expected: 9.1 Observed: 11	-0.7 Expected: 19.0 Observed: 16	-1.1 Expected: 15.3 Observed: 11
High 3D *N*=171	**-2.8** **Expected: 38.2** **Observed: 21**	-1.3 Expected: 27.9 Observed: 21	1.7 Expected: 58.1 Observed: 71	1.6 Expected: 46.7 Observed: 58

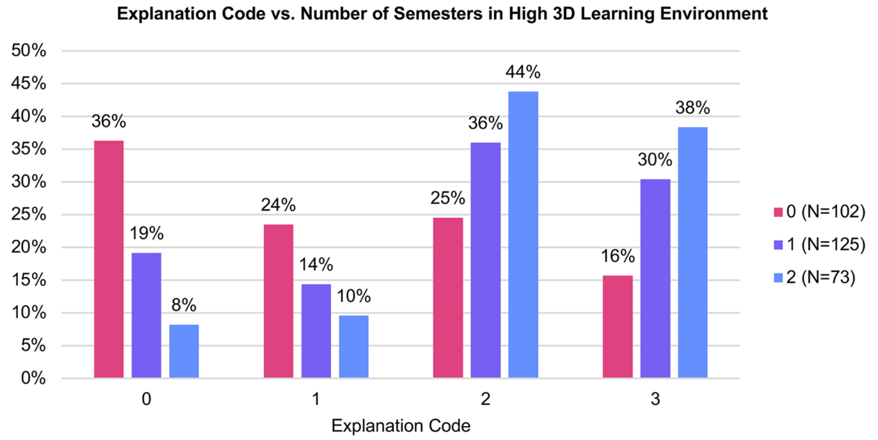
Figure 8. Distribution of explanation codes according to OChem I learning environment (top). Contingency table for the χ^2 test examining the association between OChem I learning environment and explanation code (bottom). Standardized residuals and expected and observed counts are reported in each cell. Standardized residuals greater than the critical value (±2.58) are in bold. Cells with positive standardized residuals are shaded blue while cells with negative standardized residuals are shaded red.

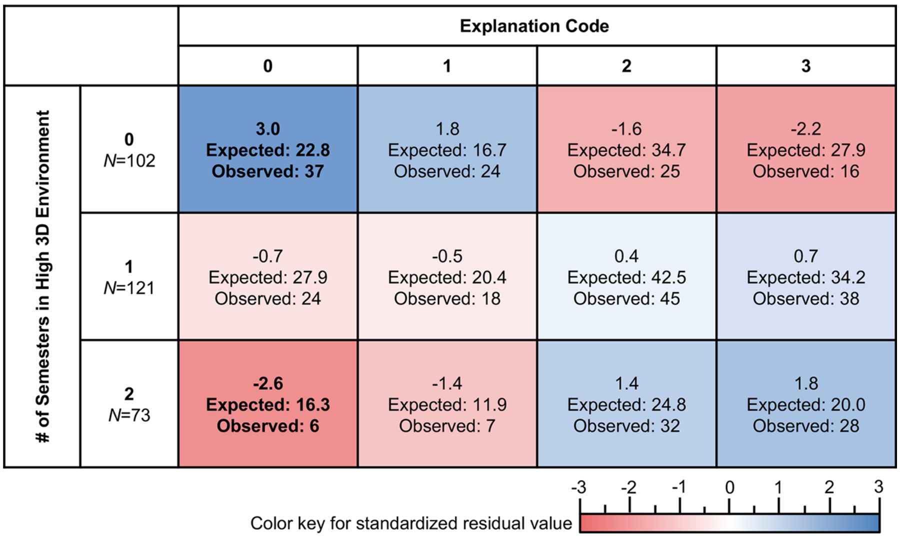
formation of Product B. Disaggregating by learning environment revealed a significant association between OChem I enrollment and mechanism code [$\chi^2(2) = 24.9$, $p < 0.001$, Cramer's V = 0.29, medium effect]. A post hoc analysis of the results of this test revealed that a negative association between enrollment in the high 3D learning environment and drawing an incorrect mechanism was the primary driver of significance. That is, students who were enrolled in the high 3D environment for OChem I were substantially *less* likely to draw an incorrect mechanism than would be expected by chance. A significant association between OChem II enrollment and mechanism code was also observed [$\chi^2(2) = 10.3$, $p < 0.001$, Cramer's V = 0.19, small effect]. A post hoc analysis of the results of this test showed no primary driver(s) of the significant association.

Next, we examined the relationship between mechanism code and explanation code. A chi-square test revealed a significant association with medium effect size: $\chi^2(3) = 44.9$, $p < 0.001$, Cramer's V = 0.387. A post hoc analysis of this chi-square test showed that the significant association was driven by students who drew incorrect mechanisms (Figure 6). Students who drew incorrect mechanisms were substantially *more* likely to provide explanations that were coded as "0" and substantially *less* likely to provide explanations that were coded as "3" than would be expected by chance. In fact, of the 51 students who drew an incorrect mechanism, only four offered an explanation coded as "3". However, for students who drew the correct mechanism, there were no substantive differences between expected and observed counts for each explanation code. This means that a student was unlikely to provide a higher code explanation without the correct mechanism, but drawing a correct mechanism was no guarantee that they could explain the meaning underpinning that mechanism. This is consistent with a large body of prior research.[6,7,9,50−52]

RQ3: How Did the Intellectual Work Emphasized and Rewarded on Assessments Given in Each Learning Environment Relate to the Structure–Energy Connections Invoked in Student Explanations?

To examine whether students who were routinely expected to construct explanations on assessments were more likely to productively connect structure- and energy-ideas when responding to our prompt, student responses were disaggregated by learning environment. Initially, we separated student responses according to OChem II learning environment enrollment, which corresponds to the semester in which the prompt was administered (Figure 7). A significant association with a small effect size was found between OChem II learning environment enrollment and the distribution of explanation codes: $\chi^2(6) = 18.2$, $p = 0.006$, Cramer's V = 0.17. A post hoc analysis of the results of this chi-square test found that none of the cells strongly drove the significant association; that is, no cells had a standardized residual greater than 2.58 in magnitude (Figure 7). We then disaggregated responses by OChem I learning environment enrollment and repeated this analysis (Figure 8). A significant association between OChem I learning environment enrollment and explanation code was found, $\chi^2(6) = 36.4$, $p < 0.001$, Cramer's V = 0.25 (medium effect size). A post





		Explanation Code			
		0	**1**	**2**	**3**
# of Semesters in High 3D Environment	**0** *N*=102	**3.0** **Expected: 22.8** **Observed: 37**	1.8 Expected: 16.7 Observed: 24	-1.6 Expected: 34.7 Observed: 25	-2.2 Expected: 27.9 Observed: 16
	1 *N*=121	-0.7 Expected: 27.9 Observed: 24	-0.5 Expected: 20.4 Observed: 18	0.4 Expected: 42.5 Observed: 45	0.7 Expected: 34.2 Observed: 38
	2 *N*=73	**-2.6** **Expected: 16.3** **Observed: 6**	-1.4 Expected: 11.9 Observed: 7	1.4 Expected: 24.8 Observed: 32	1.8 Expected: 20.0 Observed: 28

Figure 9. Distribution of explanation codes according to number of semesters enrolled in the high 3D learning environment (top). Contingency table for the χ^2 test examining the association between number of semesters enrolled in high 3D learning environment and explanation code (bottom). Standardized residuals and expected and observed counts are reported in each cell. Standardized residuals greater than the critical value (±2.58) are in bold. Cells with positive standardized residuals are shaded blue while cells with negative standardized residuals are shaded red.

hoc analysis of the results of the chi-square test showed that students enrolled in the low 3D learning environment were substantially *more* likely than would be expected by chance to construct an explanation coded as "0" while students enrolled in the high 3D learning environment were substantially *less* likely than would be expected by chance to construct an explanation coded as "0" (Figure 8).

We hypothesize that the more substantive association between OChem I section enrollment and explanation code distribution is due to the foundational nature of the first semester course. Most broadly useful explanatory ideas are introduced in OChem I, including π-conjugation, hyperconjugation, and steric repulsion. In addition, although the reaction used in our diagnostic assessment is revisited in OChem II in the context of benzylic reactivity, electrophilic addition to alkenes is first introduced early in OChem I and could be explained at that time. We should note that students who were enrolled in the low 3D learning environment for OChem I did not have the opportunity to use π-conjugation as an explanatory idea until the end of the course. It is conceivable that these students might be less inclined to integrate π-conjugation into their explanations on account of having less experience doing so than other study participants.

Finally, we wanted to examine how learning environment enrollment over the two-semester introductory organic sequence related to the distribution of codes describing student explanations. Specifically, we looked at how the number of semesters students spent in the high 3D learning environment related to the distribution of codes describing their explanations (Figure 9). A significant association with medium effect size was observed: $\chi^2(6) = 35.2$, $p < 0.001$, Cramer's V = 0.24. A post hoc analysis of the results of this test revealed that the vast majority of students who were enrolled in the high 3D learning environment for both semesters (82%) connected π-conjugation to intermediate or transition state energy in their explanation (Figure 9). By contrast, students who spent two semesters in a course that rarely (or never) emphasized explanations on exams (i.e., low or mid 3D learning environment) were substantially less likely to connect differences in π-conjugation to differences in transition state energies in their explanations.

The differences in explanation code distributions among learning environments point toward the importance of assessment emphasis in messaging course priorities. While the general format of the learning environments was the same (i.e., recorded lectures, virtual instructor-led problem-solving sessions, weekly discussions, four exams), and all sections made use of the same order of topics and textbook, the performances signaled as important differed markedly. To earn high marks in the high 3D learning environment, students were required to use models to predict and explain reaction outcomes. By contrast, students could succeed perfectly well in the low 3D environment without ever connecting structure- and energy-ideas to why phenomena happen. Given this difference in what was emphasized and rewarded on exams, it makes sense that we found that students who were enrolled in the high 3D learning environment were substantially *less* likely to receive a code of "0" on their explanations. Conversely, students who enrolled in a lecture environment where they were never asked to provide explanations for phenomena on assessments were significantly *more* likely to receive a code of "0" on their explanations. While our study design prevents us from determining if differences in assessments *caused* the differences in explanation code distributions, this work supports the notion that assessments convey strong messages to students regarding the intellectual work central to organic chemistry. If we want to support students in constructing explanations for how and why observable events happen the way they do, we likely need to reward them for doing so throughout the course.

Overall, most students were able to generate an explanation for the outcome of the focal benzylic alkene hydrobromination reaction by using charge/electron delocalization to account for differences in intermediate or transition state energies. Since many phenomena central to organic chemistry can be explained, at least in part, by charge delocalization, it is encouraging that most students activated these ideas when justifying their claim as to the reaction outcome. Whether the stabilizing influence of electron delocalization is viewed as a useful explanatory resource in the context of other reactions remains to be seen. Some of the other factors invoked by students, such as the stabilization from increased substitution on a carbocation or the destabilizing influence of steric repulsion, demonstrate that many of the students who did not discuss π-conjugation possess useful resources for understanding other reaction systems.

LIMITATIONS

We cannot directly observe students' thinking and must use their drawings and writings to infer what they know and can do in a given context. Due to the inherent restrictions of a virtual semester, we do not know how many of the students worked together to complete the assignment, so not all responses may be indicative of an individual student's thinking. Furthermore, this study focused on student responses to a single prompt at a single point in the semester. If we were to administer this prompt at a different time, under different conditions, or with different wording, students may exhibit new patterns of resource activation. As an example, the potential energy surface question (part 3) may have cued different resources for students in the high 3D learning environment, which emphasized use of potential energy surfaces, compared to students in the other learning environments, which did not. Similarly, if we were to repeat this study using a diagnostic prompt focused on a different phenomenon, we do not know if we would observe the same associations between learning environment enrollment and student explanations. Thus, we are not suggesting that one learning environment is definitively "better" than another. Furthermore, our analysis of the different learning environments only characterized assessment emphasis. While the formats of the courses were similar (e.g., video lectures, weekly discussion sections), we did not examine the emphases instructors placed on 3D performances during lecture. It is possible that the instructors differed in how they modeled the practice of constructing explanations, which may have contributed to the differences we observed. Finally, we have no evidence as to how students were framing their engagement with the assessment item. The way students understand the aim of explanation construction, the appropriate sources of knowledge to draw from, and what constitutes a credible justification shape the degree to which they perceive the task as a "school science" exercise versus an opportunity to make sense of a phenomenon.[53,54] Given that one's framing is largely dependent on past experience, it is likely that students viewed the diagnostic task in a similar manner to other problems given during the course, in which the goal is rapidly producing a "correct answer".[55,56] Future work will focus on methods for characterizing and influencing students' frames in order to support students in engaging more authentically with organic chemistry.

CONCLUSION

In this study, we identified the structure−energy connections students made to explain the outcome of an alkene addition reaction, explored how their explanations related to their mechanisms, and examined the relationship between course assessment emphasis and explanations. We found that most students justified their outcome predictions based on the relative energies of intermediates or transition states caused by differences in the extent of electron delocalization. Generally useful ideas relating to steric repulsion and degree of substitution on carbocation intermediates were also observed, illustrating that students who do not arrive at a canonically correct answer may possess productive ideas upon which to build. Unsurprisingly, several of these ideas (e.g., steric repulsion, carbocation substitution, activation energy) matched the concepts Bodé et al. found embedded in students' arguments pertaining to an S_N1 reaction, which also proceeds though a carbocation intermediate.[42] Our analysis indicates that we should not assume that students who draw a correct mechanism understand why the reaction proceeds in the manner illustrated. In-line with the previously published literature, we found no substantive association between a correctly drawn mechanism and any particular explanation code.[7−9] We did, however, find a significant association between learning environment enrollment and explanation code with a medium effect size. Students who were routinely expected to construct explanations on assessments tended to connect electron delocalization to intermediate or transition state energy in the context of our prompt.

Since constructing explanations is central to the work of chemists, we should design learning environments to support students in this practice.[4] Studies have provided evidence demonstrating the positive impact curricular reforms can have toward achieving this goal, but the impact transforming assessments in an otherwise traditional course cannot be determined from these studies.[57,19,58] Our study begins to address that gap in the literature. Based on our findings, we will continue to probe the relationship between assessment emphasis and students' propensity to construct productive causal accounts for phenomena, as transformation of assessments appears to be a potentially productive avenue for reform. In particular, we are interested in examining students' reasoning using structure−energy connections across multiple prompts over a year of introductory organic chemistry instruction.

Implications for Research

This study lends further support to the oft-repeated claim that assessments signal to students "what counts" in a learning environment.[13−18] Relatedly, the measures researchers use to determine the success of interventions or transformations message the sorts of performances those researchers think are important. We argue that researchers should take the nature of assessments into account when evaluating reform efforts. Noting some sort of improvement in course grade or exam performance means very little if we do not know what "success" meant in the course or on the exam. As the format and content of exams may vary considerably, the nature of the assessment items on those exams should be reported so that the reader knows precisely what sort of performances are influenced by the intervention. Relatedly, we argue that researchers should prioritize the use of assessment items that have the potential to elicit detailed evidence of engagement in aspects of "doing science", such as constructed-response items that ask for the reasoning supporting a claim. Items that require only a claim, including those that instruct students to draw a mechanism, cannot support inferences as to why students claimed what they did.

Implications for Instruction

Assessments are powerful tools that instructors of chemistry can use to shape learning. We urge instructors to reflect on whether the performances they award points to align with intellectual work central to the discipline under study. Do assessment tasks promote the construction of causal accounts for phenomena or a reliance on pattern recognition? We also suggest that instructors who want to improve students' engagement in "doing chemistry" but are unable to implement whole curricular reforms consider how they might modify their formative and summative assessments in order to support a coherent emphasis on constructing explanations for phenomena. Although assessment reform is nontrivial and may increase the grading burden, it is vital that "success" in organic chemistry courses aligns with productive engagement in intellectual work characteristic of the discipline.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.1c01080.

Example 3D and non-3D assessment items, a canonical answer to the assessment item used in this study, and a detailed description of the explanation coding process (PDF, DOCX)

OChem I and OChem II exams from a high 3D Learning Environment (PDF)

Assessment response codes and inter-rater agreement calculations (XLSX)

3D-LAP item analysis (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Ryan L. Stowe − *Department of Chemistry, University of Wisconsin−Madison, Madison, Wisconsin 53706, United States;* orcid.org/0000-0002-5548-495X; Email: rstowe@chem.wisc.edu

Authors

Kimberly S. DeGlopper − *Department of Chemistry, University of Wisconsin−Madison, Madison, Wisconsin 53706, United States*

Cara E. Schwarz − *Department of Chemistry, University of Wisconsin−Madison, Madison, Wisconsin 53706, United States;* orcid.org/0000-0003-0552-8886

Niall J. Ellias − *Department of Chemistry, University of Wisconsin−Madison, Madison, Wisconsin 53706, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jchemed.1c01080

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Support for this work was provided by the Office of the Vice-Chancellor for Research and Graduate Study at the University of Wisconsin—Madison with funding from the Wisconsin Alumni Research Foundation. We would like to thank all instructors who provided their exams for us to analyze.

■ REFERENCES

(1) The Nobel Committee for Chemistry. *Enamine and Iminium Ion-Mediated Organocatalysis*; Royal Swedish Society of Chemistry, 2021.

(2) Manz, E.; Lehrer, R.; Schauble, L. Rethinking the Classroom Science Investigation. *Journal of Research in Science Teaching* **2020**, *57* (7), 1148−1174.

(3) Schwarz, C. V.; Passmore, C.; Reiser, B. J. *Helping Students Make Sense of the World Using Next Generation Science and Engineering Practices*; NSTA Press, 2017.

(4) The National Research Council. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*; National Academies Press: Washington, DC, 2012.

(5) Stowe, R. L.; Cooper, M. M. Practicing What We Preach: Assessing "Critical Thinking" in Organic Chemistry. *J. Chem. Educ.* **2017**, *94* (12), 1852−1859.

(6) Grove, N. P.; Cooper, M. M.; Rush, K. M. Decorating with Arrows: Toward the Development of Representational Competence in Organic Chemistry. *J. Chem. Educ.* **2012**, *89* (7), 844−849.

(7) Houchlei, S. K.; Bloch, R. R.; Cooper, M. M. Mechanisms, Models, and Explanations: Analyzing the Mechanistic Paths Students Take to Reach a Product for Familiar and Unfamiliar Organic Reactions. *J. Chem. Educ.* **2021**, *98*, 2751.

(8) Caspari, I.; Weinrich, M. L.; Sevian, H.; Graulich, N. This Mechanistic Step Is "Productive": Organic Chemistry Students' Backward-Oriented Reasoning. *Chem. Educ. Res. Pract.* **2018**, *19* (1), 42−59.

(9) Bhattacharyya, G.; Bodner, G. M. It Gets Me to the Product": How Students Propose Organic Mechanisms. *J. Chem. Educ.* **2005**, *82* (9), 1402.

(10) Popova, M.; Bretz, S. L. Organic Chemistry Students' Interpretations of the Surface Features of Reaction Coordinate Diagrams. *Chem. Educ. Res. Pract.* **2018**, *19* (3), 919−931.

(11) Popova, M.; Bretz, S. L. Organic Chemistry Students' Challenges with Coherence Formation between Reactions and Reaction Coordinate Diagrams. *Chem. Educ. Res. Pract.* **2018**, *19* (3), 732−745.

(12) Lamichhane, R.; Reck, C.; Maltese, A. V. Undergraduate Chemistry Students' Misconceptions about Reaction Coordinate Diagrams. *Chem. Educ. Res. Pract.* **2018**, *19* (3), 834−845.

(13) Momsen, J.; Offerdahl, E.; Kryjevskaia, M.; Montplaisir, L.; Anderson, E.; Grosz, N. Using Assessments to Investigate and Compare the Nature of Learning in Undergraduate Science Courses. *LSE* **2013**, *12* (2), 239−249.

(14) Scouller, K. The Influence of Assessment Method on Students' Learning Approaches: Multiple Choice Question Examination versus Assignment Essay. *Higher Education* **1998**, *35* (4), 453−472.

(15) Scouller, K. M.; Prosser, M. Students' Experiences in Studying for Multiple Choice Question Examinations. *Studies in Higher Education* **1994**, *19* (3), 267−279.

(16) Snyder, B. *The Hidden Curriculum*; The MIT Press: Cambridge, MA, 1973.

(17) Entwistle, N. J. Approaches to Learning and Perceptions of the Learning Environment. *High Educ* **1991**, *22* (3), 201−204.

(18) Crooks, T. J. The Impact of Classroom Evaluation Practices on Students. *Review of Educational Research* **1988**, *58* (4), 438−481.

(19) Crandell, O. M.; Lockhart, M. A.; Cooper, M. M. Arrows on the Page Are Not a Good Gauge: Evidence for the Importance of Causal Mechanistic Explanations about Nucleophilic Substitution in Organic Chemistry. *J. Chem. Educ.* **2020**, *97* (2), 313−327.

(20) Cooper, M. M.; Stowe, R. L.; Crandell, O. M.; Klymkowsky, M. W. Organic Chemistry, Life, the Universe and Everything (OCLUE): A Transformed Organic Chemistry Curriculum. *J. Chem. Educ.* **2019**, *96* (9), 1858−1872.

(21) Stowe, R. L.; Scharlott, L. J.; Ralph, V. R.; Becker, N. M.; Cooper, M. M. You Are What You Assess: The Case for Emphasizing Chemistry on Chemistry Assessments. *J. Chem. Educ.* **2021**, *98*, 2490.

(22) Cooper, M.; Klymkowsky, M. Chemistry, Life, the Universe, and Everything: A New Approach to General Chemistry, and a Model for Curriculum Reform. *J. Chem. Educ.* **2013**, *90* (9), 1116−1122.

(23) Talanquer, V.; Pollard, J. Let's Teach How We Think Instead of What We Know. *Chem. Educ. Res. Pract.* **2010**, *11* (2), 74−83.

(24) Flynn, A. B.; Ogilvie, W. W. Mechanisms before Reactions: A Mechanistic Approach to the Organic Chemistry Curriculum Based on Patterns of Electron Flow. *J. Chem. Educ.* **2015**, *92* (5), 803−810.

(25) National Research Council. *Knowing What Students Know: The Science and Design of Educational Assessment*; National Academies Press: Washington, D.C., 2001; p 10019. DOI: 10.17226/10019.

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

(27) DiSessa, A. Knowledge in Pieces. In *Constructivism in the Computer Age*; Forman, G. E., Pufall, P. B., Eds.; Lawrence Erlbaum Associates: Hillsdale, NJ, 1998; pp 49−70.

(28) DiSessa, A. A. A History of Conceptual Change Research: Threads and Fault Lines. In *The Cambridge Handbook of the Learning Sciences*; Sawyer, R. K., Ed; Cambridge University Press: New York, 2006; pp 265−281.

(29) Cooper, M. M.; Corley, L. M.; Underwood, S. M. An Investigation of College Chemistry Students' Understanding of Structure−Property Relationships. *Journal of Research in Science Teaching* **2013**, *50* (6), 699−721.

(30) Rosenberg, S.; Hammer, D.; Phelan, J. Multiple Epistemological Coherences in an Eighth-Grade Discussion of the Rock Cycle. *Journal of the Learning Sciences* **2006**, *15* (2), 261−292.

(31) Robertson, A. D.; Atkins, L. J.; Levin, D. M.; Richards, J. L. What Is Responsive Teaching? In *Responsive teaching in science and mathematics*; Robertson, A. D., Scherr, R. E., Hammer, D., Eds.; Routledge, 2015; pp 19−53.

(32) Posner, G. J.; Strike, K. A.; Hewson, P. W.; Gertzog, W. A. Accommodation of a Scientific Conception: Toward a Theory of Conceptual Change. *Science Education* **1982**, *66* (2), 211−227.

(33) Brown, D. E. Students' Conceptions as Dynamically Emergent Structures. *Sci. & Educ* **2014**, *23* (7), 1463−1483.

(34) Loudon, M.; Parise, J. *Organic Chemistry*, 6th ed.; W. H. Freeman and Company, 2016.

(35) Laverty, J. T.; Underwood, S. M.; Matz, R. L.; Posey, L. A.; Carmel, J. H.; Caballero, M. D.; Fata-Hartley, C. L.; Ebert-May, D.; Jardeleza, S. E.; Cooper, M. M. Characterizing College Science Assessments: The Three-Dimensional Learning Assessment Protocol. *PLoS One* **2016**, *11* (9), No. e0162333.

(36) Stowe, R. L.; Esselman, B. J.; Ralph, V. R.; Ellison, A. J.; Martell, J. D.; DeGlopper, K. S.; Schwarz, C. E. Impact of Maintaining Assessment Emphasis on Three-Dimensional Learning as Organic Chemistry Moved Online. *J. Chem. Educ.* **2020**, *97* (9), 2408−2420.

(37) Mislevy, R. J.; Haertel, G. D. Implications of Evidence-Centered Design for Educational Testing. *Educational Measurement: Issues and Practice* **2006**, *25* (4), 6−20.

(38) Arjoon, J. A.; Xu, X.; Lewis, J. E. Understanding the State of the Art for Measurement in Chemistry Education Research: Examining the Psychometric Evidence. *J. Chem. Educ.* **2013**, *90* (5), 536−545.

(39) Fleiss, J. L. Measuring Nominal Scale Agreement among Many Raters. *Psychological Bulletin* **1971**, *76* (5), 378−382.

(40) McHugh, M. L. Interrater Reliability: The Kappa Statistic. *Biochem Med. (Zagreb)* **2012**, *22* (3), 276−282.

(41) Caspari, I.; Kranz, D.; Graulich, N. Resolving the Complexity of Organic Chemistry Students' Reasoning through the Lens of a Mechanistic Framework. *Chem. Educ. Res. Pract.* **2018**, *19* (4), 1117−1141.

(42) Bodé, N. E.; Deng, J. M.; Flynn, A. B. Getting Past the Rules and to the WHY: Causal Mechanistic Arguments When Judging the Plausibility of Organic Reaction Mechanisms. *J. Chem. Educ.* **2019**, *96* (6), 1068−1082.

(43) Goodwin, W. Explanation in Organic Chemistry. *Ann. N.Y. Acad. Sci.* **2003**, *988* (1), 141−153.

(44) Goodwin, W. M. Structural Formulas and Explanation in Organic Chemistry. *Found Chem.* **2008**, *10* (2), 117−127.

(45) *SPSS Statistics for Mac*; IBM Corp: Armonk, NY, 2017.

(46) Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates, 1988.

(47) Agresti, A. Inference for Two-Way Contingency Tables. In *Categorical Data Analysis*; Wiley: Hoboken, NJ, 2013.

(48) MacDonald, P. L.; Gardner, R. C. Type I Error Rate Comparisons of Post Hoc Procedures for I j Chi-Square Tables. *Educational and Psychological Measurement* **2000**, *60* (5), 735−754.

(49) Talanquer, V. Chemistry Education: Ten Heuristics To Tame. *J. Chem. Educ.* **2014**, *91* (8), 1091−1097.

(50) Bhattacharyya, G. Trials and Tribulations: Student Approaches and Difficulties with Proposing Mechanisms Using the Electron-Pushing Formalism. *Chem. Educ. Res. Pract.* **2014**, *15* (4), 594−609.

(51) Ferguson, R.; Bodner, G. M. Making Sense of the Arrow-Pushing Formalism among Chemistry Majors Enrolled in Organic Chemistry. *Chemistry Education Research and Practice* **2008**, *9*, 102−113.

(52) Anzovino, M. E.; Lowery Bretz, S. Organic Chemistry Students' Ideas about Nucleophiles and Electrophiles: The Role of Charges and Mechanisms. *Chem. Educ. Res. Pract.* **2015**, *16* (4), 797−810.

(53) Chinn, C. A.; Buckland, L. A.; Samarapungavan, A. Expanding the Dimensions of Epistemic Cognition: Arguments From Philosophy and Psychology. *Educational Psychologist* **2011**, *46* (3), 141−167.

(54) Lemke, J. L. *Talking Science: Language, Learning, and Values*; Ablex Publishing Corporation: Norwood, NJ, 1990.

(55) Tannen, D. *Framing in Discourse*; Oxford University Press: New York, 1993.

(56) Hammer, D.; Elby, A. On the Form of a Personal Epistemology. In *Personal Epistemology: The Psychology of Beliefs About Knowledge and Knowing*; Hofer, B. K., Pintrich, P. R., Eds.; Lawrence Erlbaum Associates: Mahwah, NJ, 2002; pp 169−190.

(57) Crandell, O. M.; Kouyoumdjian, H.; Underwood, S. M.; Cooper, M. M. Reasoning about Reactions in Organic Chemistry: Starting It in General Chemistry. *J. Chem. Educ.* **2019**, *96* (2), 213−226.

(58) Webber, D. M.; Flynn, A. B. How Are Students Solving Familiar and Unfamiliar Organic Chemistry Mechanism Questions in a New Curriculum? *J. Chem. Educ.* **2018**, *95* (9), 1451−1467.

Chemistry Education Research and Practice

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Chem. Educ. Res. Pract.*, 2023, **24**, 768

Beliefs *versus* resources: a tale of two models of epistemology

Kimberly S. DeGlopper, [a] Rosemary S. Russ, [b] Prayas K. Sutar [b] and Ryan L. Stowe *[a]

Received 27th October 2022,
Accepted 13th February 2023

DOI: 10.1039/d2rp00290f

rsc.li/cerp

Compelling evidence, from multiple levels of schooling, suggests that teachers' knowledge and beliefs about knowledge, knowing, and learning (*i.e.*, epistemologies) play a strong role in shaping their approaches to teaching and learning. Given the importance of epistemologies in science teaching, we as researchers must pay careful attention to how we model them in our work. That is, we must work to explicitly and cogently develop theoretical models of epistemology that account for the learning phenomena we observe in classrooms and other settings. Here, we use interpretation of instructor interview data to explore the constraints and affordances of two models of epistemology common in chemistry and science education scholarship: epistemological beliefs and epistemological resources. Epistemological beliefs are typically assumed to be stable across time and place and to lie somewhere on a continuum from "instructor-centered" (worse) to "student-centered" (better). By contrast, a resources model of epistemology contends that one's view on knowledge and knowing is compiled in-the-moment from small-grain units of cognition called *resources*. Thus, one's epistemology may change one moment to the next. Further, the resources model explicitly rejects the notion that there is one "best" epistemology, instead positing that different epistemologies are useful in different contexts. Using both epistemological models to infer instructors' epistemologies from dialogue about their approaches to teaching and learning, we demonstrate that how one models epistemology impacts the kind of analyses possible as well as reasonable implications for supporting instructor learning. Adoption of a beliefs model enables claims about which instructors have "better" or "worse" beliefs and suggests the value of interventions aimed at shifting toward "better" beliefs. By contrast, modeling epistemology as *in situ* activation of resources enables us to explain observed instability in instructors' views on knowing and learning, surface and describe potentially productive epistemological resources, and consider instructor learning as refining valuable intuition rather than "fixing" "wrong beliefs".

Introduction

It goes without saying that chemistry instructors at the undergraduate level have a great deal of knowledge about chemistry. The content they teach is rich and complex and requires nuanced understandings of an incredible array of concepts and phenomena (Boothe *et al.*, 2018; Zotos *et al.*, 2021). However, in addition to this knowledge of chemistry, instructors also have a great deal of knowledge and beliefs – albeit potentially tacit – about teaching and learning (Hora, 2014; Gibbons *et al.*, 2018; Popova *et al.*, 2020). For example, consider two different instructors' understandings of teaching and learning chemistry.

One of my most important roles as an instructor was to show people how the ideas interconnected… I should be doing something that goes, I guess, beyond just following the textbook because that's information they already can get. – Liam

The process of learning what a model is, what it applies to, and going through the practice of application of that model to explain an outcome and seeing that those things can be connected is the powerful thing we want our science students to do. – James

From these quotes, we might infer that Liam conceptualizes knowledge as consisting of many pieces of information that must be connected and that James sees learning as constructing, applying, and connecting models to explain phenomena. But what can these quotes tell us about their teaching?

Research in teaching and teacher education demonstrates that teacher thinking about teaching and learning has a substantial impact on teacher practice (*e.g.* Clark and Peterson, 1986; Pajares, 1992; Mansour, 2009; Baldwin and Orgill, 2019; Popova *et al.*, 2020). Teachers' implementations of curricular

---

[a] *Department of Chemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA. E-mail: rstowe@chem.wisc.edu*

[b] *Department of Education, University of Wisconsin–Madison, Madison, Wisconsin, 53706, USA*

reforms are influenced by beliefs about teaching and learning as are smaller day-to-day decisions like how much time to spend on a particular topic or their interaction with curricular materials (Cronin-Jones, 1991; Haney *et al.*, 1996; Wallace and Kang, 2004; Remillard, 2005; Roehrig *et al.*, 2007). The relationship between beliefs and practice is complex and its strength may vary depending on contextual factors (Fang, 1996). Nevertheless, if we wish to support chemistry instructors in improving their teaching practices, the literature suggests we should attend to instructor thinking.

In this work then, we examine and unpack existing research on instructors' knowledge about teaching and learning in chemistry. First, we recast that work in terms of what has been referred to elsewhere in the science education literature as epistemologies (Hofer and Pintrich, 1997; Lising and Elby, 2005; Sandoval, 2005; Smith and Wenk, 2006; Havdala and Ashkenazi, 2007; Oliveira *et al.*, 2012) or, more recently, epistemic cognitions (Greene *et al.*, 2016). Specifically, epistemologies "consist of [people's] systems of beliefs [tacit or explicit] about (1) the nature of knowledge and (2) the processes of knowing" (Hofer and Pintrich, 1997). Second, we compare and contrast two models of instructor thinking, particularly in regard to their underlying assumptions about the stability and hierarchy of beliefs. We then analyze our interview data according to each model and discuss affordances and limitations of each. Finally, we consider the implications of each model on instructor professional development.

Literature background

Education researchers have long sought to understand aspects of instructors' thinking that give rise to their teaching practice (Clark and Peterson, 1986; Shulman, 1986; Kagan, 1992; Schoenfeld, 1998; Abell, 2008). This approach to studying teaching practice is rooted in a cognitive paradigm that "conceptualizes teaching largely in terms of [teachers'] mental life and focuses on teaching as a way of thinking with a particular set of specialized knowledge and cognitive processes" (Russ *et al.*, 2016). Within this tradition, scholars have examined teacher's knowledge, beliefs, identities, and goals in an attempt to get "under the hood" of teacher practice (*e.g.*, Pajares, 1992; Abd-El-Khalick *et al.*, 1998; Lederman, 1999; Loughran *et al.*, 2004; Remillard, 2005; Avraamidou and Zembal-Saul, 2010; Orgill *et al.*, 2015; Connor and Shultz, 2018; Kradtap Hartwell, 2019; Lutter *et al.*, 2019; Posey *et al.*, 2019; Connor *et al.*, 2023). Further, scholars have similarly focused on instructors' attitudes, beliefs, and orientations toward teaching (*e.g.*, Mutambuki and Fynewever, 2012; Towns, 2015, 2016; Gibbons *et al.*, 2018; Popova *et al.*, 2020; Heidbrink and Weinrich, 2021; Vo *et al.*, 2022).

Of specific concern within science education has been the set of knowledge and beliefs that teachers possess that is associated with knowledge, knowing, and learning. For example, participants may view knowledge as constructed from things they already know or knowledge as transferred from authority. Further, they may view science learning as either an opportunity to make sense of phenomena or to memorize information. Although researchers use a range of constructs to conceptualize these knowledge and beliefs, here we follow work in science education that characterizes them as epistemologies (Hofer and Pintrich, 1997) or, more recently, epistemic cognitions (Greene *et al.*, 2016).

Tracing back to the 1970s, scholars have worked both to identify participants' epistemologies and also to tie those views to classroom practices of teaching and learning. Both correlational and case-study evidence suggests that epistemologies play an important role in school settings (Rosenberg *et al.*, 2006; Liang and Tsai, 2010; Greene *et al.*, 2018). A range of researchers across both K-12 and undergraduate settings have explored how instructors' tacit views of knowledge and knowing impact the ways they engage in teaching (Wendell *et al.*, 2019). For example, Russ and Luna (2013) followed a high school teacher across multiple class sessions to identify how her teaching practice shifted depending on whether she viewed teaching as an opportunity to Connect Biological Ideas or Use Procedural Knowledge. Similarly, Chari *et al.* (2019) analyzed 50 episodes of upper-division, undergraduate physics instruction to demonstrate how differing behavior of instructors was shaped by their two-dimensional epistemological understanding of problem-solving as being algorithmic/conceptual and mathematics/physics. Likewise, within chemistry education, researchers have probed the link between instructor thinking and practice. Gibbons *et al.* (2018) conducted a large scale study of chemistry instructors and found correlations between the instructors' beliefs about teaching and learning and reported pedagogical practices. Popova *et al.* (2020) focused specifically on assistant chemistry professors and similarly found some alignment between beliefs and practices.

These findings from across science education bear out the assumption that epistemology plays a strong role in shaping the teaching practices of instructors in science courses. As such, here we take as a given that epistemologies are an important piece of what lies "under the hood" in chemistry instructors' approaches to teaching and learning. Further, given the importance of epistemologies in science teaching, we as researchers must pay careful attention to how we model them in our work. That is, we must work to explicitly and cogently develop theoretical models of epistemology that account for the learning phenomena we observe in classrooms.

Theoretical framework

In our review of the literature, we identified two distinct approaches to modeling epistemology. In one approach, epistemologies are seen as "theories" that people consciously possess and apply in their lives (Hashweh, 1996; Hofer and Pintrich, 1997; Davis, 2003; Havdala and Ashkenazi, 2007; Kittleson, 2011). These are often referred to as "epistemological beliefs" (Hofer and Pintrich, 1997; Schommer-Aikins, 2004). The other approach views epistemology as constructed in-the-moment from "epistemological resources" – fine-grained knowledge elements concerning knowledge and the nature of knowing (Hammer and Elby, 2002). These models differ from each other in two key aspects: the extent to which epistemologies are assumed to be stable and whether or not epistemologies develop hierarchically over time. Here, we compare and contrast the two models by describing each model and its underlying assumptions. In doing so, our work on epistemology parallels prior scholarship comparing models of student conceptual learning as grounded in (mis)conceptions *versus* knowledge-in-pieces (*e.g.* Smith III *et al.*, 1994; Scherr, 2007).

A focus on beliefs

Modeling epistemologies as beliefs is common across science education literature and is especially prominent in chemistry education research. For example, Popova *et al.* (2020) interviewed assistant chemistry professors about their beliefs and checked in two years later to see how these beliefs changed (Popova *et al.*, 2021). Mack and Towns (2016) focused on physical chemistry instructors and interviewed them about their approach to teaching, which revealed beliefs about the purpose of their courses and the nature of knowledge in their discipline. Other studies have described instructors' beliefs in the context of specific topics, such as systems thinking (Szozda *et al.*, 2022) and grading (Mutambuki and Fynewever, 2012).

Although studies on instructor beliefs have uncovered a variety of beliefs regarding teaching and learning, many further classify their beliefs (and/or practices) as instructor-centered or student-centered (*e.g.*, Gibbons *et al.*, 2018; Popova *et al.*, 2020; Popova *et al.*, 2021). Instructor-centered beliefs are associated with a transmission view of learning and include beliefs that students learn chemistry most effectively by taking notes during lecture or doing homework problems. In contrast, believing that students learn chemistry most effectively by working in groups or making connections between chemistry and everyday life is considered student-centered and is associated with a constructivist view of learning. In their implications, the authors of these studies discussed ways to shift instructors' epistemological beliefs *and their practice* from instructor-centered to student-centered.

Modeling epistemologies as beliefs brings with it a set of common features, which include: (1) beliefs are stable and (2) beliefs develop hierarchically over time. These assumptions are rarely stated explicitly in the literature; rather, we infer their existence by examining the methods of data collection and analysis used (see below). In this paper, we aim to bring these assumptions to the forefront so that we can determine how they impact our understanding of instructor thinking.

Beliefs are stable. Chemistry instructor beliefs are often treated as stable over time. We can infer this feature from the methodologies – commonly longitudinal studies – used to study these beliefs. If beliefs are assumed to be unstable over the period of minutes or hours, we would expect to see studies looking at changes during this time scale. However, if beliefs are assumed to be stable over longer periods of time (*e.g.*, months or years), then it would be logical to collect data less frequently, perhaps once a semester or once a year. In the chemistry education literature, we mostly observe the latter. For example, Popova *et al.* (2021) conducted a study on assistant chemistry professors in which they compared participants' initial beliefs to their beliefs two years later, implying that changes were expected to occur on a longer time scale. Similarly, using a pre/post study design, in which beliefs are measured before and after an intervention, is reasonable if one assumes that the participants' beliefs would be essentially unchanged in the absence of the intervention for the duration of the study. Stains *et al.* (2015) have conducted such a study to measure the impact of a professional development program on assistant chemistry professors' beliefs. Conversely, we are not aware of any studies that characterize how chemistry instructors' thinking changes moment-to-moment.

Beliefs develop hierarchically over time. In the tradition of Piagetian stages of the 1960s (Piaget, 1969; Piaget, 1970) or the Expert-Novice studies of the 1980s (see Chi *et al.*, 1988), beliefs are often modeled as moving through a progression in which they become more sophisticated over long periods of time. For example, in order to develop a chemistry version of the Colorado Learning Attitudes about Science Survey (CLASS), originally developed for physics education research, Adams *et al.* (2006) interviewed non-major introductory chemistry students and chemistry faculty to establish the novice and expert responses, respectively, for survey items. This method makes sense if one expects differences in beliefs between these populations and similarities within each population. Furthermore, using their survey, the authors observed a "regression in beliefs" over a semester of general chemistry. The use of the term "regression" is consistent with a hierarchical, developmental model. Returning to the example of student-centered and instructor-centered beliefs, Popova *et al.* (2020) identified a cluster of beliefs they labeled "transitional and consistent," which contained a mixture of student-centered and instructor-centered beliefs. The label "transitional" implies an intermediate stage within a progression. While this continuum could be utilized in a purely descriptive manner, it has typically been presented in an evaluative manner. In their implication sections, the authors of these studies discuss ways to shift instructors from instructor-centered to student-centered, communicating that the latter is more desirable than the former.

A focus on epistemological resources

In contrast to the model of epistemological beliefs commonly used in the chemistry education literature, another model of epistemology contends that it is made up of a range of smaller units of cognition known as *resources* (Hammer, 2000). Below we detail the features of this model, presenting them in contrast to the features embedded in a beliefs model of epistemology.

Epistemological resources are unstable. Rather than understanding epistemologies as beliefs that are relatively stable across time and place, epistemological resources are taken to be unstable across contexts. As in the case with beliefs, this assumption shows up in the methods researchers use to study and document epistemologies. Specifically, researchers will use methods that allow them to capture rich data over relatively short time spans on the order of minutes. For example, in a case study of a group of 8th graders reasoning about the rock cycle, Rosenberg *et al.* (2006) use classroom video to demonstrate how students transition from one epistemology to another in a matter of moments based on a single comment from their teacher. Similarly, transitions in epistemologies that occur over minutes (rather than the hours, days, or years assumed in more stable models of cognition) have been documented in short excerpts (as few as 5–10 lines of transcript) in college physics classes (Scherr and Hammer, 2009; Irving *et al.*, 2013; Dini and Hammer, 2017; Modir *et al.*, 2017). The ''framework of epistemological resources, smaller and more general than theories or traits'' accommodates this dynamic contextual dependence (Hammer and Elby, 2002).

This unstable model of epistemology is rooted in a similar model of mind for conceptual understanding that may be more familiar to the reader (diSessa, 1993). Although science education began by comparing student thinking to scientific paradigms or robust scientific theories (McCloskey, 1983; Hewson and Hewson, 1984; Strike and Posner, 1985), a commitment to the notion of constructivism has demanded a move away from this (mis)conceptions model (Smith III *et al.*, 1994). Specifically, the field is now ''skeptical of treating knowledge or abilities as things one acquires and manipulates as intact units'' (Hammer *et al.*, 2005). Instead, we now think of conceptual knowledge as a complex system of many ''pieces'' (diSessa, 1993) which students unconsciously and dynamically assemble and disassemble in moments of thinking (Minstrell, 1989; Sherin, 2006; Philip, 2011). An epistemological resources model assumes the same is true for epistemology (Hammer, 2000; Hammer and Elby, 2002). Instead of people having ''precompiled'' (Hammer *et al.*, 2005) views of knowledge that they call up in learning situations, an epistemological resource model assumes people compile their view of knowledge dynamically in real time by drawing on many small epistemological elements.

Epistemological resources are differentially useful in different contexts. One of the key premises of a model of epistemological resources is that different situations call for different epistemologies (Elby and Hammer, 2001). For example, while the NGSS (NGSS Lead States, 2013) may encourage us to have students construct their own models for phenomena, we do not necessarily want the lay public to construct their own models for the spread of COVID (in fact the state of our public health may be drastically different if fewer people had done so!). In the former context (the classroom) we may want students to adopt a view that they can be the authority on knowledge, whereas in the sphere of COVID we want people to adopt a view that the scientific community is the authority. But even this grain size is not sufficient; it is not the case that the NGSS always wants students to believe they are the knowledge authority in classrooms. There are times in which we want students to adopt a view of learning where their teachers, or the textbook, are the authority – for example, when they are told a value like Avogadro's number.

Given the diversity and variability of epistemological resources that can be useful across the contexts of teaching and learning, researchers that adopt this model of epistemology explicitly reject a hierarchical model of progressive sophistication. Instead, this model assumes that there is no ''more correct'' or ''more expert'' epistemology but that instead epistemological resources are differentially productive for learning in context. Sophistication then is not merely adopting a set of expert views but is instead the ability to ''explore and discuss the differences between knowledge in multiple contexts'' (Elby and Hammer, 2001). In the case of teachers, epistemological expertise involves the ''awareness and judicious use of'' (Russ, 2018) a range of epistemological resources. Stated differently, epistemological sophistication means possessing a suite of epistemological resources as well as a finely tuned mechanism for identifying which contexts call for which resources.

Research questions

In the proceeding sections, we have described assumptions that underpin two common models for epistemology (epistemological beliefs and epistemological resources). Here, we take a look at what these models let us infer about chemistry instructors' epistemologies from dialogue about their approaches to teaching and learning. Specifically, we examine whether modeling instructors' epistemologies as resources supports different implications for instructor learning than modeling instructors' epistemologies as hierarchical, stable beliefs. The following research questions guided our efforts:

(1) What epistemologies do chemistry instructors articulate when talking about their approaches to teaching and learning in undergraduate organic chemistry?

(2) What are the affordances and limitations of modeling instructor thinking as beliefs and as epistemological resources?

Our purpose here is to show that the model of epistemology researchers chose powerfully influences the kind of analysis they conduct on their data and what they can infer about useful approaches to supporting instructor learning.

Methods

Our goal in this work is to examine the different tacit models of epistemology that exist in the literature to make sense of the kinds of claims each one can make about the nature of instructor thinking and learning. At its core then, we seek to contribute to the development of a cogent and well-specified framework (or theory) of instructor epistemologies; ours is a theoretical manuscript. In that way, our work parallels research by conceptual change scholars who have sought to understand the nature of student content knowledge (diSessa, 1993; Smith *et al.*, 1993; Scherr, 2007).

Given our parallel aims, we adopt empirical methods – particularly analytic methods – similar to those used by those scholars. Like them, we do not collect data from large-numbers of chemistry instructors and then summarize across it. Such an approach is more consistent with the goal of exhaustively mapping the terrain of knowledge elements used by participants (*i.e.*, Taber, 2010). Instead, we use a relatively small sample of instructors and carefully examine key moments in which their epistemologies are both active and inferrable. This small number of key moments allows us to investigate the features of epistemology we hypothesize in the previous sections. In doing so, we collect and use empirical data as a ''testing ground on which to refine [theoretical ideas]'' (diSessa, 117) about models of epistemology in chemistry. diSessa (1993) describes this as one of the key roles that data can play in theory development.

Further, our methodological approach provides data that lends intuitive plausibility (Russ, 2018) to our claims and arguments. It is our sincere hope that, once the theory of epistemology in chemistry is better specified, other scholars (or perhaps ourselves) will take up what we have begun and begin to map the terrain using large scale studies. But first we must decide the best way to carve up the terrain; that is our goal in this theoretical work.

Context and participants

This study focused on introductory organic chemistry instructors at a large public university in the Midwest. Although much of the chemistry education research focuses on general chemistry, here we choose to focus on organic chemistry for two reasons – one opportunistic and one substantive. First, many discussions were taking place in the department regarding changing and/or unifying the course. As a result, there was a pre-existing need to understand the goals instructors have for their students' knowledge construction and the means by which they believe these goals can be achieved. Second, and perhaps more importantly for our argument here, organic chemistry instructors have considerable autonomy in how they teach. Thus, we expected that more of their decisions would be based on their own epistemologies rather than institutional constraints (*e.g.*, ''I do this because my department says I have to''). This autonomy allows us to examine epistemologies more directly.

The introductory organic chemistry course at this university consists of two semesters (OChem I and OChem II). As this is a required course for chemistry, biology, and chemical engineering majors and anyone intending to pursue a career in the health field, it serves approximately 1,000 students each semester.

Organic chemistry instruction is divided among tenured professors, pre-tenure professors, and non-tenure track professors. The non-tenure track professors typically teach both OChem I and OChem II while most tenured and pre-tenure professors teach only one of these courses. All instructors use the same textbook and there is general agreement regarding the content that should be covered in each course, but each instructor has the freedom to choose their own teaching practices, author their own exams, and determine how points are allocated in their course. Some instructors have chosen to teach jointly with shared course materials and exams.

Interview requests were sent to everyone involved in teaching introductory organic chemistry over the last five years. We chose to restrict invitations to instructors who taught in the last five years because presumably these people would still remember details of how they approach(ed) teaching the course and would be involved in teaching it for several more years. Ten organic chemistry instructors responded. Informed consent was obtained from each respondent as approved by the University of Wisconsin – Madison Institutional Review Board. They included tenured professors, pre-tenure professors, and non-tenure track professors with teaching experience ranging from one year to approximately thirty years. Four of these instructors teach both OChem I and OChem II while the other six typically only teach one of these courses.

During data analysis, we utilized an intensity sampling approach (Creswell, 2007) to select ''information-rich cases that manifest [teacher beliefs] intensely but not extremely'' (p. 159). This approach allowed us to select a relatively small number of cases that provided in depth information for analysis; here we focus on three of the ten professors interviewed (Table 1). These instructors represent different roles within the department and exhibit a range of epistemological resources. James is a non-tenure track professor whose interview elicited fairly frequent and consistent epistemological ideas. Liam is a pre-tenure professor who demonstrated more inconsistency in his epistemic cognition. Mark is a tenured professor whose interview was most notable for the focus on logistical aspects of teaching rather than epistemological aspects.

Data collection

We chose to use interviews to infer instructor epistemologies. Interviews allowed the instructors to respond to the questions in their own words and in a more detailed manner than surveys typically allow. In recognition of the context-dependency of epistemic cognition, the interview questions were written to elicit reflections on the instructors' particular courses rather than general thoughts on teaching. Instructors could also supply context through the use of anecdotes and examples from their experiences. Additionally, the interview questions probed a range of teaching activities and contexts, from planning to assessment to student performance. However, such reflections are still filtered through the perceptions of the

Table 1 Relevant characteristics of instructors at the time they were interviewed

Instructor[a]	Position	Courses taught	Years of teaching experience
James	Non-tenure track	OChem I, OChem II	8
Liam	Pre-tenure	OChem I	1
Mark	Tenured	OChem II	15

[a] Actual names have been replaced with pseudonyms to protect identities.

interviewees; thus, they are not equivalent to direct observations of the instructors as they lecture or author assessments (Alshenqeeti, 2014). Ideally, the interviews would be coupled with observations of the instructors as they taught, wrote assessments, graded assessments, *etc.* In the future, we hope to collect this data. Nevertheless, we believe that interviews can help us figure out productive ways to model epistemology and can prompt instructors to consider multiple contexts for their teaching practices.

Semi-structured interviews were conducted over Zoom by the third author and lasted approximately one hour. The interviews began with questions regarding how the instructor got interested in chemistry and why they chose to stay in academia following graduate school (Q1 & Q2). Then the instructors were asked why students should take organic chemistry, what the students should learn from the course, and how the students can maximize their learning (Q3, Q4, Q8). The interview also included a discussion of assessment: how the instructors evaluate learning, what they aim to assess, and how they interpret assessment responses (Q6 & Q9). The interviews concluded with questions about if and/or how the instructors make use of teaching resources, including advice from peers and chemistry education research, and the role of evidence in changing teaching practices (Q11–Q13). The full interview protocol can be found in the appendix.

The interviews were transcribed by Zoom, and the first author corrected these transcriptions as needed to ensure they were accurate. The first author also broke up longer sections of dialogue into utterances that focused on a particular idea. These served as the units of analysis while coding.

Data analysis

Strand 1: analyzing instructor beliefs. In our first strand of analysis, we sought to understand instructors' epistemologies using a beliefs model. To do so we developed an analytic scheme by looking across the work of multiple authors who seek to characterize teacher beliefs from interviews or from their practice (Simmons *et al.*, 1999; Luft and Roehrig, 2007; Popova *et al.*, 2021). Although all of the authors identify a range of beliefs (*e.g.*, beliefs about student learning/actions, beliefs about the role/actions of teachers, beliefs about content, *etc.*), their analyses ultimately cluster teachers by patterns of responses. Further, although they each have several different clusters, ultimately the clusters are placed along a continuum where the two ends are student-centered and instructor-centered beliefs. Synthesizing across the papers we identified some common key elements of these two ends of the spectrum.

Student-centered. Instructors believe students learn by doing and not by listening; thus, the role of instructors involves collaborating with, facilitating, and guiding students as they construct ideas that are relevant to their lived experience from their prior knowledge.

Instructor-centered. Instructors believe that students learn by paying attention and listening to the instructor; thus, the role of the instructor is to provide content and experiences so they can assess if students know a set of pre-defined facts.

In addition to these two ends of the spectrum, researchers typically also included a transitional or inconsistent category when an instructor evidenced beliefs from both ends of the spectrum.

In our work here, we used the two ends as a guide for our analysis; the first two authors read through the transcripts and together assigned a code of "student-centered" or "instructor-centered" to each utterance. (Recall that an utterance was a section of dialogue concerning a single topic, typically 5–8 sentences.) We restricted our analysis to questions 3–10 of the interview protocol because these questions surfaced reflections on their own teaching rather than their perceptions of the department and the field of chemistry education. Furthermore, we ignored utterances that were not epistemic in nature (*e.g.*, "Say that one more time."). For this analysis, we relied heavily on which pronouns ("I" *versus* "they/them") were used in the active voice and which were used in the passive voice when referencing teaching and learning. If an utterance included both student-centered and instructor-centered beliefs, it was labeled as "both." Table 2 provides some examples from our data for each of the two clusters.

Strand 2: analyzing instructor resources

To describe the epistemological resources of chemistry instructors, we ground our analysis in the five-dimensional model proposed by Chinn and his colleagues (2011). We chose this model because it is, for us, the most comprehensive of all the existing models and is consistent with insights and components from other prominent scholars of epistemology (Hofer and Pintrich, 1997; Hammer and Elby, 2002; Schommer-Aikins, 2004). Below we will briefly describe each of the five dimensions.

Epistemic aims and values. Epistemic aims are the goals relating to inquiry, and epistemic values describe the relative worth of particular aims. Aims, or what others call goals (Berland *et al.*, 2016) are an important part of characterizing

Table 2 Examples of utterances coded as student-centered and instructor-centered

Code	Example
Student-centered	*What we need to be as educators, as teachers, is people who set up the students to have those experiences I just described. We need to be creating environments where somehow students are engaged in thinking about models, using models, writing about them to explain why something happens.*
Instructor-centered	*I felt like one of my most important roles as an instructor was to show people how the ideas interconnected. So whenever we introduce a new idea, be very clear about what is new in this idea... with kind of a very brief review of whatever that concept is. So I think that's something that is much harder to do when you're kind of working through, um, something kind of on your own.*

a person's epistemology because they are the ends to which other aspects of epistemic cognition are directed.

Structure of knowledge. The structure of knowledge refers to how knowledge is organized and answers questions like "What kind of answer should our [learning] provide?" (Berland *et al.*, 2016). They are akin to epistemic forms (Collins and Ferguson, 1993; Hammer and Elby, 2002) which are "target structures that guide inquiry."

Reliable and unreliable processes for achieving epistemic aims. Processes refer to the actions one takes to achieve one's epistemic aims. Epistemic processes are similar to Hammer and colleagues' (Hammer and Elby, 2002; Rosenberg *et al.*, 2006) epistemological activities that help people (tacitly!) answer the question, "What are you doing?" in terms of knowledge construction or use. Processes are also consistent with what researchers in undergraduate physics education (Tuminaro and Redish, 2007; Chen *et al.*, 2013; Odden and Russ, 2018) have called the moves in an epistemic game (Collins and Ferguson, 1993).

Sources, justifications, and stances. Sources of knowledge refers to where knowledge was obtained from, such as an expert, authority figure, textbook, or one's direct experience. Justifications for knowledge are the criteria by which a person evaluates knowledge, such as coherence with prior knowledge, logical consistency, or support with acceptable evidence. Stances toward knowledge describe a person's view on a given knowledge claim. Although Chinn *et al.* (2011) put these together because they are tightly linked in practice, other scholars treat these dimensions independently (Hammer and Elby; 2002; Tuminaro and Redish, 2007; Berland *et al.*, 2016).

Virtues and vices. Epistemic virtues and vices encompass personal characteristics that either support or hinder epistemic endeavors. Few other scholars in science education discuss this dimension.

Using these categories as a guide for our coding process, the authors read each utterance of dialogue and discussed whether they saw anything that would fall into the categories. Once this was completed for the three interviews, the authors summarized their observations for each category into a succinct list of resource codes. Once individual codes were defined, the authors used the constant comparison method (Glaser and Strauss, 1967) to confirm that all utterances were coded with a stable codebook (Table 3).

Results and discussion

In the section that follows, we consider the affordances and limitations of beliefs and resources models of epistemology in describing instructors' views on knowing and learning manifest during our interviews. We begin by briefly describing the instructor epistemologies elicited during the interviews, first using instructor-centered and student-centered descriptors. Then we will summarize the epistemological resources we observed using Chinn *et al.*'s (2011) multidimensional framework. Then we will consider the extent to which epistemologies embedded in interview dialogue were stable and the implications of treating these epistemologies as hierarchical.

RQ1: What epistemologies do chemistry instructors articulate when talking about their approaches to teaching and learning in undergraduate chemistry?

When we coded our instructors' beliefs as student-centered, instructor-centered, or both, we observed three qualitatively distinct profiles for our three instructors (Fig. 1). Approximately three quarters of Mark's beliefs were deemed instructor-centered while the remaining were student-centered. The reverse was observed for James; the vast majority of his beliefs were student-centered while a few were instructor-centered. Liam's beliefs were distributed almost equally among student-centered and instructor-centered. Therefore, if we were to adopt this model of describing instructor thinking, we would label Mark as instructor-centered, James as student-centered, and Liam as transitional.

When we coded for epistemological resources and organized them according to Chinn *et al.*'s multidimensional model, we identified several aims, reliable processes, sources, *etc.* (2011). These epistemological resources are summarized, along with examples from our data, in Table 3. The epistemic aims expressed by our instructors included memorization, explanations, and problem-solving, along with the value of usefulness. Reliable (or unreliable) processes for achieving these aims included forming (*i.e.*, constructing one's own knowledge based on prior knowledge), accumulating, and connecting. Connecting could be further described based on whether the instructor described how different topics relate (structural) or how causes give rise to effects (functional). These different ways of connecting knowledge were closely related to how the instructors discussed the structure of knowledge in their courses. They referenced "pieces" or "building blocks" of knowledge and articulated how making connections between them could result in more complex knowledge structures. Other times they described how the complexity could be reduced down to a few underlying pieces or fundamental ideas. Sources of knowledge referenced included the instructors themselves, the textbook, and data. We identified correctness as a commonly invoked justification for whether or not an aim had been achieved. Stances toward knowledge and virtues and vices were not observed in our dataset. Although we have described the epistemological resources observed amongst our instructors in aggregate, each of our instructors activated a variety of resources over the course of the interview. We will characterize each instructor's ideas in more detail when we explore the extent to which they were stable in the following section.

RQ2: What are the affordances and limitations of modeling instructor thinking as beliefs and as epistemological resources?

Throughout our interviews, instructors expressed epistemological ideas relating to course design, lecture practices, assessment strategies, and interactions with students. This enabled

Table 3 Epistemic resource codes and examples

Epistemic resource codes	Examples
Aims and values	
Memorization	"The students then have to memorize these factoids and memorize these patterns instead of understanding the model where they don't have to memorize anything."
Explanation	"I'd like to do a better job of assessing, um, is, um, actually getting some feedback myself about where they're deriving their explanations. So like when they say this would go through SN2, um, basically how can it be explained, um, like why, why did you say SN2, or what sort of factors do you think are at play here?"
Problem-solving	"So, uh, for somebody interested in, um, medicine, um, first of all, I guess like a large fraction of people taking the class, I think that, um, there are sort of aspects of the, the type of problem solving we do in organic chemistry that's really important. So, um, and sort of as specifically as I can, I guess what I feel like we're talking about is, uh, taking like a set of, uh, I guess, kind of starting criteria, like sort of the simple ideas, like steric bulk, um, electronic sort of perturbations, that have these principles and then trying to figure out how to sort of interconnect them to come up with an answer to a new sort of problem."
Usefulness	"Now I have no belief that most of my students will do a distillation again after they leave my class. And I do not care if they ever do a distillation again, that's irrelevant. But I know that 100% of my students are going to apply models to explain systems. They're going to use models to predict outcomes. They're gonna use models to rationalize outcomes. We should have them engaged in doing that."
Structure of knowledge	
Pieces	"I think the important things for us to be actually getting from [the students] are like connecting concepts and that's not connecting any concepts. Um, but I think hitting at some of the individual concepts on their own is also important. Um, so making sure that they're getting those building blocks and that we're not only assessing them on connecting the building blocks. I find that's also important."
Connections	"And so I felt like one of my most important roles as an instructor was to show people how the ideas interconnected. So whenever we introduce a new idea, be very clear about what is new in this idea and what is drawing on things they've already learned, with kind of a very brief review of whatever that concept is."
Hierarchy (building up)	"I think being able to connect independent concepts to address a more complex question, um, I think that's sort of a fundamental learning objective for organic chemistry."
Hierarchy (underlying)	"And so, especially for these pre-professional students who may never take another science class beyond second semester organic chemistry, um, this teaches them how you master a complicated topic that demands more than just rote memorization, right? This, it really does kind of, uh, teach you that, um, cramming isn't feasible at, um, you do have to understand underlying mechanisms to really succeed in a class like this."
Reliable processes	
Accumulating	"But one of the challenges to, um, doing formative assessment, in my view, is that because we put so much content in the class, I think it, I found it very difficult to adjust, to sort of respond to the students. Um, 'cause I would like to, if they're really struggling with the question, be able to dig in a little bit more, um, and sort of give that a little bit more time, and on some of the times that was okay and possible. Um, there certainly were other times where that wasn't going to be feasible because I had to get to the next sort of set of content."
Connecting (structural)	"But even for those [students] who are reading the book, I think that my job as an instructor is to, uh, put all of this information into a, into a package that's digestible so that they can see how the inferences get drawn, to see analogies from one unit to another one."
Connecting (functional)	"But I'm trying to assess, uh, whether [students] can predict reactivity or properties like acidity from molecular structure. Uh, and there, the sort of like a sub version of that, that's sort of predicting relative behavior of different structures, so being able to predict how two different mol-, how two different structure will result in two different activities."
Forming	"Um, and that, I think, there's sort of a trap in organic chemistry for those students because our content is, the learning objectives are about figuring out which of these principles to be thinking about and then thinking about them properly. And stuff can seem clear when you have the answer, where you really wouldn't be able to derive that answer yourself... learning the process of actually solving the problem is, I think, the most important thing for being successful"
Sources	
Instructor	"In practice, I think most of [the students] use the lecture as the main source of information, so, you know, as much as I would like for them all to be reading the book, I think that I am a primary source of information for them."
Textbook/online	So I didn't feel like my role was to define what the content was. And also there are very good resources; online textbooks are pretty good. There are lots of places [students] can get kind of that most basic information."
Data	"And so I try to, I try to convey the point that this class is such a conceptual one, such a theoretical one that that learning modality [of cramming] is going to fail, and I show [the students] data from previous years to show exactly why this fails."
Justifications	
Correctness	"And so then if they get [the question] right, or by and large get it right as a class, um, then I feel like I'm safe to move on [to the next topic]. If not, then that means I devote a little bit more time in the lecture to trying to clarify whatever that specific problem was."

us to explore instructor epistemologies across different topics over the span of the interview. For this sort of analysis, we focus on each instructor individually rather than looking across them.

Stability

We begin with James, who repeatedly espoused student-centered beliefs during his interview. James' primary goal for his students is that they engage with the practice of scientific modeling.

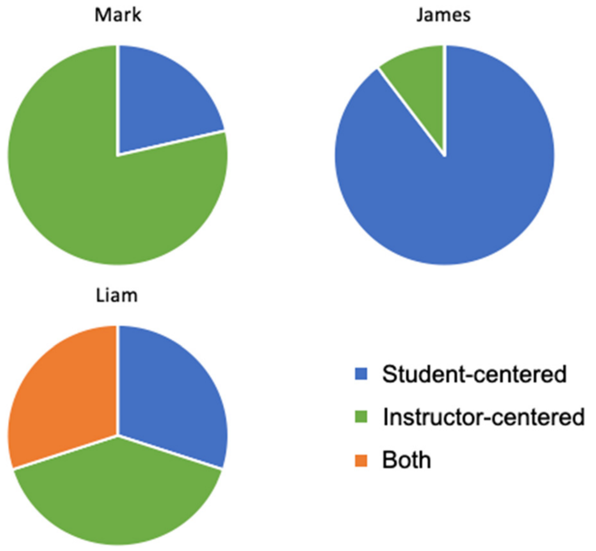


Fig. 1 Pie charts showing the proportions of student-centered and instructor-centered beliefs expressed by the instructors in this study.

So like the process of learning what a model is, what it applies to, and going through the practice of application of that model to explain an outcome and seeing that those things can be connected is the powerful thing we want our science students to do. Because what it finally does is it gets people thinking scientifically in a meaningful way, in that they understand, "Oh, people have seen data. People have generated models that explain those data. They have then tested those models and refined them over time. And here's the best understanding we have right now. Now, there might need to be some tweaks to that down the road, but this is the best understanding we have right now. And I can take that understanding and apply it to these cases and work out what's likely to happen. And I can then test that with spectroscopy. I can test that with some tool."

In this description of modeling, James positions his students as the constructors of knowledge – a hallmark of student-centered instruction. He wants his students to "*think scientifically*," "*to see that... things can be connected*," and "*take that understanding and apply it*."

When looking across James' interview, he repeatedly expresses this goal for his students' learning. When asked why students should take organic chemistry James' answer mirrors the one above.

The reason you should take organic chemistry...is that taking organic chemistry, if taught right, will help you understand that we can use some very simple, straightforward models that are accessible to students and to experts and they're the same model... We can use those same models to explain why chemical reactions happen. We can use those same simple models to rationalize why you get a particular regiochemistry or particular stereochemistry, why one product is major and one's minor, why one is seen and one is not observed in the data... That's incredibly empowering use of models to explain outcomes.

James centralizes modeling again when referencing his role as an instructor.

What we need to be as educators, as teachers, is people who set up the students to have those experiences I just described. We need to be creating environments where somehow students are engaged in thinking about models, using models, writing about them to explain why something happens.

Both a beliefs model and a resources model of epistemology work quite well for making sense of James' thinking. James' thinking about teaching appears to be consistently, and stably, student-centered since students are positioned as the modelers. Alternatively, we can state that the epistemic aim of modelling was repeatedly activated by James when reflecting on his course.

However, the other two instructors' interviews demonstrate more instability. Liam said he did not think it was his job to determine the course content and that students could obtain content from a variety of sources. In reflecting on his role as an instructor, he said,

We have, although we have some differences in what we teach across the different instructors, we teach a lot of the same reactions and basic principles. So I didn't feel like my role was to define what the content was. And also there are very good resources; online textbooks are pretty good. There are lots of places [the students] can get kind of that most basic information.

Rather, he described his job as follows: "So my role was one, I guess, make sure [the students] got some of the basic information, so sort of reviewing it a little bit, but more so than that, I thought my role was to show them how to connect concepts." His primary goal was not to deliver knowledge but to have students connect and use that knowledge. This goal would be considered student-centered. However, moments later in the interview, Liam shared the challenges he experienced with implementing clicker questions in lecture. He said,

But one of the challenges to doing formative assessment, in my view, is that because we put so much content in the class, I think it, I found it very difficult to adjust, to sort of respond to the students. 'Cause I would like to, if they're really struggling with the question, be able to dig in a little bit more and sort of give that a little bit more time, and on some of the times that was okay and possible. There certainly were other times where that wasn't going to be feasible because I had to get to the next sort of set of content.

Liam felt pressure to – and in fact states that he does – cover the content in lecture. That goal is part of an instructor-centered mindset.

As a result of these two different statements, the beliefs model of epistemology might place Liam into a "transitional" or "inconsistent" category of beliefs. Using the lens of a resources model, we would account for the instability by noting that Liam possesses epistemic resources for understanding both himself and outside resources as a sources of knowledge and that these were activated at different times.

Liam is not an anomaly in terms of this instability. Mark similarly demonstrates instability in his thinking about teaching and learning. For example, when talking about how students can maximize their learning in his course, Mark defaults to a practical, rather than knowledge-based, perspective.

And so I try to, I try to convey the point that this class is such a conceptual one, such a theoretical one that that [cramming] is *going to fail, and I show them data from previous years to show exactly why this fails... And I also try to tell them that, you know, that they should gear their, their studying around what the assessment is. And so if the assessment is, um, some kind of problem and a certain kinds of problems, then they should be doing those problems and those kinds of problems as part of their studying.*

In this quote, learning looks like ''time on task;'' his focus is on students engaging in activities (*e.g.*, assessments) that he has designed for them, a hallmark of instructor-centered teaching (Simmons *et al.*, 1999; Popova *et al.*, 2020; Luft and Roehrig, 2007). However, when asked why students should take organic chemistry, Mark articulates the following:

...the reason that I think everyone ought to take it is that it teaches you how to deal in a more sophisticated way with drawing influences, uh, inferences from data, uh, from using data to support an argument...And so if everybody took organic chemistry, then it would sort of help them to think about how you, um, how you, uh, use data to make informed decisions, which seems like a really important thing just in general.

He believes organic chemistry enables students to make decisions outside of the classroom by teaching them reliable processes characteristic of science such as using data. Further, he positions students as authorities who can interpret data, craft arguments, and make decisions. Making the class ''relevant'' for students who then engage in substantive intellectual work is a hallmark of student-centered instruction (Popova *et al.*, 2021). Like Liam, these two quotes of Mark's would lead beliefs researchers to characterize him as ''inconsistent'' or transitional.''

The resources model on the other hand expects this variability and treats it as something that can provide insight into Mark's teaching practice. For example, later in the interview Mark describes his approach to writing assessments which he summarizes as consisting of three general types of questions.

''The lowest level [type of question] is simply, you know, if you have some starting molecule, um, what reagents do you use to do some kind of a transformation?

... [The second type of question] is what I call circle-square, um, kinds of questions, which is circle the most acidic compound, square the least, uh, acidic compound, circle the most nucleophilic compound, square the least nucleophilic compound. Right? And so, so these sorts of questions are trying to get students to think through structure-reactivity principles, to get a sense of the character of the compounds.

...And then there are compound, there are, um, uh, questions that put everything together that has people to, uh, essentially explain an observable phenomenon, whether that is showing them a reaction and asking them to propose an arrow-pushing mechanism or giving them a phenomenon and asking them to explain why that phenomenon occurs or to rationalize the outcome. So it is very much along the lines of trying to model what a scientist does, right? If you are given an observable piece of data, how do you use theoretical constructs to rationalize that outcome?''

The first two types of questions simply ask students for a claim, whether it's providing the correct reagents or circling the correct molecule. In the last type of question, students are asked to do intellectual work of using data, drawing inferences, and making arguments. What we see here is again instability in his epistemologies; he has resources both for seeing the epistemic aim of knowing facts (*i.e.*, obtaining true beliefs) and the epistemic aim of having a rational model for how the world works. Rather than categorizing Mark as merely ''inconsistent,'' the resources model encourages us to explore the range and depth of his thinking and to view that range as potentially productive for his teaching practice (see below).

Hierarchy

Recall that if we examine the instructors' beliefs in aggregate, we see that James has mostly student-centered beliefs, Mark has mostly instructor-centered beliefs, and Liam has a mix of both. An implication of this analysis might be that James is the best instructor and that interventions are needed to shift Mark and Liam towards more student-centered beliefs.

Using an epistemological resources model of instructor thinking, we would come to a different conclusion. Since activation of epistemological resources is assumed to be context dependent, we would not treat individual resources as ''good'' or ''bad.'' Rather, we would recognize situations in which they might be more or less productive. For example, consider the following quote from Mark:

In practice, I think most of [the students] use the lecture as the main source of information, so, you know, as much as I would like for them all to be reading the book, I think that I am a primary source of information for them. But even for those who are reading the book, I think that my job as an instructor is to put all of this information into a package that's digestible so that they can see how the inferences get drawn, to see analogies from one unit to another one.

Since Mark positioned himself as the source of knowledge in the course, he would be described as instructor-centered and therefore less epistemologically sophisticated. Alternatively, we might notice that Mark possessed an epistemological resource for the instructor as a source of knowledge. Depending on the particular information Mark wanted to impart, this resource may be considered productive or unproductive. For example, creating space for students to ''figure out'' correspondence between features of spectroscopic traces and molecular structure may not be a good use of time. Organic chemists and organic chemistry learners need to be able to effectively analyze and interpret spectroscopic data (Stowe and Cooper, 2019), but they can do so by using skills and rules they are told (*e.g.*, the $n + 1$ rule). The goal of pulling information from spectra is to inform arguments about component(s) of a system under study. It would therefore be better to spend more class time considering consistency between possible claims and spectroscopic evidence rather than, for example, ''figuring out'' the $n + 1$ rule *via* numerous pattern matching exercises. From this and related examples, we can conclude that viewing the instructor as a source of knowledge is neither good nor bad but more or less appropriate depending on the particular circumstances.

An epistemological resource model also allows for a much more detailed characterization of instructor ideas, which enables us to recognize the variety of ideas each individual instructor holds, rather than reduce them to a single dimension. Even though James would overall be considered student-centered, some of his beliefs are more instructor-centered. For example, he states *"[The students] have to be explaining chemical phenomena using correct models, those models have to be based on core ideas, it all has to tie together, they have to be able to do that on course-wide assessments."* The standard of justification conveyed here is correctness, which based on our knowledge of his course, means agreement with scientific canon (*i.e.*, authority). A student-centered approach to justifying models might be consistency with data as judged by the classroom community. By labeling James as student-centered, we might not recognize the aspects of his teaching that could still be improved.

On the other hand, consider Mark, who expressed mostly instructor-centered beliefs. A closer examination reveals some student-centered beliefs. For example, when he articulated how he thinks organic chemistry aids pre-med students, he said:

> *And so, especially for these pre-professional students who may never take another science class beyond second semester organic chemistry, this teaches them how you master a complicated topic that demands more than just rote memorization, right? This, it really does kind of teach you that cramming isn't feasible, you do have to understand underlying mechanisms to really succeed in a class like this. And I think that's important, especially for the people who are going on to these higher education where they are going to have to start learning things like medicine, where, you know, simply memorizing a list of, you know, characteristics of a disease is much less important than understanding the underlying mechanism. So it really is the same kind of thought process.*

In this response, Mark stressed the importance of understanding rather than simply memorizing information. A resources model allows us to attend to these ideas.

The example of Liam arguably provides the most interesting case. Recall that Liam exhibited a mix of student-centered and instructor-centered beliefs. One method of analysis might be to place him in a "transitional" category. But treating the variability as noise ignores the interesting tensions Liam himself identified and prevents us from gaining insight into how we could support his teaching. For example, consider this quote from Liam.

> *I guess something I do a bad job, I think, of assessing, but I'd like to do a better job of assessing is actually getting some feedback myself about where they're deriving their explanations. So like when they say this would go through SN2, basically how can it be explained, like why did you say SN2, or what sort of factors do you think are at play here? Again, I worry about grading burden.*

Because he framed assessment improvement as a feedback tool for himself as the instructor and noted the implication in terms of the grading burden for himself and his TAs, we coded this as instructor-centered. But if we look more closely, we can infer that Liam was not satisfied with the epistemic aim of correct answers for his students. Rather, he wanted to know that his students understood the "why." Liam's desire to improve assessment practices to gain more insight into students' thinking and extend justifications beyond simply correct claims is an excellent starting point for improving his teaching. In this case, the barriers are not epistemic but are instead logistical. A supportive approach for Liam would be to reinforce the aim of understanding for students and provide additional graders to help him assess understanding.

Implications and conclusion

In this study, we have examined how two different models of epistemology lend themselves to different sorts of analysis of data on instructor thinking about learning in chemistry. Our findings suggest that each model has different constraints and affordances. We conclude by exploring those differences and their implications for professional development and research in chemistry education.

Beliefs model of epistemology

Limitations. Our analysis above indicates a number of limitations of the beliefs model. First, the beliefs model does not account for the impact of context on instructor thinking. As our examples and literature in science education (*e.g.* Coffey *et al.*, 2009; Russ and Luna, 2013) both show, the way that instructors think about learning in one context (for example when thinking about their assessments) can differ dramatically from how they think about learning in other contexts (for example when thinking about their course plans). By collapsing across time and context in analyses of instructor beliefs, this beliefs model loses this variation. While this may not be problematic in all cases (there are some instructors with stable beliefs), there are cases even in their own data (*e.g.*, "instructor-centered and inconsistent beliefs" category in Popova, 2020) in which instructors express contradictory beliefs. In such cases, collapsing their thinking into a single belief may obscure important portions of their thinking. We can make the analogy to the use of descriptive statistics in data analysis. A beliefs model is akin to using an average – and only the average – to define a data set that may itself be fairly disparate. An average can be insufficient at best (as in the case of a distribution with long tails) and misleading at worst (as in the case of a bimodal distribution). In either case, the beliefs model misses out on portions of instructor thinking that may have relevance to their practice.

A second limitation of the beliefs model is the assumed link between those beliefs and practice. The research on beliefs uses decontextualized surveys or interviews (Luft and Roehrig, 2007; Gibbons *et al.*, 2018; Popova *et al.*, 2020) to elicit and analyze instructor espoused beliefs (Sandoval, 2003). In doing so, this model assumes a relationship between these decontextualized beliefs and how those beliefs will be enacted in practice. However, there is little empirical data from the literature to support such a connection between espoused beliefs and practice. As Hora (2014) articulates, "Given the lack of evidence regarding the causal relations among faculty beliefs, teaching, and student outcomes, faculty developers would be well served to not focus solely on faculty beliefs but instead to adopt a more comprehensive view of teacher growth and development'' (p. 64–65).

Finally, if we assume instructor beliefs are more-or-less stable across time and place and fall on a continuum from ''worse'' (*i.e.*, instructor-centered beliefs) to ''better'' (*i.e.*, student-centered beliefs), a focus on characterizing and improving ''bad'' beliefs makes a great deal of sense. Indeed, it is common for scholars who assume beliefs are stable to categorize these beliefs *via* self-report surveys (Gibbons *et al.*, 2018), concept maps (Fletcher and Luft, 2011; Lee, 2019), or interviews (Luft and Roehrig, 2007; Popova *et al.*, 2020), and propose interventions meant to support shifts toward ''better'' beliefs (Fletcher and Luft, 2011; Mattheis and Jensen, 2014; Moore *et al.*, 2015; Pelch and McConnell, 2016; Czajika and McConnell, 2019; Lee, 2019). Unfortunately, this approach potentially positions instructors as having ''wrong'' beliefs which require ''fixing'' and largely ignores potentially productive, if nascent, ways instructors have for thinking about teaching and learning. Taking such an evaluative approach can have unwanted implications for supporting instructors. Specifically, treating their thinking as ''wrong'' and in need of fixing can elicit defensive behavior from instructors with whom we work, making them unreceptive to our suggestions. Thus a final limitation is that a beliefs model lends itself to a deficit model of instructors.

Affordances. However, there are multiple affordances of the beliefs model. First, it has intuitive appeal in that it is consistent with how we talk about thinking in the everyday world. No one would argue with the notion that instructors hold beliefs about learning, and designing a survey or interview protocol asking instructors what they believe is relatively straightforward. Data analysis can be similarly straightforward since it does not require much inference to interpret statements like ''I believe students should learn by reading the textbook.'' The clear and straightforward methodological implications of using a beliefs model means that it is more accessible to researchers or practitioners who want to study their local context.

In addition, a beliefs model of epistemology lends itself to large-scale studies. If beliefs are assumed to be relatively stable, surveys can be used to collect data, making it feasible to obtain large samples sizes and conduct statistical analyses. For example, Gibbons *et al.* (2018) obtained over a thousand responses to their survey probing chemistry instructors beliefs and practices and conducted a factor analysis to distinguish types of instructional styles from the responses. Furthermore, large scale studies like these seem to be particularly persuasive to administrators who might be more familiar with how to interpret quantitative rather than qualitative results.

A beliefs model also aligns with studies that utilize a pre/post design. If beliefs are assumed to be stable over time in the absence of any intervention, then any changes in beliefs can be tentatively attributed to the intervention rather than the dynamic nature of epistemology. Stains *et al.* (2015) conducted such a study to examine the impact of a professional development workshop on new faculty members. They concluded that their workshop was successful in shifting faculty beliefs from instructor-centered to student-centered. Like the large scale studies described above, pre/post studies in which there is a clear link between intervention and outcome may also be persuasive to outside stakeholders.

Resources model of epistemology

Despite these affordances of the beliefs model, we argue for adopting a resources model that shifts away from the standard, tacit epistemology model used in chemistry education in which teachers slowly develop ''better'' (*i.e.*, more student-centered) beliefs. Below we articulate the limitations and constraints of the resources model with an eye toward highlighting its generative aspects for both professional development and research.

Limitations. First, modeling epistemology as activation of resources is conceptually complex. Foundational work upon which the resources model is based suggests epistemologies are often tacit, variable, and context-dependent (*e.g.* Hammer and Elby, 2001; Russ and Luna, 2013). Thus, scholars employing a resources model have to consider what inferences about epistemology can be reasonably made from behavior, how variability should be attended to, and what features of context have the potential to send consequential messages about knowledge and knowing.

Second, analyzing the epistemological resources activated in- and across-moments is time consuming work. Researchers interested in such analyses need to carefully consider how dimensions of epistemology might be manifest in what study participants do and say. This sort of work is highly inferential. For example, some of the dimensions in Chinn and colleague's model, such as *structure of knowledge*, are almost always implicit in speech about other topics (*e.g.*, ''it is important for students to connect what they learned in the last chapter to problems they do in this unit''). Calibrating what sort of inferences about epistemology may be reasonably made from messy data takes time, training, and an understanding of learning theory – this is not an analysis that can readily be done by non-researchers. Consequently, resource analyses are low-throughput relative to analyses that rely on multiple-choice surveys. It is not practical to conduct such analyses with hundreds of instructors.

Finally, there is no ''best epistemology'' according to the resources model. Instead, different epistemologies have differential utility depending on the goals of the learner. Accordingly, implications from resources analyses are highly context bound and not intended to be generalized. We would not say, for example, that Mark's epistemologies are worse than James, so he needs to take workshop B to improve. Nuanced implications about, for example, the features of context that relate to participants' in-the-moment epistemologies do not lend themselves to quick summaries that can be readily digested by policymakers and administrators.

Affordances. The first affordance of the resources model deals with professional development. When we view instructor epistemic cognition as in-the-moment compilation of small-grain ''pieces'' related to knowledge and knowing, it becomes clear that no ''piece'' is inherently ''right'' or ''wrong'' (diSessa, 1993; Smith III *et al.*, 1994; Hammer, 1996; Hammer, 2000; Hammer *et al.*, 2005). Instead, clusters of epistemological resources may be more or less productive in progressing toward certain knowledge construction aims in a given moment; the resources model considers the context when assigning value to an idea. By attending to the context, we can avoid labeling instructors' beliefs as good or bad and perpetuating a deficit view of instructors. A focus on instructors' epistemological resources allows us to shift toward surfacing potentially productive resources and connections and creating contexts that signal the utility of desirable in-the-moment epistemologies. Stated differently, a resources perspective allows us to approach instructor learning as constructivists (Schafer *et al.*, 2022). In doing so, we place instructors firmly in the role of having some expertise and ways of thinking that contribute to new ways of teaching, putting them in a position of power rather than a position of defensiveness. Our analysis of Liam and Mark in particular points to the ''nuggets'' of productive epistemologies that we could draw on in professional development to support their learning to teach.

Further, we know that, in principle, instructors possess productive epistemological resources for doing science that they could bring to the classroom. The instructors we interviewed are all practicing scientists with years of experience constructing, revising and communicating evidence-based causal accounts for phenomena they care about. While our interviews suggest that instructors activate some epistemological resources for doing science in the context of teaching, they do not activate others. For example, in a research setting, a model of how and why a reaction occurs is typically evaluated by consistency with experimental data, but in the classroom, such models are typically evaluated by alignment with expert models and deemed ''correct'' or ''incorrect.'' We hypothesize that supporting instructors in adopting *doing science* epistemologies in school contexts could lead to enactment of more authentic, meaningful chemistry learning environments. Thus, as a potential first step toward supporting epistemologies for science (Russ, 2014) in the classroom, we advocate for providing opportunities for instructors to reflect on how they approach science and how they approach teaching. Importantly, the goal should be to make use of their experiences as scientists rather than ''fix'' their teaching.

The second affordance of the resources model involves its implications for research. A host of intriguing questions come into focus when one adopts a resources perspective on epistemic cognition, including: what leads instructors to compile their epistemologies in a certain way in a given moment? How can we influence those resources instructors (tacitly) see as productive? How does activation of certain epistemologies influence instructor decisions about curriculum and assessment? Seeking answers to these and related questions will allow us to understand mechanisms by which instructor epistemologies evolve, which will in turn support approaches to instructor epistemological learning that surface and build on productive resources. Such approaches would focus on helping instructors identify which epistemological resources are productive in which contexts.

Understanding, and ultimately influencing, instructor epistemologies in-the-moment and across moments is non-trivial. Epistemologies arise from a dynamic and complex system of interactions between people and materials inside and outside of the classroom. Furthermore, epistemologies cannot be understood in terms of discrete levels a person progresses through but rather as in-the-moment confluences of epistemological resources pertaining to aims for knowledge use, processes for achieving aims, sources of knowledge, and justifications for evaluating knowledge. We think the analysis described in this paper is a useful means of characterizing these resources, and the interviews we conducted surfaced some of the specific resources that might be observed. However, more research is needed to understand how instructor epistemologies arise and how they influence the design of course materials and evaluation of student knowledge products. Once we generate a working model of the relationship between instructor epistemologies and the actions they take in the context of teaching, we can study strategies for productively ''tipping'' instructors toward activating epistemological resources that have the potential to support students in engaging with science for the purpose of making sense of phenomena.

Limitations of the study

We conceive of this work as the beginning of an investigation into instructor epistemologies, and there is still much we plan to explore. The data we analyzed were collected through interviews and therefore are filtered through the perceptions of the instructors. We do not know the extent to which the epistemologies elicited through our interview protocol align with the epistemologies which shape in-the-moment instructional or assessment decisions. We would need to observe instructor's behavior as they talk to students or grade exams in order to infer these epistemologies. Furthermore, we do not claim that the epistemologies we have identified are representative of chemistry instructors everywhere nor do we claim to have uncovered all the epistemologies for teaching and learning that the interviewed instructors possess. Rather, we offer this analysis as an illustration of how one might elicit and characterize instructor ideas.

Author contributions

Conceptualization: KSD, RSR, PS, RLS. Data curation: KSD, PS. Formal analysis: KSD, RSR. Investigation: PS. Supervision: RSR, RLS. Visualization: KSD. Writing – original draft: KSD, RSR. Writing – review & editing: KSD, RSR, PS, RLS.

Conflicts of interest

There are no conflicts to declare.

Appendix: interview protocol

One-on-one interviews with faculty and staff are intended to elicit evidence as what interviewees view as the aims of organic chemistry learning environments, interviewee perspectives on how those environments are best designed, and how they view the role of the learner. Evidence collected from these interviews will support inferences about how the beliefs of the faculty and staff influence their approach to the curricular deliberations.

The interviews will be conducted in person or *via* video conference using WebEx and the audio would be recorded either *via* the use of a recorder or *via* using the record function of WebEx. Informed consent will be obtained prior to collection of any interview data.

The interview will begin with a period in which the teacher is asked to reflect on their experience as a chemist, an educator and as a student. Following this initial reflection, the interviewer will ask the questions including those written below. As this is intended as a semi-structured interview, additional questions may be added on the spot in response to ideas brought up by the interviewee. Following the period of initial reflection, the interviewer will ask the following questions, following up where needed to elicit examples salient to each question.

1. Why should students take organic chemistry as undergraduates?

2. What are the most important things that students should learn in an undergraduate organic chemistry course?

3. How do you describe your role as a teacher?

Description: What are the things you focus on as a teacher in order to achieve these aims?

4. How do you know when your students understand?

Description: How do you find out if your students are learning the things are the focus of the course?

5. How do you decide what works and what does not work for you as a teacher?

Description: How do decide if you need to change the course content, assessments or your instruction? What kind of feedback do you look for?

6. What should students do to maximize their learning during the course?

7. What are the things that you aim to test students on while designing the assessments and why?

8. Do you think the experience of your students in your course is similar or different from your experience as an undergraduate student? Why do you think so?

9. What kind of resources do you refer to inform your teaching? Do you refer to chemistry education research to inform your teaching? Why/Why not?

10. Do you think chemistry faculty reflect on and revise their practice in response to evidence? What kind of evidence convinces them that change is needed/beneficial?

Description: As scientists chemistry researchers must modify their beliefs and assumption if thy find evidence to the contrary. When it comes to teaching, what kind of evidence convinces the faculty that they need to modify their teaching practices? How is the nature of this evidence similar or different from the evidence in chemistry research?

11. If a chemistry education researcher were to approach you to modify your teaching practices, what kind of evidence would he need to produce in order to convince you?

Acknowledgements

The authors wish to thank the organic instructors who generously shared their time and expertise with us. This study was supported by the U.S. National Science Foundation under Grant No. DUE-2225025. Funding was also provided by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin – Madison with funding from the Wisconsin Alumni Research Foundation.

References

Abd-El-Khalick F., Bell R. L. and Lederman N. G., (1998), The nature of science and instructional practice: making the unnatural natural, *Sci. Educ.*, **82**(4), 417–436, DOI: **10.1002/(SICI)1098-237X(199807)82:4<417::AID-SCE1>3.0.CO;2-E**.

Abell S. K., (2008), Twenty years later: does pedagogical content knowledge remain a useful idea? *Int. J. Sci. Educ.*, **30**(10), 1405–1416, DOI: **10.1080/09500690802187041**.

Adams W. K., Perkins K. K., Podolefsky N. S., Dubson M., Finkelstein N. D. and Wieman C. E., (2006), New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey, *Phys. Rev. Special Topics – Phys. Educ. Res.*, **2**(1), 010101, DOI: **10.1103/PhysRevSTPER.2.010101**.

Alshenqeeti H., (2014), Interviewing as a data collection method: a critical review, *English Linguistics Res.*, **3**(1), 39–45, DOI: **10.5430/elr.v3nlp39**.

Avraamidou L. and Zembal-Saul C., (2010), In search of well-started beginning science teachers: insights from two first-year elementary teachers, *J. Res. Sci. Teach.*, **47**(6), 661–686, DOI: **10.1002/tea.20359**.

Baldwin N. and Orgill M., (2019), Relationship between teaching assistants' perceptions of student learning challenges and their use of external representations when teaching acid–base titrations in introductory chemistry laboratory courses, *Chem. Educ. Res. Pract.*, **20**(4), 821–836, DOI: **10.1039/C9RP00013E**.

Berland L. K., Schwarz C. V., Krist C., Kenyon L., Lo A. S. and Reiser B. J., (2016), Epistemologies in practice: making scientific practices meaningful for students, *J. Res. Sci. Teach.*, **53**(7), 1082–1112, DOI: **10.1002/tea.21257**.

Boothe J. R., Barnard R. A., Peterson L. J. and Coppola B. P., (2018), The relationship between subject matter knowledge and teaching effectiveness of undergraduate chemistry peer facilitators, *Chem. Educ. Res. Pract.*, **19**(1), 276–304, DOI: **10.1039/C7RP00171A**.

Chari D. N., Nguyen H. D., Zollman D. A. and Sayre E. C., (2019), Student and instructor framing in upper-division physics, *Am. J. Phys.*, **87**(11), 875–884, DOI: **10.1119/1.5120392**.

Chen Y., Irving P. W. and Sayre E. C., (2013), Epistemic game for answer making in learning about hydrostatics, *Phys. Rev. Special Topics – Phys. Educ. Res.*, **9**(1), 010108, DOI: **10.1103/PhysRevSTPER.9.010108**.

Chi M. T. H., Glaser R. and Farr M. J., (1988), *The nature of expertise*, Psychology Press, DOI: **10.4324/9781315799681**.

Chinn C. A., Buckland L. A. and Samarapungavan A., (2011), Expanding the dimensions of epistemic cognition: arguments from philosophy and psychology, *Educational Psychologist*, **46**(3), 141–167, DOI: **10.1080/00461520.2011.587722**.

Clark C. M. and Peterson P. L., (1986), Teachers' thought processes, in M. C. Wittrock (ed.), *Handbook on research in teaching*, Macmillan Publishing Co, pp. 255–296.

Collins A. and Ferguson W., (1993), Epistemic forms and epistemic games: structures and strategies to guide inquiry, *Educ. Psychol.*, **28**(1), 25–42, DOI: **10.1207/s15326985ep2801_3**.

Connor M. C. and Shultz G. V., (2018), Teaching assistants' topic-specific pedagogical content knowledge in 1H NMR spectroscopy, *Chem. Educ. Res. Pract.*, **19**(3), 653–669, DOI: **10.1039/C7RP00204A**.

Connor M. C., Rocabado G. A. and Raker J. R., (2023), Revisiting faculty members' goals for the undergraduate chemistry laboratory, *Chem. Educ. Res. Pract.*, **24**, 217–233, DOI: **10.1039/D2RP00215A**.

Creswell J. W., (2007), *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*, SAGE Publications.

Cronin-Jones L. L., (1991), Science teacher beliefs and their influence on curriculum implementation: two case studies, *J. Res. Sci. Teach.*, **28**(3), 235–250, DOI: **10.1002/tea.3660280305**.

Czajika C. D. and McConnell D., (2019), The adoption of student-centered teaching materials as a professional development experience for college faculty, *Int. J. Sci. Educ.*, **41**(5), 693–711, DOI: **10.1080/09500693.2019.1578908**.

Davis E., (2003), Untangling dimensions of middle school students' beliefs about scientific knowledge and science learning, *Int. J. Sci. Educ.*, **25**(4), 439–468, DOI: **10.1080/09500690210145765**.

Dini V. and Hammer D., (2017), Case study of a successful learner's epistemological framings of quantum mechanics, *Phys. Rev. Phys. Educ. Res.*, **13**(1), 010124, DOI: **10.1103/PhysRevPhysEducRes.13.010124**.

diSessa A. A., (1993), Toward an epistemology of physics, *Cognition Instruct.*, **10**(2–3), 105–225, DOI: **10.1080/07370008.1985.9649008**.

Elby A. and Hammer D., (2001), On the substance of a sophisticated epistemology, *Sci. Educ.*, **85**(5), 554–567, DOI: **10.1002/sce.1023**.

Fang Z., (1996), A review of research on teacher beliefs and practices, *Educ. Res.*, **38**(1), 47–64, DOI: **10.1080/0013188960380104**.

Fletcher S. S. and Luft J. A., (2011), Early career secondary science teachers: a longitudinal study of beliefs in relation to field experiences, *Sci. Educ.*, **95**(6), 1124–1146, DOI: **10.1002/sce.20450**.

Gibbons R. E., Villafañe S. M., Stains M., Murphy K. L. and Raker J. R., (2018), Beliefs about learning and enacted instructional practices: an investigation in postsecondary chemistry education, *J. Res. Sci. Teach.*, **55**(8), 1111–1133, DOI: **10.1002/tea.21444**.

Glaser B. and StraussA., (1967), *The Discovery of Grounded Theory*, Aldine Publishing Company.

Greene J. A., Sandoval W. A. and Bråten I., (2016), An introduction to epistemic cognition, in J. A. Greene, W. A. Sandoval and I. Bråten (ed.), *Handbook of Epistemic Cognition*, Routledge, pp. 1–15, DOI: **10.4324/9781315795225**.

Greene J. A., Cartiff B. M. and Duke R. F., (2018), A meta-analytic review of the relationship between epistemic cognition and academic achievement, *J. Educ. Psychol.*, **110**(8), 1084–1111, DOI: **10.1037/edu0000263**.

Hammer, D., (1996), Misconceptions or p-prims: how may alternative perspectives of cognitive structure influence instructional perceptions and intentions, *J. Learn. Sci.*, **5**(2), 97–127, DOI: **10.1207/s15327809jls0302_1**.

Hammer D., (2000), Student resources for learning introductory physics, *Am. J. Phys.*, **68**(S1), S52–S59, DOI: **10.1119/1.19520**.

Hammer D. and Elby A., (2002), On the form of a personal epistemology, in B. K. Hofer and P. R. Pintrich (ed.), *Personal epistemology: the psychology of beliefs about knowledge and knowing*, Lawrence Erlbaum Associates Publishers, pp. 169–190.

Hammer D., Elby A., Scherr R. E. and Redish E. F., (2005), Resources, framing, and transfer, in J. P. Mestre (ed.), *Transfer of Learning from a Modern Multidisciplinary Perspective*, Information Age Publishing, pp. 89–119.

Haney J. J., Czerniak C. M. and Lumpe A. T., (1996), Teacher beliefs and intentions regarding the implementation of science education reform strands, *J. Res. Sci. Teach.*, **33**(9), 971–993, DOI: **10.1002/(SICI)1098-2736(199611)33:9<971::AID-TEA2>3.0.CI;2-S**.

Hashweh M. Z., (1996), Effects of science teachers' epistemological beliefs in teaching, *J. Res. Sci. Teach.*, **33**(1), 47–63, DOI: **10.1002/(SICI)1098-2736(199601)33:1<47::AID-TEA3>3.0.CO;2-P**.

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, DOI: **10.1002/tea.20215**.

Heidbrink A. and Weinrich M., (2021), Undergraduate chemistry instructors' perspectives on their students' metacognitive development, *Chem. Educ. Res. Pract.*, **22**(1), 182–198, DOI: **10.1039/D0RP00136H**.

Hewson P. W. and Hewson M. G. A., (1984), The role of conceptual conflict in conceptual change and the design of science instruction, *Instruct. Sci.*, **13**(1), 1–13, DOI: **10.1007/BF00051837**.

Hofer B. K. and Pintrich P. R., (1997), The development of epistemological theories: beliefs about knowledge and knowing and their relation to learning, *Rev. Educ. Res.*, **67**(1), 88–140, DOI: **10.3102/00346543067001088**.

Hora M. T., (2014), Exploring faculty beliefs about student learning and their role in instructional decision-making, *Rev. Higher Educ.*, **18**(1), 37–70, DOI: **10.1353/rhe.2014.0047**.

Irving P. W., Martinuk M. S. and Sayre E. C., (2013), Transitions in students' epistemic framing along two axes, *Phys. Rev. Special Topics – Phys. Educ. Res.*, **9**(1), 010111, DOI: **10.1103/PhysRevSTPER.9.010111**.

Kagan D. M., (1992), Implication of research on teacher belief, *Educ. Psychol.*, **27**(1), 65–90, DOI: **10.1207/s15326985ep2701_6**.

Kittleson J. M., (2011), Epistemological beliefs of third-grade students in an investigation-rich classroom, *Sci. Educ.*, **95**(6), 1026–1048, DOI: **10.1002/sce.20457**.

Kradtap Hartwell S., (2019), Changing instructor attitudes and behaviors to support student learning and retention, in *Enhancing retention in introductory chemistry courses: teaching practices and assessments*, American Chemical Society, **vol. 1330**, pp. 169–186, DOI: **10.1021/bk-2019-1330.ch011**.

Lederman N. G., (1999), Teachers' understanding of the nature of science and classroom practice: factors that facilitate or impede the relationship, *J. Res. Sci. Teach.*, **36**(8), 916–929, DOI: **10.1002/(SICI)1098-2736(199910)36:8＜916::AID-TEA2＞3.0.CO;2-A**.

Lee S. W., (2019) The impact of a pedagogy course on the teaching beliefs of inexperienced graduate teaching assistants, *CBE—Life Sci. Educ.*, **18**(1), 18:ar5, 1–18:ar5, 12, DOI: **10.1187/cbe.18-07-0137**.

Liang J. and Tsai C., (2010), Relational analysis of college science-major students' epistemological beliefs toward science and conceptions of learning science, *Int. J. Sci. Educ.*, **32**(17), 2273–2289, DOI: **10.1080/09500690903397796**.

Lising L. and Elby A., (2005), The impact of epistemology on learning: a case study from introductory physics, *Am. J. Phys.*, **73**(4), 372, DOI: **10.1119/1.1848115**.

Loughran J., Mulhall P. and Berry A., (2004), In search of pedagogical content knowledge in science: developing ways of articulating and documenting professional practice, *J. Res. Sci. Teach.*, **41**(4), 370–391, DOI: **10.1002/tea.20007**.

Luft J. A. and Roehrig G. H., (2007), Capturing science teachers' epistemological beliefs: the development of the teacher beliefs interview, *Electron. J. Res. Sci. Math. Educ.*, **11**(2), 38–63.

Lutter J. C., Hale L. V. A. and Shultz G. V., (2019), Unpacking graduate students' knowledge for teaching solution chemistry concepts, *Chem. Educ. Res. Pract.*, **20**(1), 258–269, DOI: **10.1039/C8RP00205C**.

Mack M. R. and Towns M. H., (2016), Faculty beliefs about the purposes for teaching undergraduate physical chemistry courses, *Chem. Educ. Res. Pract.*, **17**(1), 80–99, DOI: **10.1039/C5RP00148J**.

Mansour N., (2009), Science teachers' beliefs and practices: issues, implications and research agenda, *Int. J. Environ. Sci. Educ.*, **4**(1), 25–48.

Mattheis A. and Jensen M., (2014), Fostering improved anatomy and physiology instructor pedagogy, *Adv. Physiol. Educ.*, **38**(4), 321–329, DOI: **10.1152/advan.00061.2014**.

McCloskey M., (1983), Naive theories of motion, in D. Gentner and A. L. Stevens (ed.), *Mental Models*, Lawrence Erlbaum Associates, Inc, pp. 299–323.

Minstrell J., (1989), Teaching science for understanding, in L. B. O. Resnick and L. Klopfer (ed.), *Toward the thinking curriculum: current cognitive research*, Association of Super-vision and Curriculum Development, pp. 133–149.

Modir B., Thompson J. D. and Sayre E. C., (2017), Students' epistemological framing in quantum mechanics problem solving, *Phys. Rev. Phys. Educ. Res.*, **13**(2), 020108, DOI: **10.1103/PhysRevPhysEducRes.13.020108**.

Moore T. J., Guzey S. S., Roehrig G. H., Stohlmann M., Park M. S., Kim Y. R., Callender H. L. and Teo H. J., (2015), Changes in faculty members' instructional beliefs while implementing model-eliciting activities, *J. Eng. Educ.*, **104**(3), 279–302, DOI: **10.1002/jee.20081**.

Mutambuki J. and Fynewever H., (2012), Comparing chemistry faculty beliefs about grading with grading practices, *J. Chem. Educ.*, **89**(3), 326–334, DOI: **10.1021/ed1000284**.

NGSS Lead States, (2013), *Next generation science standards: for states, by states*, The National Academies Press.

Odden T. O. B. and Russ R. S., (2018), Sensemaking epistemic game: a model of student sensemaking processes in introductory physics, *Phys. Rev. Phys. Educ. Res.*, **14**(2), 020122, DOI: **10.1103/PhysRevPhysEducRes.14.020122**.

Oliveira A. W., Akerson V. L., Colak H., Pongsanon K. and Genel A., (2012), The implicit communication of nature of science and epistemology during inquiry discussion, *Sci. Educ.*, **96**(4), 652–684, DOI: **10.1002/sce.21005**.

Orgill M., Bussey T. J. and Bodner G. M., (2015), Biochemistry instructors' perceptions of analogies and their classroom use, *Chem. Educ. Res. Pract.*, **16**(4), 731–746, DOI: **10.1039/C4RP00256C**.

Pajares M. F., (1992), Teachers' beliefs and educational research: cleaning up a messy construct, *Rev. Educ. Res.*, **62**(3), 307–332, DOI: **10.3102/00346543062003307**.

Pelch M. A. and McConnell D. A., (2016), Challenging instructors to change: a mixed methods investigation on the effects of material development on the pedagogical beliefs of geoscience instructors, *Int. J. STEM Educ.*, **3**(5), 1–18, DOI: **10.1186/s40594-016-0039-y**.

Philip T. M., (2011), An ''ideology in pieces'' approach to studying change in teachers' sensemaking about race, racism, and racial justice, *Cognition Instruct.*, **29**(3), 297–329, DOI: **10.1080/07370008.2011.583369**.

Piaget J., (1969), *Psychology of intelligence*, Littlefield Adams.

Piaget J., (1970), Piaget's theory, in P. Mussen (ed.), *Carmichael's handbook of child psychology*, Wiley, pp. 703–732.

Popova M., Shi L., Harshman J., Kraft A. and Stains M., (2020), Untangling a complex relationship: teaching beliefs and instructional practices of assistant chemistry faculty at research-intensive institutions, *Chem. Educ. Res. Pract.*, **21**(2), 513–527, DOI: **10.1039/C9RP00217K**.

Popova M., Kraft A., Harshman J. and Stains M., (2021), Changes in teaching beliefs of early-career chemistry faculty: a longitudinal investigation, *Chem. Educ. Res. Pract.*, **22**(2), 431–442, DOI: **10.1039/D0RP00313A**.

Posey L. A., Bieda K. N., Mosley P. L., Fessler C. J. and Kuechle V. A. B., (2019), Mathematical knowledge for teaching in chemistry: identifying opportunities to advance instruction, in *It's just math: Research on students' understanding of* *chemistry and mathematics*, American Chemical Society, **vol. 1316**, pp. 135–155, DOI: 10.1021/bk-2019-1316.ch009.

Remillard J. T., (2005), Examining key concepts in research on teachers' use of mathematics curricula, *Rev. Educ. Res.*, **75**(2), 211–246, DOI: 10.3102/00346543075002211.

Roehrig G. H., Kruse R. A. and Kern A., (2007), Teacher and school characteristics and their influence on curriculum implementation, *J. Res. Sci. Teach.*, **44**(7), 883–907., DOI: 10.1002/tea.20180.

Rosenberg S., Hammer D. and Phelan J., (2006), Multiple epistemological coherences in an eighth-grade discussion of the rock cycle, *J. Learn. Sci.*, **15**(2), 261–292, DOI: 10.1207/s15327809jls1502_4.

Russ R. S., (2014), Epistemology of science vs. epistemology for science, *Sci. Educ.*, **98**(3), 388–396, DOI: 10.1002/sce.21106.

Russ R. S., (2018), Characterizing teacher attention to student thinking: a role for epistemological messages, *J. Res. Sci. Teach.*, **55**(1), 94–120, DOI: 10.1002/tea.21414.

Russ R. S. and Luna M. J., (2013), Inferring teacher epistemological framing from local patterns in teacher noticing, *J. Res. Sci. Teach.*, **50**(3), 284–314, DOI: 10.1002/tea.21063.

Russ R. S., Sherin B. L. and Sherin M. G., (2016), What constitutes teacher learning? in D. H. Gitomer and C. A. Bell (ed.), *Handbook of Research on Teaching*, 5th edn, American Educational Research Association, pp. 391–438.

Sandoval W. A., (2003), The inquiry paradox: Why doing science doesn't necessarily change ideas about science, in *CBLIS conference proceedings 2003: New technologies and their applications in education*, Department of Educational Sciences, University of Cyprus, vol. I, pp. 825–834. http://gnosis.library.ucy.ac.cy/handle/7/64579.

Sandoval W. A., (2005), Understanding students' practical epistemologies and their influence on learning through inquiry, *Sci. Educ.*, **89**, 634–656, DOI: 10.1002/sce.20065.

Schafer A. G. L., Kuborn T. M., Schwarz C. E., Deshaye M. Y. and Stowe R. L. (2023), Messages about valued knowledge products and processes embedded within a suite of transformed high school chemistry curricular materials, *Chem. Educ. Res. Pract.*, **24**, 71–78, DOI: 10.1039/D2RP00124A.

Scherr R. E., (2007), Modeling student thinking: an example from special relativity, *Am. J. Phys.*, **75**(3), 272–280, DOI: 10.1119/1.2410013.

Scherr R. E. and Hammer D., (2009), Student behavior and epistemological framing: examples from collaborative active-learning activities in physics, *Cognition Instruct.*, **27**(2), 147–174, DOI: 10.1080/07370000902797379.

Schoenfeld A. H., (1998), Toward a theory of teaching-in-context, *Issues Educ.*, **4**(1), 1–94, DOI: 10.1016/S1080-9724(99)80076-7.

Schommer-Aikins M., (2004), Explaining the epistemological belief system: introducing the embedded systemic model and coordinated research approach, *Educ. Psychol.*, **39**(1), 19–29, DOI: 10.1207/s15326985ep3901_3.

Sherin B., (2006), Common sense clarified: the role of intuitive knowledge in physics problem solving, *J. Res. Sci. Teach.*, **43**(6), 535–555, DOI: 10.1002/tea.20136.

Shulman L. S., (1986), Those who understand: knowledge growth in teaching, *Educ. Res.*, **15**(2), 4–14, DOI: 10.3102/0013189X015002004.

Simmons P. E., Emory A., Carter T., Coker T., Finnegan B., Crockett D., Richardson L., Yager R., Craven J., Tillotson J., Brunkhorst H., Twiest M., Hossain K., Gallagher J., Duggan-Haas D., Parker J., Cajas F., Alshannag Q., McGlamery S. and Labuda K., (1999), Beginning teachers: beliefs and classroom actions, *J. Res. Sci. Teach.*, **36**(8), 930–954, DOI: 10.1002/(SICI)1098-2736(199910)36:8<930::AID-TEA3>3.0.CO;2-N.

Smith C. L. and Wenk L., (2006), Relations among three aspects of first-year college students' epistemologies of science, *J. Res. Sci. Teach.*, **43**(8), 747–785, DOI: 10.1002/tea.20113.

Smith III J. P., diSessa A. A. and Roschelle J., (1994), Misconceptions reconceived: a constructivist analysis of knowledge in transition, *J. Learn. Sci.*, **3**(2), 115–163, DOI: 10.1207/s15327809jls0302_1.

Stains M., Pilarz M. and Chakraverty D., (2015), Short and long-term impacts of the Cottrell Scholars Collaborative New Faculty Workshop, *J. Chem. Educ.*, **92**(9), 1466–1476, DOI: 10.1021/acs.jchemed.5b00324.

Stowe R. L. and Cooper M. M., (2019), Arguing from spectroscopic evidence, *J. Chem. Educ.*, **96**(10), 2072–2085, DOI: 10.1021/acs.jchemed.9b00550.

Strike K. A. and Posner G. J., (1985), A conceptual change view of learning and understanding, in L. H. T. West and A. L. Pines (ed.), *Cognitive structure and conceptual change*, Academic, pp. 211–231.

Szozda A. R., Bruyere K., Lee H., Mahaffy P. G. and Flynn A. B., (2022), Investigating educators' perspectives toward systems thinking in chemistry education from international contexts, *J. Chem. Educ.*, **99**(7), 2474–2483, DOI: 10.1021/acs.jchemed.2c00138.

Tuminaro J. and Redish E. F., (2007), Elements of a cognitive model of physics problem solving: epistemic games, *Phys. Rev. Special Topics – Phys. Educ. Res.*, **3**(2), 020101, DOI: 10.1103/PhysRevSTPER.3.020101.

Vo K., Sarkar M., White P. J. and Yuriev E., (2022), Problem solving in chemistry supported by metacognitive scaffolding: teaching associates' perspectives and practices, *Chem. Educ. Res. Pract.*, **23**(2), 436–451, DOI: 10.1039/D1RP00242B.

Wallace C. S. and Kang N., (2004), An investigation of experienced secondary science teachers' beliefs about inquiry: an examination of competing belief sets, *J. Res. Sci. Teach.*, **41**, 936–960, DOI: 10.1002/tea.20032.

Wendell K. B., Swenson J. E. S. and Dalvi T. S., (2019), Epistemological framing and novice elementary teachers' approaches to learning and teaching engineering design, *J. Res. Sci. Teach.*, **56**(7), 956–982, DOI: 10.1002/tea.21541.

Zotos E. K., Tyo J. J. and Shultz G. V., (2021), University instructors' knowledge for teaching organic chemistry mechanisms, *Chem. Educ. Res. Pract.*, **22**(3), 715–732, DOI: 10.1039/D0RP00300J.

JOURNAL OF CHEMICAL EDUCATION

Editors' Choice

pubs.acs.org/jchemeduc

Article

Authenticity-Driven Design of a High-Enrollment Organic Laboratory Course

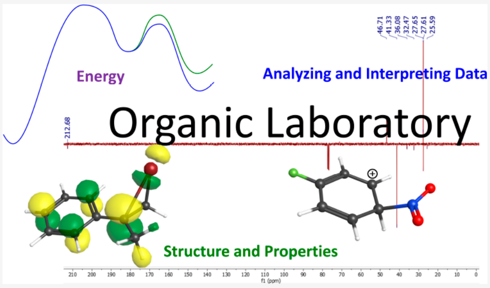
Brian J. Esselman,* Nicholas J. Hill,* Kimberly S. DeGlopper, Aubrey J. Ellison, Ryan L. Stowe,* Cara E. Schwarz, and Niall J. Ellias

Cite This: *J. Chem. Educ.* 2023, 100, 4674−4685

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We have developed a curriculum for the organic chemistry laboratory in which students draw on authentic usage of spectroscopy, spectrometry, and computational chemistry to explain chemical phenomena. This curriculum, which has been continuously refined over a decade, has been explored by many thousands of students and is scalable to small and large institutions. Herein, we articulate our design philosophy of engaging students in explaining chemical phenomena using authentic data and describe how this philosophy informs our curricular choices. We present an overview of the entire curricular system that includes well-scaffolded activities to support student learning throughout course activities, including laboratory assessments and written exams. To assess the extent to which our course materials align with our goals, we analyzed those materials through the lens of three-dimensional learning. Our laboratory assessments and written exams are highly 3D (∼20−30%) and emphasize the science and engineering practice of analyzing and interpreting data (∼50−60%). We demonstrate that it is possible to have a highly 3D laboratory curriculum that supports students in a high enrollment course (>500 students/term). Future work will explore how students experience these 3D tasks that require them to analyze and use multiple sources of data to construct explanations for chemical phenomena.



KEYWORDS: *Organic Chemistry, Second-Year Undergraduate, Curriculum, Assessment, Testing, Spectroscopy, Computational Chemistry*

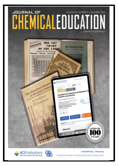
OUR VISION FOR LABORATORY CURRICULUM DESIGN

Authenticity is the theme of our philosophical approach to curricular design in laboratory courses.[1] As we design and refine our course materials, we continually ask ourselves, "How would a practicing organic chemist gain deeper understanding of these phenomena? How would a practicing organic chemist studying these chemical phenomena think and what would they do? Which analyses would be meaningful to other chemists? How can we engage students in a similar fashion?" The answers to these questions guide our core pedagogical choices, a few of which are described here:

(1) Students should perform reactions to generate a defined target molecule from an available procedure. The procedure should include the necessary reagents and amounts, conditions, and purification processes.[2]
(2) Students should employ as many modern instrumental techniques (IR, NMR, GC-MS) as possible, practical, and beneficial for analysis of each reaction. When feasible and enriching, both the crude reaction mixture and the purified product should be analyzed via NMR and/or GC-MS techniques. Students should always receive data for the samples they generate to promote a connection between their own experiences in the laboratory and the spectroscopic outcome.[3]
(3) Students should engage in rationalizing the observed outcome of a chemical reaction using the most appropriate model. Students should use a modern understanding of structure and reactivity, ideally supported by computational chemistry, to articulate why the observed reaction occurs.[4]

An organic chemistry laboratory course designed according to these pedagogical choices is a rich environment in which students can apply conceptual models to authentic data to make sense of chemical phenomena. Students explore complex systems, apply relevant models, and argue from evidence to reach chemically sound conclusions that allow them to make sense of the systems studied. Students focus on engaging in scientific practices to make sense of chemical phenomena using their data, rather than focusing on acquiring good laboratory technique. These practices and their underlying reasoning are authentic to all scientists and should be reflected in students'

Received: August 21, 2023
Revised: October 30, 2023
Accepted: November 2, 2023
Published: November 15, 2023



https://doi.org/10.1021/acs.jchemed.3c00845

experience in undergraduate laboratory courses that adopt this philosophy.

Practicing chemists rationalize chemical phenomena and communicate their results and insights via group meetings, conference presentations, and journal articles. Unfortunately, rationalizing chemical phenomena has not traditionally been the assessment focus of many organic laboratory curricula. Rather, the focus has often been on students learning how to successfully perform a reaction or use an apparatus. While important, the techniques used to generate the data are generally not the primary concern of synthetic organic (or other subspecialty) chemists.[5] Understanding the purpose, benefits, limitations, and functionality of any technique is often more important than the physical implementation of the technique itself. Likewise, few students will need to use organic-specific techniques in their future endeavors. Many, however, will need to understand how and why techniques are used and to engage in analyzing data to understand a relevant chemical reaction. Thus, authentic organic laboratory curricula and assessments should emphasize rationalizing phenomena and de-emphasize the acquisition of skills related to manipulative techniques and/or memorization of facts. Additionally, laboratory assessments should use appropriate experimental data to focus on selected phenomena. Assessments should not be an avenue to ask general questions about loosely related reactions for which data are not provided to students. We strongly believe that assessments (exams and postlaboratory analyses) clearly communicate to students the desired learning objectives, thus we have explicitly focused our assessments on how students use authentic spectral and computational data to explain chemical phenomena. Stated plainly, we believe, "You are what you assess."[6]

THEORETICAL FRAMEWORK FOR CURRICULUM DEVELOPMENT IN THE ORGANIC LABORATORY

As stated within our philosophy, organic chemistry learning environments, and laboratory courses in particular, should engage students in making sense of chemical phenomena through use of authentic data and the most appropriate models. Though not explicit in our minds as we began our journey toward a more authentic organic laboratory curriculum, our philosophical choices are well-aligned with priorities expressed by the National Research Council (NRC) in *The Framework for K-12 Science Education*.[7,8] There is broad consensus that K-12 learning environments should engage students in making sense of the natural world and that this practice can position them as *knowers and doers* of science.[9] There have been many similar calls to focus science education on students *doing science*, rather than emphasizing the acquisition of facts and skills.[10,11] Members of the chemistry and science education community have conceptualized *doing science* as using a collection of practices (e.g., argumentation from evidence, developing and using models) for the purpose of determining how and why phenomena happen.[8,12−15] As part of doing science, students should describe, connect, explain, and predict phenomena and systems of interest.[16] Our vision of an authentic organic laboratory course is well-aligned with the community's belief that students should engage in practices and reasoning similar to those used by professional scientists.

These scientific practices have been described in *The Framework for K-12 Science Education*.[8] Though intended for K-12 science education, this document has also informed a variety of college STEM learning environment transformations, including general and organic chemistry.[17−21] These efforts have integrated *science and engineering practices* (SEPs)[22] with the *disciplinary core ideas* (DCIs)[23] of chemistry and the *crosscutting concepts* (CCCs)[24] that focus attention on specific aspects of phenomena. Collectively, these have been used to construct "three-dimensional" (3D) performance expectations which have guided construction, assessment, and refinement of learning environments. Our approach to authenticity in the laboratory curriculum, much like these other efforts, can be described in the language of these three dimensions of science learning (SEPs, DCIs, and CCCs).

The extent to which a curriculum, classroom environment, or assessment provides opportunities for students to engage in the type of learning described by the *Framework* can be quantified by analyzing their 3D-learning emphasis. In order for an activity or assessment to meet the criteria for 3D learning in this context, students or instructors must be engaged in using SEPs, DCIs, and CCCs.[22−24] Taken together, these three dimensions of learning provide a framework for instructors to design experiences in which students make sense of natural phenomena. The three-dimensional learning assessment protocol (3D-LAP) has been used extensively by Cooper and co-workers[20−29] to evaluate whether an assessment item has the potential to elicit evidence of 3D learning in student responses. The curricula of only two organic chemistry lecture courses have been analyzed for their emphasis and use of 3D learning.[20,30] There are many individual laboratory exercises and inquiry-focused courses (e.g., REActivities,[31] problem-based learning,[32,33] CUREs,[34−37] etc.) that may accomplish some aspect of 3D learning in the chemistry education literature, but to our knowledge this report is the first to detail an organic laboratory curriculum intentionally designed to engage students in 3D learning across an entire course serving a large number of students.

LABORATORY CONTEXT

Institution and Curricular Setting

The laboratory course described in this work (CHEM 344) is offered at the University of Wisconsin−Madison as a standalone, two-credit course associated with the content of the Organic Chemistry I and II lecture courses.[30,38] The CHEM 344 course is designed, administered, and assessed separately from the lecture courses. At UW−Madison, a single organic chemistry sequence serves the overwhelming majority of students in chemistry and related STEM majors. All courses in this sequence use Organic Chemistry 7th Edition by Loudon and Parise[39] as the textbook, and most follow the book's conceptual progression. The Organic Chemistry I and II lecture courses enroll approximately 2000 and 1600 students each calendar year, respectively. Students may enroll in the laboratory course concurrently with or subsequent to completion of Organic Chemistry II lecture. Annually, approximately 1200 students complete CHEM 344, which is offered in the fall, spring, and summer terms. The course is taught by up to three principal instructors and up to 38 teaching assistants (TAs, typically graduate students). Each TA is assigned to a laboratory section of 18 students. The typical TA workload for the course (discussion, laboratory, grading, office hours, weekly staff meeting, etc.) is typically less than 16 h per week. The laboratory operation is assisted by stockroom

Table 1. CHEM 344 Laboratory Experiments[a]

phase	experiment	spectroscopy and spectrometry	computational chemistry[b]	% SEP	% 3D[c]
1	Computational Chemistry[40,41]	^{1}H, ^{13}C, HSQC, IR	opt, IRC, NMR, NBO	73	55
2	Separation of a Mixture by Acid−Base Extraction	^{1}H, ^{13}C, HSQC, IR, GC-MS	opt, NMR	41	0
	Nucleophilic Substitution Reactions (S_N1/S_N2)[42,43]	^{1}H, ^{13}C, HSQC, IR, GC-MS	opt, NMR, NBO	60/44[d]	30/38[d]
	Elimination Reactions (E1/E2)[42]	^{1}H, ^{13}C, HSQC, GC-MS	opt, NMR, NBO	70	28
	Oxidation of 4-*tert*-Butylcyclohexanol[44−46]	^{1}H, ^{13}C, HSQC, IR, GC-MS	opt, NMR, NBO	48	17
	Bromohydrin Reaction of Alpha-Methylstyrene[47,48]	^{1}H, ^{13}C, HSQC, GC-MS	opt, NBO	68	32
3	Reactions of Nitrogen Functional Groups[49]	^{1}H, ^{13}C, HSQC, IR, GC-MS	opt, NMR, NBO	72	43
	Wittig Synthesis of Ethyl Cinnamate[50]	^{1}H, IR, GC-MS	opt, NMR	59	0
	Wittig & Diels−Alder Reaction[51,52]	^{1}H, IR, GC-MS	opt, NMR	56	16
	Stereochemistry of a Carbonyl Reduction[53]	^{1}H, IR, GC-MS	opt, NMR	55	32
	Electrophilic Aromatic Substitution - Nitration[54]	^{1}H, ^{19}F, GC-MS	opt, NMR, NBO	67	35
	Electrophilic Aromatic Substitution - Acylation[55,56]	^{1}H, ^{13}C, HSQC, IR, GC-MS	opt, NMR	55	34
	Grignard Reaction[57,58]	^{1}H, ^{13}C, ^{19}F, HSQC, IR, GC-MS	opt, NMR, NBO	67/42[d]	39/20[d]
	Suzuki-Miyaura Coupling[59]	^{1}H, ^{13}C, IR	opt, NBO	67	44
	Cu/TEMPO Catalyzed Aerobic Oxidation of a Primary Alcohol[60e]	^{1}H, GC-MS	opt, NBO	70	32
	Biginelli Reaction[61]	^{1}H, ^{13}C, HSQC, IR	opt, NMR, NBO	74	38
	Aldol Condensation[62]	^{1}H, ^{13}C, HSQC, IR, GC-MS	opt, NMR, NBO	67	27
	S_NAr Reaction of a Fluorinated Aromatic Compound[63]	^{1}H, ^{13}C, ^{19}F, HSQC, IR, GC-MS	opt, NMR, NBO	64	40
	Radical Bromination of Bibenzyl[64]	^{1}H, ^{13}C, HSQC, IR, GC-MS	opt, NMR, NBO	60	60

[a] Not all listed experiments are included each semester. [b] The details of the a geometry optimization (opt), intrinsic reaction coordinate (IRC), NMR, and natural bond orbital (NBO) calculations are available in the computational chemistry experiment.[40,41] [c] Analysis of the credit associated with three-dimensional learning using 3D-LAP. [d] At least two substantially different versions of this experiment have been implemented. See SI for details. [e] Cu/TEMPO Oxidation has been implemented in phases 2 and 3.

technicians and staff who maintain the departmental NMR and computational chemistry facilities.

CHEM 344 Course Structure

The three broadly targeted phases in CHEM 344 are (1) spectroscopy, spectrometry, and computational chemistry, (2) experiments associated with concepts from Organic I lecture, and (3) experiments associated with Organic II lecture. Each phase is punctuated by a cumulative written assessment. Phase 1 is dedicated to students gaining expertise in analysis of increasingly complex NMR, GC-MS, and IR data and learning how to use computational software to rationalize chemical phenomena. All of the acquired skills are critical for the analyses utilized continually throughout the remainder of the course. As shown in Table 1, this phase includes a single laboratory experiment on computational chemistry.[40,41] Students engage with experiments that correspond to content typically found in either the Organic I or Organic II lecture courses during phases 2 and 3, respectively. As mentioned above, students may co-enroll in Organic II lecture and the laboratory course; thus, the content directly related to Organic II appears in phase 3 to ensure students have encountered relevant material in their lecture courses. Typically, students meet twice weekly with a TA for a 4 h session that includes both a classroom discussion and a laboratory session during the fall and spring terms. The daily schedule is modified slightly during the summer term to accommodate shorter class periods.

Prior to attending the TA-led classroom discussion session, students submit prelab exercises designed to inform their forthcoming experiences in the laboratory and assist their postlab analysis. The prelab assignments intentionally provide scaffolding for the computational or spectroscopic analysis required in the postlab analysis; see Supporting Information (SI) for more details. As mentioned above, these items are supplemented by a 20−40 min TA-led prelaboratory discussion that focuses on the background of the chemical reactions, the answers to the prelab exercises, the purpose and setup of the apparatus, and safety reminders regarding the experiment. The combination of prelab exercises and TA support is sufficient for students to enter the laboratory well-prepared.

During the experiment, TAs and laboratory directors assist students as they carry out the written procedure. Most students obtain usable data from the experiments listed in Table 1 during the laboratory period. There is, however, no course credit assigned to students isolating the target product or obtaining usable data. All students have access to stock data taken from a previous student's experiment and can use that data to supplement their analysis or as a replacement for their data if they were unable to obtain their own. In our experience, students are highly motivated to obtain their own usable data from the experiment even without any credit associated with their in-laboratory success. This aspect comports with our belief that the most desirable learning outcome of a lab course is for students to construct coherent arguments about the outcome of a reaction based on conceptual models supplemented by authentic experimental and theoretical data ("you are what you assess").

The bulk of the credit-based work for each experiment takes place postlab, where students complete a scaffolded analysis that requires them to collate multiple sources of spectroscopic data and computational results. While students do not have a direct incentive to generate high-quality data, it quickly becomes clear to them that the better they perform in the laboratory, the easier it will be to interpret their data. A commitment to authenticity on the part of the instructor requires that students generate and learn to interpret imperfect spectra (e.g., from products containing residual starting material, byproducts, and/or solvents). The experimental data and computational results are frequently interpreted using structure-energy relationships that challenge students to explain how and why the observed outcome was obtained. Thus, it becomes natural that our students are engaged in tasks that are defined as 3D learning by the *Framework*.[8]

Phase 1—Spectroscopy, Spectrometry, and Computational Chemistry

Prior to performing laboratory experiments, students receive three sequential periods of instruction focusing upon molecular structure determination via IR, GC-MS, and NMR data. Using a flipped-classroom approach, students view introductory lectures on spectrometry and spectroscopy prior to meeting as a group with their TA and work though exercises in groups to master the data analysis during the class session. Focus is placed on the use of available empirical resources (e.g., chemical shift ranges, common coupling constants) to make spectroscopic assignments, rather than memorizing facts/trends or explaining how the instruments obtain data. Even during this skill-building portion of the course, strong emphasis is placed on making an argument from spectroscopic evidence regarding the structure of a molecule, the source of impurities in a sample, the major/minor products, and the selectivity of a chemical reaction. The NMR coverage includes ^{1}H, ^{13}C with and without attached proton test (APT), and $^{1}H-^{13}C$ Heteronuclear Single Quantum Coherence (HSQC) NMR techniques.[42] The latter two techniques are not uniformly discussed in the organic chemistry lecture sequence, and many students are therefore introduced to them via the lab course. Students become proficient at assigning NMR spectra of mixtures and determining ratios of molecules present in the sample. The GC-MS material, which also may not have been introduced in any previous coursework, focuses upon identification of molecular ions generated by electron−ionization and understanding their fundamental fragmentation processes. The GC-MS data help students identify how many components are present in a mixture, as well as providing insight into their structures. Students use IR spectroscopy for the identification of functional groups and changes in functionality upon reaction. We are privileged to have access to a 400 MHz NMR spectrometer, two GC-MS instruments, and nine AT-IR spectrometers. While all our exercises and experiments could be adapted to other instructional laboratories with different resources, the combination of these data sources allows the in-depth exploration of organic reactions during phases 2 and 3. Examples of the in-class activities, practice exercises, and spectroscopic resources for phase 1 can be found in the SI.

The first experiment uses WebMO[65] to complete a series of computational chemistry exercises that allow students to master the required computational software and use the results to explain chemical phenomena. We share the perspective argued by others[66] that the role of computational chemistry in the undergraduate curriculum is not primarily about using computers or about how the underlying mathematics are implemented. Operationally, we treat computational chemistry very similarly to spectroscopy and spectrometry, in that we do not assess the methodology or underlying quantum mechanics, nor the difference between various basis sets and levels of theory. Rather, we assess students' abilities to use the results to make arguments about and explain a chemical phenomenon based upon core ideas and/or their ability to justify/rationalize their spectroscopic analysis. Students complete the computational chemistry exercise during two laboratory sessions with the support of prelaboratory videos, a TA-led prelaboratory discussion, and instructors available to help with conceptual and technical issues. This suite of exercises has undergone several iterations, two of which have been published[40,41] and an additional example is provided in the SI. All computational jobs submitted by students can be easily accommodated by a computer cluster with approximately one node (with 12 to 20 processors) per five students.[67] Alternately, the computational data can be provided to students in a manner similar to recently described implementations of computational chemistry in the organic laboratory at UW−Waukesha County and Kirkwood Community College.[68]

Phase 2—Organic I Experiments

Following the skill-building of phase 1, students complete experiments that explore reactions relevant to the Organic I lecture, e.g., acid/base reactions, oxidation,[44−46,60] alkene addition, nucleophilic substitution (S_N1/S_N2),[42,43] and elimination (E1/E2).[42] These experiments, which undergo minor changes each term, form the core learning of CHEM 344 and introduce the practical techniques of separation of biphasic solutions, thin-layer chromatography, heating under reflux conditions, distillation, rotary evaporation, recrystallization, and isolation of a solid by filtration. With the exception of recrystallization and filtration, these techniques are new to virtually all of our students. Detailed explanations of each process are provided, and students are expected to demonstrate an understanding of the underlying principles on course assessments. Their skill with carrying out each technique, however, is not directly assessed. Rather, the experimental techniques are used as a means to obtain authentic spectroscopic data for the students to analyze using the skills developed during phase 1. Through this analysis, combined with the use of computational chemistry, students explore fundamental organic chemistry concepts related to acidity/basicity, conformational analysis, stereochemistry, nucleophilicity, electrophilicity, leaving group stability, solvent effects, regioselectivity, stereoselectivity, reversibility, and Le Châtelier's principle. Recent examples of each of these experiments can be found in the SI.

Phase 3—Organic II Experiments

Following the midterm exam that signifies the transition from Organic I to Organic II content, students engage in a set of experiments that use few new techniques but feature more complex chemical systems. These experiments include electrophilic aromatic substitution (EAS),[54−56] nucleophilic aromatic substitution (S_NAr),[63] reactions of benzyl groups,[64] reactions of aryl halides, pericyclic reactions,[51] and various reactions of carbonyl-containing molecules.[49−53,57−59,61,62] Each semester, students complete up to five experiments from those listed in phase 3 of Table 1. Regardless of the exact experiments chosen, students grow their proficiency with interpreting spectroscopic data and rationalizing the outcome of reactions using computational data due to the increased complexity of these chemical systems. For example, as revealed by NMR and GC-MS data, the crude product obtained from the EAS nitration of bromobenzene[54] contains up to five discernible nitration products. This result is much more complicated than those obtained in phase 1 experiments, which generate one or two products in the reaction mixture. We are not concerned by the

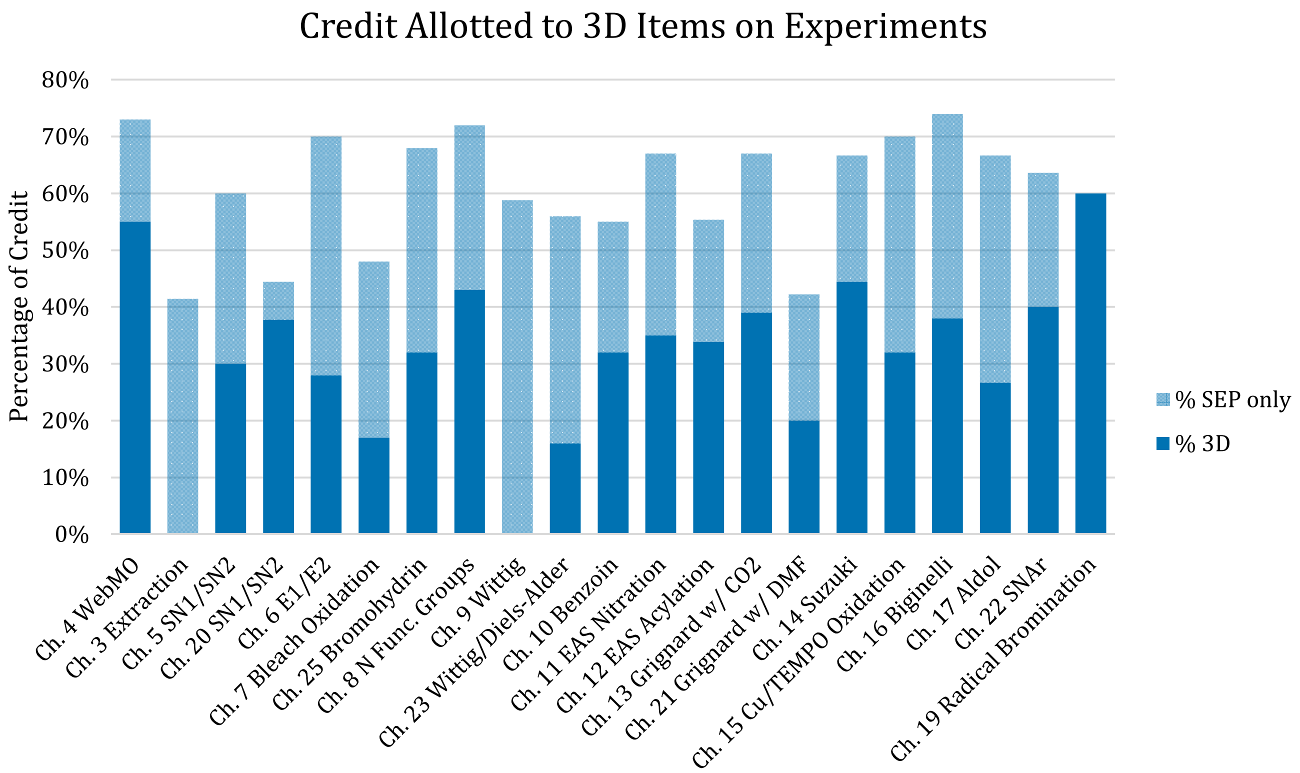


Figure 1. Percentage of credit associated with science and engineering practices and three-dimensional learning on the most recent versions of each CHEM 344 laboratory experiment. The details of the application of the 3D-LAP and each experimental handout can be found in the SI with the chapter numbers listed for each. The order of the listed experiments corresponds to the order presented in Table 1.

totality of reaction coverage in any particular semester but are instead committed to having a variety of reactions that span the typical breadth of the undergraduate organic chemistry curriculum.

Support of Student Learning

Making sense of chemical phenomena is challenging for students and professional chemists alike. Students require substantial support to engage in these activities. As described above, students attend a prelaboratory talk from their TA that encompasses the safety, techniques, data analysis, and theoretical background necessary to complete each experiment. Additionally, in all three phases of the course, students are supported with active-learning group review activities, typically during sessions in which no laboratory work is scheduled. These are well-scaffolded exercises featuring data-focused, 3D questions designed to review the analyses of the previous laboratory experiments and allow students to refine their understanding of the underlying concepts. Students are provided a detailed learning management software course page that includes sample spectra and safety data sheets for all reagent and product molecules, stock data (NMR, IR, GC-MS, TLC plate, etc.) for submission with lab reports if they fail to generate product, lab-specific instructional materials, all in-class activities and their answer keys, and a link to the course Piazza page. Piazza[69] is a crucial resource for students to receive assistance in navigating the course content and logistics. Additionally, students receive assistance via regularly scheduled instructor and TA office hours.

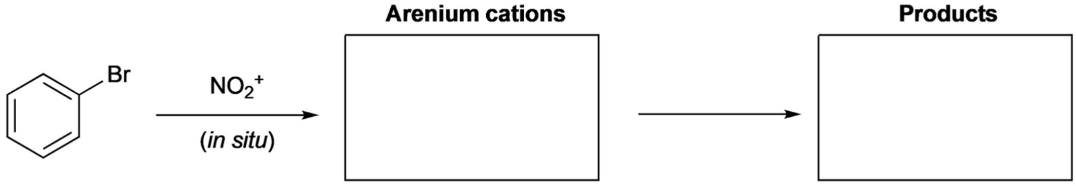
Postlaboratory work submitted for grading is typically due 1 week after the end of the lab session, allowing sufficient time for all NMR and GC-MS samples to be processed and the data uploaded to servers, and for students to generate coherent responses to the postlab exercises. Students' pre- and postlaboratory work is graded promptly by their TAs, providing students with important feedback regarding their interpretation of data and analyses of the chemical reaction. The feedback and subsequent conversations on those assessments ensure that students are able to improve their understanding of chemical systems and data analysis as the term progresses. Our periodic exams provide students feedback on their ability to interpret results and make sense of chemical phenomena in another context. Exams are graded by all course instructors and provide valuable information about students' individual and collective abilities. Throughout the term, students receive regular feedback regarding their overall course grades via emailed grade reports. The grade reports are generated using a custom spreadsheet, similar to that previously described and provided for the organic lecture courses at UW-Madison.[30] Regular grade reports ensure that the students are aware of their current standing in the course, reducing the uncertainty often associated with final course grades.

■ THREE-DIMENSIONAL ANALYSIS OF CURRICULUM

As described above, students are assessed through their laboratory work (pre- and postlab analysis) and written exams. The laboratory work generally accounts for 50−75% of a student's final grade; the exact percentages that each of those items contribute to the overall grade vary semester to semester. All graded items include the analysis of spectra and computational results. Many of these assessments are inherently 3D (i.e., integrate science practices, core ideas, and crosscutting concepts) because students are required to

Phenomenon

Up to 3 different arenium cation intermediates could be formed from the reaction of a mono-substituted benzene with NO_2^+, and thus potentially 3 regioisomers of bromonitrobenzene could be isolated. Drawn an *ortho*, a *meta*, and a *para* arenium cation intermediate and the corresponding product formed during the mono-nitration of bromobenzene.



Obtain the optimized structures of each of the *ortho*, *meta*, and *para* arenium cation intermediates ($C_6H_5NO_2Br^+$) shown below. Calculate the relative energies (B3LYP/6-31G(d), kcal/mol) of the optimized arenium cation intermediates and display the results in a table.

Construct a representation **Provide reasoning**

Draw resonance structures of all the arenium cation intermediates and use them to **explain the stability trend** of the three optimized regioisomeric arenium cation intermediates obtained above.

Figure 2. An example of a 3D task completed in the EAS nitration of bromobenzene experiment.

use structure-energy relationships to rationalize how and why the observed results were obtained.[70]

The 3D-LAP was modified slightly when applied to the CHEM 344 materials because of the structure of the assessment items in the course.[25,28] CHEM 344 exams usually consist of 2−3 questions, each of which consists of 10−15 intentionally scaffolded subquestions. While all of the subquestions can be answered independently, they are linked by a narrative where students analyze the outcome of a single reaction. The authors of the 3D-LAP[25,28] recommend grouping together questions that "are explicitly linked by a diagram, context, question stem, or similar construct," which, if applied to the CHEM 344 materials, would result in 100% of points dedicated to 3D tasks for nearly all the exams and experiments. To estimate a lower bound on the percentages dedicated to 3D questions for exams and experiments, each subquestion was analyzed independently via the 3D-LAP. This results in much lower percent-3D values, due to many questions only containing science and engineering practices, e.g., analyzing and interpreting data. Either the traditional approach or our modified 3D-LAP provides evidence that a substantial amount of course credit is devoted to items that have the potential to elicit evidence of 3D-learning. Our lower values reflect a reasonable estimate of the instructors' intent to have students contextualize assessment items in a manner that causes them to use SEPs through CCCs to apply DCIs to make sense of chemical phenomena. The extent to which students see individual subquestions as connected in this manner, and thus useful for making sense of each reaction, is unknown and almost certainly varies from student to student.

Laboratory Experiments

To support claims about the emphasis our curriculum places on integrating core chemistry concepts and practices, we analyzed the pre- and postlab exercises from each experiment using the 3D-LAP developed by Cooper and co-workers.[25,28] The results of the 3D-LAP are summarized in Table 1 and Figure 1. It is readily apparent that the amount of pre- and postlab analysis that involves 3D learning is highly variable from experiment to experiment. It is also clear that our curriculum places substantial emphasis on scientific practices, especially analyzing and interpreting data, but not all of these practices are directly related to a core idea. This is consistent with the work of practicing chemists; it is common to analyze and interpret data to make claims about what happened without necessarily rationalizing the origin of the outcome. Generally, about 1/3 of the credit is associated with 3D tasks and about 1/3 of the credit comes from SEPs that are not directly tied to a DCI.[71] This is a consequence of the need for students to analyze their data sufficiently to be able to make connections between theory and the experimental outcome. Students receive credit for their spectroscopic and/or computational analyses, which is an important scaffold, despite it lowering the perceived 3D nature of the course.

The 3D tasks often involve rationalizing the experimental outcome based upon the reaction mechanism (explored further below). Provided that the rationalization involves a written description of how an energy difference related to the potential energy surface controls the outcome, the task will almost certainly be 3D. The incorporation of computational chemistry throughout the curriculum provides theoretical support for students to approximate these energy differences. For example, the nitration of bromobenzene[54] is a kinetically controlled reaction for which the regiochemical outcome can be rationalized by comparing the relative energies of the arenium cation intermediates. The prompt shown in Figure 2 and the subsequent spectroscopic analysis involve each component of 3D learning (SEPs, CCCs, and DCIs). Students optimize the *para*-, *meta*-, and *ortho*-nitrated arenium cation intermediates and rationalize their relative energies by describing the differences in charge delocalization and steric interactions among the cations. In the subsequent spectroscopic analysis, knowledge of the relative energies of the arenium cations and

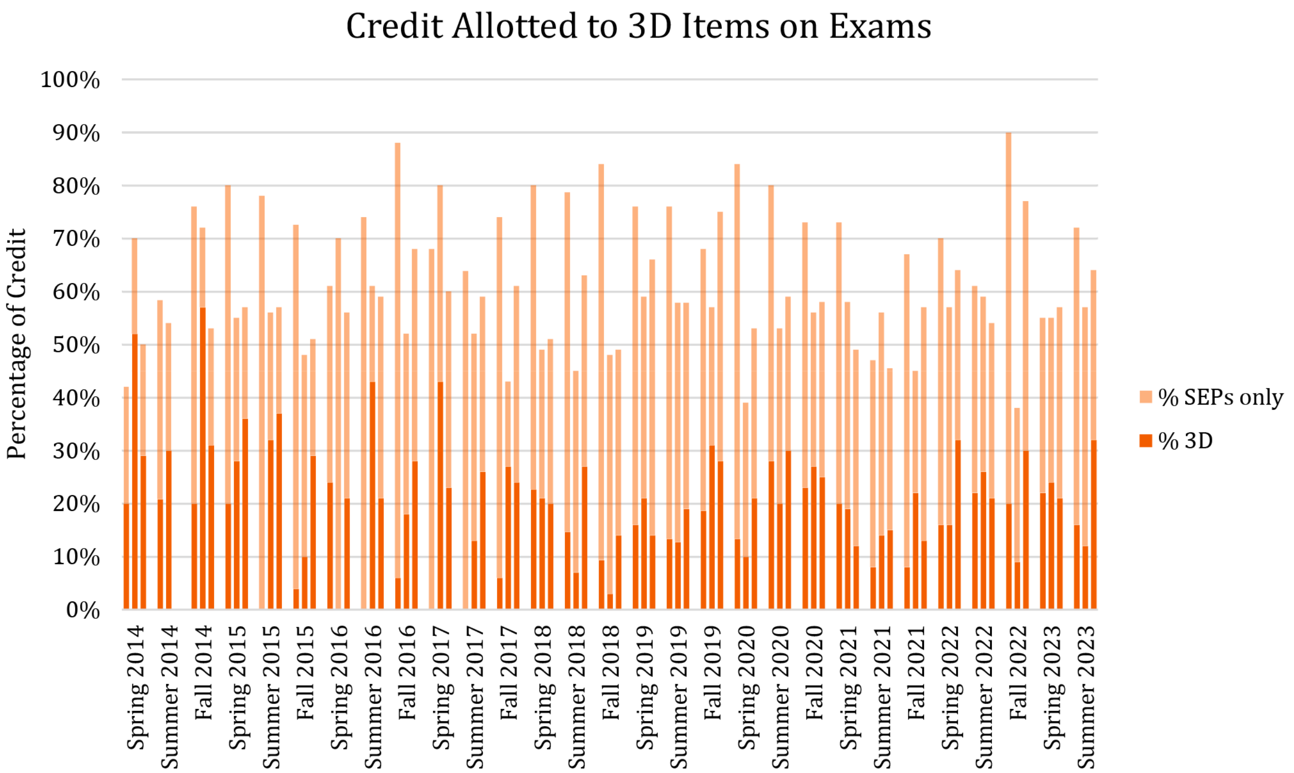


Figure 3. Percentage of credit associated with 3D items (solid orange) and SEPs (transparent orange) by term, with spectroscopy, spectrometry, and WebMO exam (left), midterm exam (middle), and final exam (right). The details of the application of the 3D-LAP and each exam can be found in the SI. Three sample exams are also provided.

the details of the reaction mechanism helps students to predict which products are likely to be present in the crude product mixture.

Our analysis revealed that the *Chapter 3 Separation of a Mixture by an Acid—Base Extraction* and the *Chapter 9 Wittig Synthesis of Ethyl Cinnamate* materials contained 0% 3D learning opportunities, despite being rich with SEPs (41% and 59%, respectively). Students had to predict and interpret a lot of spectroscopic data and computational results in each of these experiments. These analyses, however, were not directly connected to DCIs. The laboratory instructors chose to address the non-3D nature of these experiments in different ways. The acid—base extraction lab, which had been used for many years, was removed from the curriculum. In order to incorporate core ideas of structure and bonding, the extraction lab pre- and postlaboratory assignments needed to become less authentic in terms of SEPs and incorporate more lecture-style questions that were less directly relevant to what students did in the laboratory. The reason for this is that the experience was less an experiment (no intentional organic reaction) and more an exercise in learning a separation technique using an arbitrary mixture of compounds whose solubility was easy to manipulate by acid—base chemistry. Such a technique-focused experiment was incongruent with the remaining course content, and ultimately not needed. In contrast, the Wittig experiment[50] was an interesting chemical reaction that generated a mixture of stereoisomers such that students could determine the stereoselectivity of the reaction via analysis of NMR and GC-MS data. Unfortunately, the structure-energy relationships that govern the stereoselectivity of the reaction are beyond the level that most undergraduate students can readily address in the context of our course. The spectroscopic analysis is sufficiently rich that the laboratory instructors have elected to keep the Wittig reaction in the regular rotation of phase 3.

Written Exams

Meaningful assessment of student learning is a substantial challenge for all educators and is particularly daunting for instructors at large institutions with high-enrollment laboratory courses where each section is directly led by a TA. While there are many assessment options in a laboratory course, we decided that a combination of pre- and postlaboratory assignments and written exams provided an efficient and effective method to assess student learning. Similar results could likely be obtained through publication-style scientific writing or cumulative experiments. However, the time limits inherent to employing graduate students as TAs and the high volume of students enrolled each term led us to focus on written exams. In keeping with the goal of authenticity, our exams provide students with a procedure, experimental spectra, and computational results pertaining to a reaction that they have not physically performed but is closely related to one of the experiments shown in Table 1. Well-scaffolded questions guide students through the required analysis, allowing them to complete a rather difficult task in an exam setting. Students are provided with extensive resources for NMR, IR, and MS that greatly reduce the need to memorize information. These exams are designed to reflect the authentic analysis completed in the laboratory assignments, and likewise typically have a substantial amount of credit associated with three-dimensional learning (Figure 3) and an even larger amount that involves scientific practices. During the preceding eight-year period, the 3D content of each exam has been around 20%. As a result of our commitment to students analyzing authentic data on the exams for reactions that they have not carried out, a large

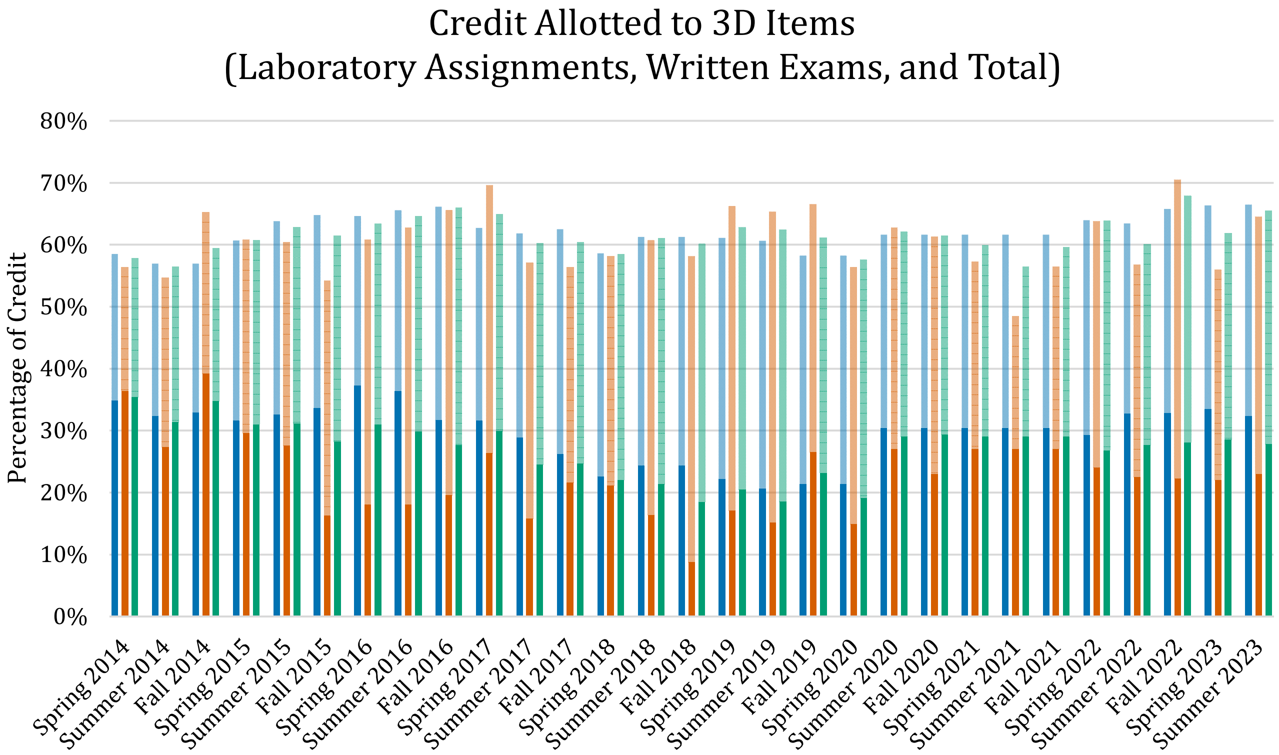


Figure 4. Percentage of credit associated with 3D items (solid colors) and SEPs (transparent) by term, with laboratory assignments (blue, left), written exams (middle, orange), and total course credit (right, green). The details of the application of the 3D-LAP and each exam can be found in the SI.

amount of the remaining portion of the credit on the exams is devoted to analysis of computational results or spectroscopic data sets. While these do not qualify as three-dimensional on their own, they are critical components to support the three-dimensional questions that follow.

In addition to performing the 3D-LAP on the most recent version of each experiment (Table 1) and the written exams (Figure 3), we analyzed all the experiments since 2014 to allow us to determine the total percent of 3D items in the course (Figure 4). Over a decade, the emphasis on 3D assessments resulted in about 20% of the credit coming from 3D items and a large amount of the remaining credit from engaging in SEPs (an additional 40%). There is an interesting trend revealed by this analysis. In our curriculum innovations, we introduced questions that were 3D in 2014, but that were not authentic to the reaction that the students carried out. In the process of removing those somewhat unrelated questions, we inadvertently reduced the % 3D of the overall laboratory course on a few assessments. This was remedied by intentionally creating more questions that asked students to explain how and why an observed chemical outcome was obtained.

■ CRITERIA FOR EVALUATING NEW EXPERIMENTS FOR INCLUSION IN THE CURRICULUM

As demonstrated by the rotation of experiments within the curriculum, the inclusion of a specific experiment or set of experiments is not critical to the outcome of the course. Many combinations of these experiments are consistent with the three pedagogical choices described above, contain many SEPs and 3D-learning opportunities for students, and increase in complexity throughout the term from phase 2 to 3. The course curriculum is ever evolving and incorporates new reactions into student laboratory experiments, review activities, and/or exam questions. In order for an experiment to be incorporated into our curriculum, it must be able to be safely conducted by novice experimentalists in the infrastructure available, be operationally simple enough that a new graduate student TA can lead novices through the procedure, generate a spectrally rich set of products, be amenable to students performing calculations via WebMO, and be conceptually appropriate such that three-dimensional assessment items are possible. As outlined by the previous discussion of the Wittig reaction, this requires that the reactions are sufficiently rich and that students can rationalize how the transformation occurs using some structure-energy relationship. The chemistry education literature has a wide range of organic experiments that are operationally simple and can be conducted safely by novice experimentalists. Many of these experiments, however, produce molecules whose spectra are either far too simple for pedagogically useful analysis by multiple modern spectroscopic techniques (IR, NMR, GC-MS) or too complicated for students to assign all the ^{1}H and most of the ^{13}C NMR signals. We prefer reactions that generate multiple molecules of moderate complexity, with the desired product in highest yield. This is a more authentic experience in terms of organic research, where spectra are rarely obtained for perfectly pure mixtures of the desired product the first time a reaction is completed. Analysis of the resulting crude reaction mixture allows for a deeper exploration of the reaction mechanism and its selectivity when more than one compound is present. Each laboratory experiment is accompanied by a computational chemistry component to assist in spectroscopic assignment and to probe the structure or reactivity of a reaction component. As stated above, the use of authentic spectra and the application of computational chemistry allows us to implement authentic three-dimensional postlab questions. If there is no clear development pathway leading to a meaningful assessment of student learning using the items outlined above, then an experiment is not considered for incorporation into the curriculum.

■ CONCLUSIONS

Reflections

Engaging students in real scientific work is intuitively appealing to educators. A great deal of thought, however, is required to turn this abstract and somewhat theoretical goal into a working curriculum. For a ten-year period, the laboratory directors have continually generated assessments rich in scientific practices, cross-cutting concepts, and core ideas. We have created an organic laboratory curriculum that provides students with continuous support as they engage in authentic analysis of organic reactions. Most importantly, we have achieved this on a very large scale, despite the obstacles inherent to high-enrollment courses. We attribute a significant amount of our ability to create a curriculum that focuses intensely on SEPs and DCIs, such as structure-energy relationships, to a few important curricular choices. Students use modern spectroscopic techniques and computational tools to determine and rationalize an authentic experimental outcome in all experiments, practice materials, and assessments. Computational chemistry is embedded throughout the curriculum, allowing students to use structure-energy relationships to explain why the observed outcome occurred. We are hopeful that, as more institutions acquire access to modern instruments and computational chemistry becomes more prominent in the undergraduate curriculum, other institutions will naturally move toward a more 3D curriculum. Our curriculum is not the only way to achieve a highly authentic and 3D curriculum, but it can provide a template for organic laboratory courses that wish to support students in *doing science.* We believe that this design approach will be useful for any college chemistry instructor wishing to move their curricula toward embodying more authentic practice.

One might look at our laboratory curriculum and dismiss its value by describing the experiments as *cookbook laboratories*, arguably a pejorative term that has become common in chemistry education literature.[72−75] Indeed, our students are always provided with a model procedure for each experiment, which does limit the decisions that students can make in the laboratory. While some may see this as a drawback, it is authentic to the experience of nearly all practicing chemists at all levels. It is rare for a researcher to run a new reaction without access to a literature precedent for a very similar process. It takes chemists a fair amount of experience before they can effectively and routinely make reasonable decisions about how to modify existing procedures. Additionally, we have found that experiments with a provided procedure are highly scalable and can be easily run by students supported by a TA who may be teaching for the first time. We believe that the most important part of a laboratory course should be how students interpret and analyze their data, not the way in which they acquire the data. We encourage instructors to focus less on how laboratory experiments are structured and whether/how students acquire certain manipulative skills and place more emphasis on the types of intellectual work in which students are engaged throughout the experiment.

Future Directions

All curricula are works-in-progress; they are never complete. As new resources become available and new intellectual infrastructure is developed, new (and better?) ways to engage students with course content will continually emerge. We plan to continue to diversify our experiments and expand the scope of reactions that students can complete during the course. As instrument availability increases and computational software and hardware continue to improve, we will maintain a state-of-the-art incorporation of these components. This will allow us to provide students with an increased breadth of chemical phenomena to study. We have provided substantial resources to support student learning and have attempted to provide an avenue to success for all students regardless of their preparation in previous courses.

We are confident that we have built a coherent, authentic, and conceptually rich laboratory curriculum, yet we acknowledge our assessment of student performance is limited to instructor-authored exams and the observations of course instructors. There are many open questions about the curriculum that require additional data and more nuanced analyses to answer. We know that a large amount of the course assessments and practice materials are devoted to students engaging in SEPs and 3D learning. We do not know the extent to which the 3D practice items support success in the 3D lab and written exam assessments, nor do we know the extent to which students' lecture course emphasis on 3D learning supports them in the laboratory. We currently do not know how transferable students' abilities to rationalize chemical phenomena are outside a laboratory report or exam question. We may gain insight into each of these by conducting cognitive clinical[76,77] and stimulated-recall[78] interviews with students. Interview data could be triangulated by embedding research-focused 3D assessments, similar to those conducted for the accompanying lecture courses.[38]

Even if such a study were to indicate that students are highly proficient at rationalizing chemical phenomena and communicating their understanding effectively, we need insight into the goals and motivations behind students engaging in those practices. Students may engage in scientific practices but for different reasons than a practicing chemist would, a phenomenon Berland and Hammer call *pseudopractices*.[79] For example, students may analyze an IR spectrum because it helps them determine the product structure or because the lab report awards points for analyzing it (or most likely for some combination of reasons). The students' view of the nature and use of knowledge on 3D tasks are currently inferred only through the observation of the course instructors. Does the narrative approach of the data sets, as demonstrated on the practice materials, laboratory assignments, and written exams, provide students with a sense that the individual analyses are creating a coherent explanation of the phenomenon? Or to what extent do they see the well-scaffolded exam items as individual prompts to address without the ability to help them make sense of a complicated chemical system? Are there changes to the style of the postlaboratory and written assessments that could move students toward viewing the data analysis as a coherent narrative? Or have we reached the limit of what the current types of assessments can provide to students? Are such changes practical given the limitations of the educational system and context of CHEM 344? Given the highly authentic, scientific practice-focused nature of the curriculum described in this work, the answers to these questions have the potential to provide meaningful direction to curriculum designers in chemistry laboratory and lecture courses.

■ ASSOCIATED CONTENT

Supporting Information

Additional course materials including staff notes and answer keys are available from the authors. The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.3c00845.

Example syllabus and three-dimensional learning and science and engineering practices for individual experiments by term (PDF)
Laboratory manual chapters (ZIP)
Sample exams (ZIP)
Problem sets, in-class activities, and review activities (ZIP)
Summary of the 3D-LAP analysis (XLSX)

■ AUTHOR INFORMATION

Corresponding Authors

Brian J. Esselman − *Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States*; orcid.org/0000-0002-9385-8078; Email: esselman@wisc.edu
Nicholas J. Hill − *Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States*; orcid.org/0000-0002-4510-6699; Email: hill@chem.wisc.edu
Ryan L. Stowe − *Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States*; orcid.org/0000-0002-5548-495X; Email: rstowe2@wisc.edu

Authors

Kimberly S. DeGlopper − *Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States*; orcid.org/0000-0002-9738-7963
Aubrey J. Ellison − *Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States*
Cara E. Schwarz − *Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States*; orcid.org/0000-0003-0552-8886
Niall J. Ellias − *Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jchemed.3c00845

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We gratefully acknowledge the staff members who have supported this course (Paul McGuire, Heike Hofstetter, Charlie Fry, Cathy Clewett, Ross Bishop, and Cristina Bahaveolos) and the numerous staff members, students, and teaching assistants that have offered helpful feedback for course improvement. We thank our collaborators on curriculum development Will Tucker, Maria Zdanovskaia, and Asif Habib. R.L.S., K.S.D., C.E.S., and N.J.E. thank the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin—Madison with funding from the Wisconsin Alumni Research Foundation. The instrumentation (IR, NMR, and GC-MS) were obtained with funds from the University of Wisconsin's Instructional Laboratory Modernization Program.

■ REFERENCES

(1) Shaffer, D. W.; Resnick, M. "Thick" authenticity: New media and authentic learning. *J. Interact. Learn.* **1999**, *10* (2), 195−216.

(2) There are many published examples of undergraduate organic chemistry experiments where students are given an unknown reagent and determine its identity after-the-fact by using spectra of the product. This is inauthentic to how virtually all practicing organic chemists carry out chemical reactions.

(3) The synthetic and analytical techniques that are important to practicing chemists' work are often far more specialized than those presented in undergraduate chemistry courses. We acknowledge that the availability of instrumentation, resources, and time limit what is practical at each institution. Wherever possible, modern analytical techniques should replace qualitative organic class tests for analyzing the outcome of a chemical reaction.

(4) This requires that students are able to sufficiently analyze spectroscopic data to identify what has occurred, employ computational chemistry to understand the molecular and electronic structure of molecules, and to relate that to an appropriate potential energy surface. All of these activities require substantial practice and ongoing support.

(5) We acknowledge that an important branch of chemistry research involves developing new analytical techniques, instrumentation, or analyses. These important endeavors are not regularly part of the undergraduate curriculum or the pursuit of many practicing chemists.

(6) Stowe, R. L.; Scharlott, L. J.; Ralph, V. R.; Becker, N. M.; Cooper, M. M. You Are What You Assess: The Case for Emphasizing Chemistry on Chemistry Assessments. *J. Chem. Educ.* **2021**, *98* (8), 2490−2495.

(7) *Next Generation Science Standards: For States, By States*; The National Academies Press: Washington, D.C., 2013 DOI: 10.17226/18290.

(8) *National Research Council A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*; The National Academies Press: Washington, D.C., 2012 DOI: 10.17226/13165.

(9) Schwarz, C. V.; Passmore, C.; Reiser, B. J. *Helping Students Make Sense of the World Using Next Generation Science and Engineering Practices*; NSTA Press: Arlington, VA, 2017.

(10) DeBoer, G. *A History of Ideas in Science Education*; Teachers College Press: New York, 2019.

(11) Odden, T. O. B.; Marin, A.; Rudolph, J. L. How has Science Education changed over the last 100 years? An analysis using natural language processing. *Sci. Educ.* **2021**, *105* (4), 653−680.

(12) Lehrer, R.; Schauble, L. The Development of Scientific Thinking. In *Handbook of Child Psychology and Developmental Science*; John Wiley & Sons: Hoboken, NJ, 2015; pp 1−44.

(13) Ford, M. J. 'Grasp of Practice' as a Reasoning Resource for Inquiry and Nature of Science Understanding. *Sci. Educ.* **2008**, *17* (2), 147−177.

(14) Manz, E. Representing Student Argumentation as Functionally Emergent From Scientific Activity. *Rev. Educ. Res.* **2015**, *85* (4), 553−590.

(15) Jiménez-Aleixandre, M. P.; Bugallo Rodríguez, A.; Duschl, R. A. Doing the lesson" or "doing science": Argument in high school genetics. *Sci. Educ.* **2000**, *84* (6), 757−792.

(16) Schwarz, C. V.; Ke, L.; Salgado, M.; Manz, E. Beyond assessing knowledge about models and modeling: Moving toward expansive, meaningful, and equitable modeling practice. *J. Res. Sci. Teach.* **2022**, *59* (6), 1086−1096.

(17) Walker, J. P.; Sampson, V.; Zimmerman, C. O. Argument-Driven Inquiry: An Introduction to a New Instructional Model for Use in Undergraduate Chemistry Labs. *J. Chem. Educ.* **2011**, *88* (8), 1048−1056.

(18) Walker, J. P.; Sampson, V.; Southerland, S.; Enderle, P. J. Using the laboratory to engage all students in science practices. *Chem. Educ. Res. Pract.* **2016**, *17* (4), 1098−1113.

(19) Cooper, M.; Klymkowsky, M. Chemistry, Life, the Universe, and Everything: A New Approach to General Chemistry, and a Model for Curriculum Reform. *J. Chem. Educ.* **2013**, *90* (9), 1116−1122.

(20) Cooper, M. M.; Stowe, R. L.; Crandell, O. M.; Klymkowsky, M. W. Organic Chemistry, Life, the Universe and Everything (OCLUE): A Transformed Organic Chemistry Curriculum. *J. Chem. Educ.* **2019**, *96* (9), 1858−1872.

(21) Carmel, J. H.; Herrington, D. G.; Posey, L. A.; Ward, J. S.; Pollock, A. M.; Cooper, M. M. Helping Students to "Do Science": Characterizing Scientific Practices in General Chemistry Laboratory Curricula. *J. Chem. Educ.* **2019**, *96* (3), 423−434.

(22) Stowe, R. L.; Cooper, M. M. Practicing What We Preach: Assessing "Critical Thinking" in Organic Chemistry. *J. Chem. Educ.* **2017**, *94* (12), 1852−1859.

(23) Cooper, M. M.; Posey, L. A.; Underwood, S. M. Core Ideas and Topics: Building Up or Drilling Down? *J. Chem. Educ.* **2017**, *94* (5), 541−548.

(24) Cooper, M. M. The Crosscutting Concepts: Critical Component or "Third Wheel" of Three-Dimensional Learning? *J. Chem. Educ.* **2020**, *97* (4), 903−909.

(25) Laverty, J. T.; Underwood, S. M.; Matz, R. L.; Posey, L. A.; Carmel, J. H.; Caballero, M. D.; Fata-Hartley, C. L.; Ebert-May, D.; Jardeleza, S. E.; Cooper, M. M. Characterizing College Science Assessments: The Three-Dimensional Learning Assessment Protocol. *PLoS One* **2016**, *11* (9), No. e0162333.

(26) Stowe, R. L.; Cooper, M. M. Arguing from Spectroscopic Evidence. *J. Chem. Educ.* **2019**, *96* (10), 2072−2085.

(27) Underwood, S. M.; Posey, L. A.; Herrington, D. G.; Carmel, J. H.; Cooper, M. M. Adapting Assessment Tasks To Support Three-Dimensional Learning. *J. Chem. Educ.* **2018**, *95* (2), 207−217.

(28) Laverty, J. T.; Underwood, S. M.; Matz, R. L.; Posey, L. A.; Carmel, J. H.; Caballero, M. D.; Fata-Hartley, C. L.; Ebert-May, D.; Jardeleza, S. E.; Cooper, M. M. Comment on "Analyzing the Role of Science Practices in ACS Exam Items. *J. Chem. Educ.* **2017**, *94* (6), 673−674.

(29) Carmel, J. H.; Ward, J. S.; Cooper, M. M. A Glowing Recommendation: A Project-Based Cooperative Laboratory Activity To Promote Use of the Scientific and Engineering Practices. *J. Chem. Educ.* **2017**, *94* (5), 626−631.

(30) Stowe, R. L.; Esselman, B. J.; Ralph, V. R.; Ellison, A. J.; Martell, J. D.; DeGlopper, K. S.; Schwarz, C. E. Impact of Maintaining Assessment Emphasis on Three-Dimensional Learning as Organic Chemistry Moved Online. *J. Chem. Educ.* **2020**, *97* (9), 2408−2420.

(31) Collison, C. G.; Kim, T.; Cody, J.; Anderson, J.; Edelbach, B.; Marmor, W.; Kipsang, R.; Ayotte, C.; Saviola, D.; Niziol, J. Transforming the Organic Chemistry Lab Experience: Design, Implementation, and Evaluation of Reformed Experimental Activities—REActivities. *J. Chem. Educ.* **2018**, *95* (1), 55−61.

(32) Flynn, A. B.; Biggs, R. The Development and Implementation of a Problem-Based Learning Format in a Fourth-Year Undergraduate Synthetic Organic and Medicinal Chemistry Laboratory Course. *J. Chem. Educ.* **2012**, *89* (1), 52−57.

(33) Costantino, L.; Barlocco, D. Teaching an Undergraduate Organic Chemistry Laboratory Course with a Tailored Problem-Based Learning Approach. *J. Chem. Educ.* **2019**, *96* (5), 888−894.

(34) Allen, W. E.; Hosbein, K. N.; Kennedy, A. M.; Whiting, B.; Walker, J. P. Design and Implementation of an Organic to Analytical CURE Sequence. *J. Chem. Educ.* **2021**, *98* (7), 2199−2208.

(35) Wilczek, L. A.; Clarke, A. J.; Guerrero Martinez, M. d. C.; Morin, J. B. Catalyzing the Development of Self-Efficacy and Science Identity: A Green Organic Chemistry CURE. *J. Chem. Educ.* **2022**, *99* (12), 3878−3887.

(36) Cruz, C. L.; Holmberg-Douglas, N.; Onuska, N. P. R.; McManus, J. B.; MacKenzie, I. A.; Hutson, B. L.; Eskew, N. A.; Nicewicz, D. A. Development of a Large-Enrollment Course-Based Research Experience in an Undergraduate Organic Chemistry Laboratory: Structure−Function Relationships in Pyrylium Photoredox Catalysts. *J. Chem. Educ.* **2020**, *97* (6), 1572−1578.

(37) Bur, S. K.; Pomerantz, W. C. K.; Bade, M. L.; Gee, C. T. Fragment-Based Ligand Discovery Using Protein-Observed 19F NMR: A Second Semester Organic Chemistry CURE Project. *J. Chem. Educ.* **2021**, *98* (6), 1963−1973.

(38) DeGlopper, K. S.; Schwarz, C. E.; Ellias, N. J.; Stowe, R. L. Impact of Assessment Emphasis on Organic Chemistry Students' Explanations for an Alkene Addition Reaction. *J. Chem. Educ.* **2022**, *99* (3), 1368−1382.

(39) Loudon, M.; Parise, J. *Organic Chemistry*; MacMillan Learning: Boston, 2021.

(40) Esselman, B. J.; Hill, N. J. Integrating Computational Chemistry into an Organic Chemistry Laboratory Curriculum Using WebMO. In *Using Computational Methods To Teach Chemical Principles*, ACS Symposium Series, Vol. *1312*; American Chemical Society: Washington, D.C., 2019; pp 139−162.

(41) Esselman, B. J.; Hill, N. J. Integration of Computational Chemistry into the Undergraduate Organic Chemistry Laboratory Curriculum. *J. Chem. Educ.* **2016**, *93* (5), 932−936.

(42) Esselman, B. J.; Hofstetter, H.; Ellison, A. J.; Fry, C. G.; Hill, N. J. $\mathrm{S_N}1$, E1, and E2 Reactions of *tert*-Amyl Compounds: Improved Analysis Using Computational Chemistry and ASAP-HSQC NMR Spectroscopy. *J. Chem. Educ.* **2020**, *97* (8), 2280−2285.

(43) Curran, T. P.; Mostovoy, A. J.; Curran, M. E.; Berger, C. Introducing Aliphatic Substitution with a Discovery Experiment Using Competing Electrophiles. *J. Chem. Educ.* **2016**, *93* (4), 757−761.

(44) Dip, I.; Gethers, C.; Rice, T.; Straub, T. S. Oxidation of 4-tert-Butylcyclohexanol with Swimming Pool Bleach. *J. Chem. Educ.* **2019**, *96* (1), 183−186.

(45) Esselman, B. J.; Hill, N. J. Revisiting the Bleach Oxidation of 4-*tert*-Butylcyclohexanol. *Chem. Educ.* **2018**, *23*, 23−27.

(46) Mohrig, J. R.; Hammond, C. N.; Schatz, P. F.; Morril, T. C. *Modern Projects and Experiments in Organic Chemistry: Miniscale and Williamson Microscale*; W. H. Freeman and Company: New York, 2003.

(47) Andersh, B.; Kilby, K. N.; Turnis, M. E.; Murphy, D. L. Regioselectivity in Organic Synthesis: Preparation of the Bromohydrin of alpha-Methylstyrene. *J. Chem. Educ.* **2008**, *85* (1), 102.

(48) Porter, D. J.; Stewart, A. T.; Wigal, C. T. Synthesis of a Bromohydrin: An Experiment Demonstrating Markovnikov Addition. *J. Chem. Educ.* **1995**, *72* (11), 1039.

(49) Touchette, K. M. Reductive Amination: A Remarkable Experiment for the Organic Laboratory. *J. Chem. Educ.* **2006**, *83* (6), 929.

(50) Nguyen, K. C.; Weizman, H. Greening Wittig Reactions: Solvent-Free Synthesis of Ethyl *trans*-Cinnamate and *trans*-3-(9-Anthryl)-2-Propenoic Acid Ethyl Ester. *J. Chem. Educ.* **2007**, *84* (1), 119−121.

(51) Davie, E. A. C. Multistep Synthesis of a Terphenyl Derivative Showcasing the Diels−Alder Reaction. *J. Chem. Educ.* **2015**, *92* (7), 1209−1213.

(52) Ballisteri, F. P.; Maccarone, E.; Perrini, G.; Tomaselli, G. A.; Torre, M. Reaction of *trans,trans*-1,4-diphenylbutadiene and maleic anhydride. Substituent and solvent effects. *J. Chem. Soc., Perkin Trans. 2* **1982**, No. 3, 273−277.

(53) Ball, D. B. Diastereoselectivity in the Reduction of α-Hydroxyketones. An Experiment for the Chemistry Major Organic Laboratory. *J. Chem. Educ.* **2006**, *83* (1), 101−105.

(54) Esselman, B. J.; Hill, N. J.; Tucker, W. B.; Zdanovskaia, M. A.; Habib, A. D. Deeper Analysis of the EAS Nitration of Bromobenzene: A Computational and Spectroscopic Study for the Undergraduate Organic Laboratory. *Chem. Educ.* **2016**, *21*, 166−172.

(55) Miles, W. H.; Nutaitis, C. F.; Anderton, C. A. Iron(III) Chloride as a Lewis Acid in the Friedel-Crafts Acylation Reaction. *J. Chem. Educ.* **1996**, *73* (3), 272.

(56) Esselman, B. J.; Hill, N. J. Proper Resonance Depiction of Acylium Cation: A High-Level and Student Computational Investigation. *J. Chem. Educ.* **2015**, *92* (4), 660−663.

(57) Hein, S. M.; Kopitzke, R. W.; Nalli, T. W.; Esselman, B. J.; Hill, N. J. Use of ^{1}H, ^{13}C, and ^{19}F-NMR Spectroscopy and Computational Modeling To Explore Chemoselectivity in the Formation of a Grignard Reagent. *J. Chem. Educ.* **2015**, *92* (3), 548−552.

(58) Maher, M. J.; Hayes, C. O.; Vaccaro, F. A.; Flynn, C. B.; Thedford, R. P.; Stephenson, C. J. Investigating a Chemoselective Grignard Reaction in an Undergraduate Discovery Lab To Predict Reactivity and Final Products. *J. Chem. Educ.* **2016**, *93* (8), 1464−1469.

(59) Hill, N. J.; Bowman, M. D.; Esselman, B. J.; Byron, S. D.; Kreitinger, J.; Leadbeater, N. E. Ligand-Free Suzuki−Miyaura Coupling Reactions Using an Inexpensive Aqueous Palladium Source: A Synthetic and Computational Exercise for the Undergraduate Organic Chemistry Laboratory. *J. Chem. Educ.* **2014**, *91* (7), 1054−1057.

(60) Hill, N. J.; Hoover, J. M.; Stahl, S. S. Aerobic Alcohol Oxidation Using a Copper(I)/TEMPO Catalyst System: A Green, Catalytic Oxidation Reaction for the Undergraduate Organic Chemistry Laboratory. *J. Chem. Educ.* **2013**, *90* (1), 102−105.

(61) Damkaci, F.; Szymaniak, A. Multicomponent Heterocyclic Chemistry for Undergraduate Organic Laboratory: Biginelli Reaction with Multiple Unknowns. *J. Chem. Educ.* **2014**, *91* (6), 943−945.

(62) Crouch, R. D.; Richardson, A.; Howard, J. L.; Harker, R. L.; Barker, K. H. The Aldol Addition and Condensation: The Effect of Conditions on Reaction Pathway. *J. Chem. Educ.* **2007**, *84* (3), 475−476.

(63) Goldstein, S. W.; Bill, A.; Dhuguru, J.; Ghoneim, O. Nucleophilic Aromatic Substitution—Addition and Identification of an Amine. *J. Chem. Educ.* **2017**, *94* (9), 1388−1390.

(64) Pavlis, R. R. An easily conducted free radical substitution for organic chemistry courses. *J. Chem. Educ.* **1982**, *59* (8), 658.

(65) WebMO Enterprise; WebMO, LLC.: Holland, MI, 2022. http://www.webmo.net (accessed 10−10−2023).

(66) Grushow, A. A Lab Course in Computational Chemistry Is Not About Computers. In *Using Computational Methods To Teach Chemical Principles*, ACS Symposium Series, Vol. *1312*; American Chemical Society: Washington, D.C., 2019; pp 211−225.

(67) We typically select B3LYP as the level of theory and 6-31G(d) for the basis set. The level of theory or complexity of the molecules, however, can be modified to accommodate the computational infrastructure at other institutions.

(68) Zdanovskaia, M. A.; Schwarz, C. E.; Habib, A. D.; Hill, N. J.; Esselman, B. J. Access to Computational Chemistry for Community Colleges via WebMO. *J. Chem. Educ.* **2018**, *95* (11), 1960−1965.

(69) Piazza. 2023. www.piazza.com (accessed 2023 08−21−2023).

(70) The style of the CHEM 344 assessments predates the instructors being introduced to the concept of three-dimensional learning. They were created in an attempt to mimic the authentic analysis conducted by students in the pre- and postlaboratory exercises on a written assessment.

(71) There is no consensus in the community regarding the needed percentage of points associated with three-dimensional tasks to focus students on three-dimensional learning. We propose that our current level of scientific practices without core ideas and three-dimensional tasks are sufficient to require students to engage in three-dimensional learning to have success in the course. Further study is needed to determine whether a greater or lesser emphasis is desirable.

(72) Hofstein, A.; Lunetta, V. N. The laboratory in science education: Foundations for the twenty-first century. *Sci. Educ.* **2004**, *88* (1), 28−54.

(73) Cracolice, M. S.; Monteyne, K. What's Wrong with Cookbooks? A Reply to Ault. *J. Chem. Educ.* **2004**, *81* (11), 1559.

(74) Ault, A. What's Wrong with Cookbooks? *J. Chem. Educ.* **2004**, *81* (11), 1569.

(75) Ault, A. What's Wrong with Cookbooks? *J. Chem. Educ.* **2002**, *79* (10), 1177.

(76) Ginsburg, H. P. *Entering the Child's Mind: The Clinical Interview In Psychological Research and Practice*; Cambridge University Press: Cambridge, U.K., 1997 DOI: 10.1017/CBO9780511527777.

(77) Russ, R. S.; Lee, V. R.; Sherin, B. L. Framing in cognitive clinical interviews about intuitive science knowledge: Dynamic student understandings of the discourse interaction. *Sci. Educ.* **2012**, *96* (4), 573−599.

(78) Dempsey, N. P. Stimulated Recall Interviews in Ethnography. *Qualitative Sociology* **2010**, *33* (3), 349−367.

(79) Berland, L. K.; Hammer, D. Framing for scientific argumentation. *J. Res. Sci. Teach.* **2012**, *49* (1), 68−94.

JOURNAL OF CHEMICAL EDUCATION

pubs.acs.org/jchemeduc

Article

Impact of Maintaining Assessment Emphasis on Three-Dimensional Learning as Organic Chemistry Moved Online

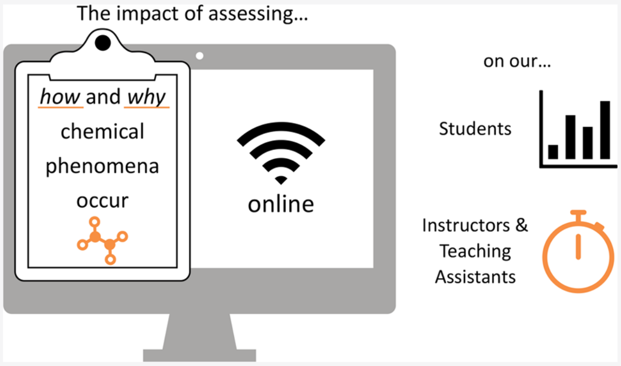
Ryan L. Stowe,* Brian J. Esselman, Vanessa R. Ralph, Aubrey J. Ellison, Jeffrey D. Martell, Kimberly S. DeGlopper, and Cara E. Schwarz

Cite This: https://dx.doi.org/10.1021/acs.jchemed.0c00757

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The affordances given to a structured, timed, and proctored paper exam are not as readily applicable in a digital medium. Accordingly, the rapid shift from in-person to online enactments may have forced instructors to consider changing their assessment practices and priorities. As assessments convey strong implicit messages about "what counts" in a given learning environment, altering what is assessed may have a profound impact on what students view as important in a course. Our four-instructor team sought to examine whether we were able to maintain emphasis on assessing *how* and *why* chemical phenomena occur online while minimizing negative impacts to students, teaching assistants, and ourselves. To support claims regarding the degree to which online assessments emphasized sensemaking relative to past exams, we characterized all summative assessments given in organic chemistry II enactments from 2016 to the present using the three-dimensional learning assessment protocol. To examine the impact of enrolling in a rapidly assembled online organic course on student outcomes, we examined the distribution of students who performed above, at, or below the final exam score predicted by their midterm performance and compared this distribution with historic norms. Results suggest that we were able to maintain emphasis on student sensemaking as our course moved online (∼50% of points on exams administered remotely were dedicated to 3D performances). Additionally, the distribution of students enrolled this past spring who scored above, at, or below the final exam score predicted by their midterm performance was in line with historic norms. When taken in aggregate, our analyses suggest that organic chemistry-enrolled students maintained their ability to make sense of chemical phenomena after the pivot to online instruction. Consistent emphasis on assessing 3D learning online was achieved without adding appreciably to the burden on instructors or teaching assistants due to our assessment writing practices, streamlined approach to online grading, and pre-existing course resources. Instructional implications for assessment design, enacting team grading, and tracking student trajectories are provided in addition to a suite of assessment items with the potential to engage students in sensemaking.



KEYWORDS: *Second-Year Undergraduate, Organic Chemistry, Testing/Assessment, Chemistry Education Research*

FEATURE: Chemical Education Research

In response to the global pandemic wrought by COVID-19 and the subsequent shuttering of in-person classes at institutions of higher education, instructors have been forced to quickly adapt their courses to an online format. Assessment in a hastily assembled, virtual learning environment presents special challenges. The once straightforward logistics of administering a printed exam to a room full of students and subsequently assembling a group of graders to review and discuss exams do not necessarily translate to a digital medium. One might imagine many possible solutions to this set of challenges, including the administration of automatically graded multiple-choice assessments or requiring students to upload files containing their responses to exam prompts.

Unfortunately, the literature on online assessments provides little in the way of precise, actionable recommendations for college chemistry instructors. For example, some studies suggest that online enactments can be as "effective" as face-to-face instruction but fail to elaborate on the characteristics of "effective online environments" in disciplinary contexts.[1] Likewise, literature indicating that timely and meaningful feedback on assessment responses is especially important in online environments does not delve into the particulars of what

Special Issue: Insights Gained While Teaching Chemistry in the Time of COVID-19

Received: June 18, 2020
Revised: July 30, 2020

this sort of feedback should look like in college chemistry courses.[2] Relatedly, very little work has been done on what could or should be assessed in online chemistry learning environments and how the messages about "what counts" might change if assessment foci shift between in-person and online enactments.

As assessments send strong implicit messages to students about the focus of a course,[3−8] modifications to assessment practices and priorities have the potential to alter students' perspective on what is important in a given learning environment. If all (or most) questions on quizzes or exams may be addressed via recall of facts, recognition of patterns, and/or application of simple algorithms, then students will receive and respond to the message that skills and factoids are "what matters".[6] We expect that the vital role of assessment emphasis in charting student priorities would be the same whether instruction is in-person or online. If, in the move to remote instruction, one turns to assessment of trivia as the sole measure of learning, then many enrolled students will likely direct their efforts toward memorizing disconnected bits of information. As we seek to prepare our students for their subsequent courses and careers, such measures of academic achievement would be potentially irrelevant or problematic.

Our four-instructor team (B.J.E., J.D.M., A.J.E., and R.L.S.) is committed to engaging students in weaving together fundamental disciplinary ideas to explain *how* and *why* chemical phenomena occur. By foregrounding making sense of phenomena in terms of atomic/molecular behavior, we hope to illumine the predictive and explanatory power of models useful in organic chemistry, as well as emphasize the broad utility of scientific ways of knowing. Due to the role of assessments in conveying to students the true values of a course, our organic chemistry II instructional team attempted to maintain assessment emphasis on making sense of phenomena as our course moved online. The work presented here examines how successful we were in emphasizing predicting, explaining, and modeling phenomena in the hastily assembled online version of our course. We operationalize "making sense of phenomena" in terms of leveraging the "big ideas" of a discipline to engage in practices characteristic of work in science as framed by cross-cutting lenses. Defining intellectual "heavy lifting" in terms of blending these three elements is consistent with the construct of three-dimensional learning put forth by the National Academies study *A Framework for K-12 Science Education* (the *Framework*).[9] We further aimed to ascertain whether our expectations of students were reasonable—that is, how student performance this semester compared to the performance of students in semesters where no crisis occurred. In pursuit of these goals, we considered the following questions:

(1) How feasible is assessing three-dimensional learning in an online organic chemistry learning environment?

(2) To what extent did the sudden shift to remote assessment impact the predicted outcomes of students?

Question (1) foregrounds the feasibility of writing, administering, grading, and returning assessment items that have the potential to elicit evidence of 3D learning. In addressing this research question, we examine whether our course assessments (i.e., exams, problem sets, discussion activities) did, in fact, emphasize making sense of phenomena and whether emphasis on 3D learning was coherent (that is, pervasive throughout all elements of the course). We further comment on how we were able to administer, grade, and return assessments without significantly increasing the burden on our instructional team. Question (2) centers around the extent to which the rapid shift to remote learning affected the degree to which students' scores on exam 2 (the last exam administered in-person) were predictive of their final exam scores. If students enrolled in organic chemistry II in Spring 2020 more often performed above or below what was predicted than their peers in prior (in-person) semesters, then concerns as to the differences of online and in-person assessments may be justified. Evidence from this study has the potential to support claims as to whether online organic chemistry learning environments can feasibly assess 3D learning without negatively impacting student performance and/or overwhelming instructors.

■ ON THREE-DIMENSIONAL ASSESSMENTS

As a focus of this piece is determining whether a rapidly assembled online organic course could feasibly emphasize making sense of phenomena on assessments, it is worthwhile to briefly discuss the nature of assessment as well as how one might determine whether an assessment has the potential to engage students in 3D performances. Assessment should be viewed as a process of evidentiary argument in which what students know and can do is inferred from their responses to assessment tasks.[10] Although these inferences are necessarily imperfect (as we cannot read minds), the strength of evidence a task can provide is powerfully affected by the structure of that task.[11] Further, interpretation of assessment evidence rests upon assumptions one makes about learning. As has been detailed in past work,[12−14] we view cognition as best modeled by context-sensitive activation of small-grain knowledge elements. Through this lens, student answers to assessment prompts may or may not indicate stable patterns of intellectual resource activation. It is therefore appropriate to provide students with many opportunities to predict, explain, and model related phenomena in order to elicit evidence of the ideas they connect and use to reason across contexts. Stated differently, if instructors wish to elicit persuasive evidence that students can weave together fundamental disciplinary ideas to make sense of a range of phenomena, then they should prompt for construction of explanations and explanatory models on a range of assessments given throughout a course.

As noted previously, "sensemaking" may be more precisely operationalized by specifying how students should draw on "big ideas" to engage in practices characteristic of work in science as framed by cross-cutting lenses. Describing science learning in terms of disciplinary core ideas (DCIs),[15] science and engineering practices (SEPs),[16] and cross-cutting concepts (CCCs)[17] has been termed "three-dimensional learning" by the *Framework*.[9] When we invoke "sensemaking" in this piece, we are referring to cognitive processes useful for figuring out how aspects of the world work.[18−20] These include figuring out "what happened" in a reaction system via analysis and interpretation of spectroscopic data and subsequent argumentation from evidence[12] as well as constructing explanations/explanatory models about how and why a phenomenon occurred.[21] We acknowledge that the construct of "sensemaking" has also been used to describe a stance toward knowledge construction[22−25] as well as discourse practices that make this stance, and the related activation of resources, visible to researchers and educators.[26−28] Our data corpus for this study cannot support claims about students' sense of "what is going on"[29] as they engaged in knowledge-in-use assessments or the discourse practices that might have characterized student work on discussion and homework activities. Accordingly, we make no claims about sensemaking frames[30] or dialogue consistent with such frames.

We refer to "3D assessments" as tasks that have the potential to elicit evidence of students drawing on resources related to "big ideas" to make sense of how and why phenomena occur. Laverty and colleagues have put forth criteria that enable researchers and practitioners to identify and design items that have the potential to elicit evidence of 3D learning.[31] This 3D learning assessment protocol (or 3D-LAP) has been used to characterize how assessments change in response to curricular transformations[32] and to guide modification of traditional assessment items in order that they might elicit stronger evidence of knowledge-in-use.[33] The 3D-LAP finds utility here as a means of characterizing the degree to which organic chemistry learning environments at UW—Madison have and continue to emphasize 3D learning. Importantly, to potentially elicit evidence of 3D learning, a task must *explicitly* ask students to connect "big ideas" to phenomena—this connection should not be inferred without written evidence. Common question types that populate organic chemistry exams, such as "predict the product", "draw the mechanism", or "determine the structure of an unknown compound from spectra" do not typically ask students *why* their claim (e.g., predicted product, mechanism, structure of an unknown) is consistent with relevant scientific principles.[16] Accordingly, we do not know whether students who predicted the proper product or drew a correct mechanism leveraged appropriate concepts. Indeed, it has been reported that students often "decorate" intermediates and products with arrows rather than using curved arrows as predictive tools.[34] Further, even graduate students often propose mechanistic steps because it "gets them to the product" rather than because that step is plausible in the system being examined.[35]

Tasks that satisfy 3D-LAP criteria for the "potential to elicit evidence of 3D Learning" do not inevitably elicit this evidence. Construction of explanations/explanatory models on assessments requires students to call to mind and connect knowledge elements that relate atomic/molecular behavior to phenomena. This is a complex and often counterintuitive task that typically requires explicit prompting.[11,36] Indeed, Cooper and colleagues found that asking students to explain *what* is happening in an acid−base reaction followed by *why* it occurs elicited more sophisticated explanations than simply asking students to explain what they "think is happening at the molecular level".[11] Calibrating the scaffolding of 3D items is nontrivial, as overly structured items might overestimate understanding whereas underspecified prompts might elicit less sophisticated responses than students are capable of constructing. Evidence of response process validity should be collected for any 3D tasks used as research instruments.

We are not advocating that all organic chemistry assessment items be 3D. There are certainly competencies that we want students to develop that do not fall under the umbrella of "making sense of a phenomenon" (e.g., the ability to draw structural representations). Additionally, there is value in prompts that ask students to predict the outcome of a reaction or devise a reasonable synthetic route without requiring that they justify their claim explicitly. We can infer that students struggled to connect appropriate resources if they proposed an extremely unreasonable reaction product or synthesis. We are, however, advocating that a *substantial* amount of points on course assessments be dedicated to 3D tasks. If we want students to receive and respond to the message that figuring out how and why things happen is a central goal of organic chemistry, we must ensure point allocations are consistent with that message. Modern examples of general and organic chemistry curricula that emphasize 3D learning have typically dedicated 35−50% of points on exams to tasks that require one weave together big ideas to engage in a scientific practice.[32]

Challenges to Assessing Learning in Online Environments

The rapid pivot of virtually all in-person university courses to online administration is unprecedented in the modern era. With precious little notice, instructors were tasked with cobbling together instructional and assessment strategies that both embodied course expectations and were flexible enough to accommodate the diverse needs and technological capabilities of enrolled students. For the assessment status quo which, in large-enrollment STEM courses, is often timed, proctored exams completed by individual students without the aid of outside resources cannot be wholly realized online. Attempts to restrict access to static resources (e.g., notes, the textbook) and/or peers while students are engaged in an online assessment can include employing intrusive and bandwidth intensive remote proctoring services and/or restricting assessment availability to a defined time window. We are unaware of literature attesting to the efficacy of these approaches in remote assessment scenarios. Further, we expect reliance on bandwidth-intensive proctoring methods has the potential to widen inequities brought about by disparities in Internet access.[37]

Prompts in which students are asked to explain how and why phenomena occur have the potential to be especially challenging to administer and grade online. Although it is possible to write 3D multiple-choice questions and automate their grading using a learning management system, far stronger evidence of student sensemaking may be elicited by 3D constructed response tasks.[38] Responding to such open-ended tasks requires either a platform that accepts students' drawing and writing (like *beSocratic*)[39,40] or a mechanism by which students can view prompts and upload files containing their responses. Once student responses are received, one then must find a way to view and grade responses to each prompt and communicate these grades to students. If desired, it may also be useful to create a mechanism for students to check correspondence between their responses and the answer key and request a regrade if merited.

■ METHODS

Course Context

This study occurred in the context of the second-semester organic chemistry course in a two-semester sequence. Here, we will refer to this course as "organic chemistry II". At UW—Madison, one organic chemistry sequence serves both chemistry majors and nonchemistry STEM majors. Organic chemistry II typically enrolls between 170 and 250 students per section. Two concurrent sessions are offered in the fall and five concurrent sessions in the spring. All course sections use the sixth edition of Organic Chemistry by Loudon and Parise[41] as their course textbook, and most enactments follow the book's conceptual progression. Students enrolled in organic chemistry II meet as a large group for 150 min each week and attend a weekly 50 min discussion section, where they work in

https://dx.doi.org/10.1021/acs.jchemed.0c00757

groups on tasks set by the instructor and/or teaching assistants (TAs). The course sections that are the focus of this study were structured according to a unified curriculum that places substantial emphasis on students constructing and using atomic/molecular models to make sense of chemical phenomena. This curriculum may be considered "big idea centered" in that, like other transformed courses,[42−44] students are tasked with using fundamental disciplinary ideas (e.g., energy, electrostatics, donor−acceptor interactions) to explain and model increasingly sophisticated systems. Discussion activities and homework problems were designed by two members of the course instructional team (B.J.E. and A.J.E.) to engage students in making sense of phenomena as well as other relevant skills. No course credit was allotted to engagement with discussion activities or work on homework problems. All homework, together with the corresponding keys, was posted to the course learning management system at the beginning of the semester. Course points derive from performance on four examinations, three approximately equally spaced midterms and a final, and three quizzes. All organic chemistry II enactments examined in this work are "on-sequence", meaning that they were offered in the spring semester. The data presented here derives from organic chemistry II enactments from Spring 2016−Spring 2020.

In response to the COVID-19 pandemic, all organic chemistry II instruction was moved online beginning on March 12, 2020. In lieu of attending live whole class meetings, students were to watch prerecorded lectures from the previous spring in which one of the course instructors (B.J.E.) narrated how big ideas could be woven together to explain and predict phenomena. Two midterm examinations were given prior to the pivot to remote instruction and two examinations, a midterm and the final, were given after. Discussion meetings were disbanded after the move to online environments as there was not time to train TAs and students in the use of the relevant software. Instead, students were provided discussion activities and corresponding keys to work through remotely during the virtual portion of the semester. Discussion activities 1−7 were the focus of weekly discussion meetings held in-person, and discussion activities 8−13 were released on Canvas (the learning management system used by UW—Madison) for students to work through remotely. Students could obtain feedback on their discussion activity and homework item responses asynchronously via Q&A platform Piazza. Piazza served as the main forum for student questions for the entirety of the spring semester. Using this platform, students may provide feedback to one another and course instructors may help clarify areas of confusion. Synchronous feedback to student questions was also accessible via regular instructor office hours facilitated by Blackboard Collaborate Ultra.

Approach to Online Assessments

Our approach to assessing student learning in an Internet-mediated course was made more straightforward by our long-standing approach to writing and refining assessment tasks. Before the beginning of the term, the three instructors who enacted organic chemistry II this spring (B.J.E., J.D.M., and R.L.S.) each took the lead on authoring one of the three midterm examinations. Esselman volunteered to draft the final exam. The three-instructor team met to discuss refinements to each exam draft prior to the start of instruction. Refinements focused on (1) the potential of prompts to elicit evidence of students using big ideas to explain phenomena, (2) whether prompts were sufficiently scaffolded, and (3) whether important performances were being assessed. Each draft was revised in response to critiques brought to the fore by the instructor group. Approximately 3 weeks before each exam was to be administered, feedback on the exam draft was solicited from all course TAs at a weekly staff meeting. TA feedback often focused on how tasks could be worded more clearly as well as how the grading rubric should be elaborated. Midterm 3 and the final exam were both given after instruction was moved online. Both assessments were changed minimally relative to the drafts prepared at the start of the semester—what we wanted students to know and be able to do did not change with the move to remote instruction. Changes were made to both remotely administered assessments to reduce the accessibility of answers on the Internet (e.g., removing the trade names of drug molecules students were asked to synthesize).

All course assessments were administered using Canvas after the move to remote instruction. A time window was selected for each exam that fit with most student schedules. Consistent with in-person enactments, conflicts with this time window were collected via a Google form and handled on a case-by-case basis by the instructor team. At the assigned time, students were to download a .pdf of their exam from Canvas, construct responses to assessment prompts, and upload their answers. In order to allow students ample time to contend with technical issues, they were allotted 1 h in addition to the normal time window provided for each exam. Accordingly, students were given 150 min for exam 3 and 180 min for the final exam. One of the course instructors was on-call for the full duration of each exam to address any student questions. All reasonable file types were accepted for upload. We accepted .pdfs annotated by tablet, pictures of responses hand-written on notebook paper, and scans of written responses on a printed exam form. Near the end of an exam submission window, all course instructors assisted students who were experiencing technical difficulties. We observed that these difficulties decreased in number as the semester progressed. Ultimately, all students who wished to submit responses to exam tasks were successful.

Grading responses to tasks which ask students to explain how and why chemical phenomena occur is a time-consuming task. However, if we would like students to receive and respond to the message that sensemaking, rather than trivia, is the focus of organic chemistry, such tasks are an essential part of our learning environments. To streamline our grading process, grading teams were dedicated to a single problem on the exam. Each group was led by regular TAs ($n = 6$) or a course instructor. All add-on graders had experience evaluating student responses to organic chemistry prompts and most were former TAs in our department. Hiring additional graders is common practice in our department whether courses are held online or in-person. We typically aim for a ratio of approximately 20 students/grader. The rubric for each task, which was written and revised along with exam prompts, was loaded into a Google Document to enable each grading team to note alterations to grading procedures. Each grading team was assigned both a primary and a secondary prompt to grade. Once a team completed grading their primary problem, they helped with their secondary assignment. Grading teams worked to ensure consistent grading by discussing any issues that arose with their team via video conference. Exam forms were displayed using the Speedgrader interface in Canvas, which enabled graders to annotate student uploads with points earned per question and other feedback. Point totals on each question were entered into a Google sheet, which tabulated students' total exam score. It took a team of approximately 30 graders around 6 h to grade 556 exams. This is approximately the same amount of time that was required for 30 graders to grade the same number of exams in-person at the beginning of the semester.

We recognize that assembling a team of 30 graders may not be practical for many instructors and that we occupy a privileged position at a well-resourced university. There are a number of strategies that can be used to reduce the grading workload while still emphasizing 3D learning on exams. Many questions in which students are asked to make a claim (e.g., major reaction products, or the absolute configuration of a stereocenter) may be asked as selected response (i.e., multiple-choice) items. Further, as we have reported previously,[16] it is possible to author 3D selected response tasks for organic chemistry. The grading of such tasks may be automated by a learning management system.

Return of graded exams, generation of reports informing students of their standing in the class, and collection of regrade requests was all made more straightforward by pre-existing infrastructure developed by B.J.E. and A.J.E. For example, our course grading spreadsheet, which has undergone numerous refinements over the past 5 years, is able to generate and send individualized grade reports that display a student's standardized grade on the recent exam as well as prior exams in the course, represent a student's grade trajectory using a graph, and list an estimated letter grade for the course, assuming grade cutoffs in-line with historical norms. Grade reports were created by this spreadsheet following each assessment administered online. In addition to releasing grade reports to each student, we also released the key to the exam and invited students to submit regrade requests via a Google form if they believe that points were apportioned incorrectly. One or more of the course instructors reviewed and responded to each regrade request. Releasing grade reports and the exam key is consistent with our long-standing aim of making assignment of grades as transparent as possible to students. The time required to return exams and respond to regrade requests was approximately the same whether enactments were in-person or online. A copy of the grading spreadsheet we used, which enables generation of grade reports and tracks students' grade trajectories throughout the course, has been appended to this article. We are hopeful it will be a useful tool to instructors enacting courses online in the fall.

Student Participants

This work was conducted as a program evaluation at UW—Madison in order to gain insight as to the impact the rapid pivot to remote instruction had on student ability to engage with 3D assessments. As such, no Institutional Review Board approval was deemed necessary.

Our sample is made up of students from five cohorts, with each cohort consisting of students who completed the final exam in sections of organic chemistry II structured according to the unified curriculum mentioned previously. Earlier versions (2016 through 2018) of the course were taught by a single instructor. For the last 2 years, organic chemistry II was taught collaboratively by a team of two (2019) or three (2020) instructors. A total of 1546 students contributed data to this work (see Table 1).

Table 1. Descriptive Statistics Suggest No Considerable Differences in Student Achievement (Cumulative GPA) and Preparation (ACT Composite Scores) Were Observed Amongst Organic Chemistry II Students at the Research Setting[a]

Instrument	Measure	Year				
		2016	2017	2018	2019	2020
$GPA_{Cumulative}$	*n*	215	171	206	398	556
	Min	1.7	1.9	2.0	2.0	1.8
	Median	3.4	3.3	3.5	3.4	3.6
	Max	4.0	4.0	4.0	4.0	4.0
$ACT_{Composite}$	*n*	187	148	192	353	490
	Min	21	22	19	22	18
	Median	29	29	30	30	30
	Max	36	36	36	36	36

[a] A subset of organic chemistry II-enrolled students did not have a composite ACT score on record with the university. Accordingly, the sample sizes listed under "ACT Composite Score" represent fewer than the total number of students enrolled in the sections examined.

To determine whether significant differences could be detected in the academic measures for our three cohorts, we ran a series of Mann−Whitney *U* tests comparing, in a pairwise fashion, cohort ACT scores and cumulative grade point averages at the conclusion of the spring term in which they were enrolled in organic chemistry II. These tests, and all statistical analyses reported in this piece, were performed using SPSS version 26 for Mac.[45] Table S1 in the Supporting Information describes the results of our pairwise Mann−Whitney *U* tests. In summary, whereas there were several statistically significant differences that emerged from these tests, none rose to the level of "practical significance". That is, all significant differences had a small effect size according to guidelines published by Cohen.[46] Our criterion for significance in this study was a p value ≤ 0.01.

RQ1: How Feasible Is Assessing Three-Dimensional Learning in an Online Organic Chemistry Learning Environment?

It is not reasonable to expect students to figure out the cause for a phenomenon on an exam unless they have had substantial prior opportunities to construct and critique explanations and explanatory models. A long chain of often counterintuitive inferences connect fundamental ideas such as bonding and energy to most phenomena.[47,48] Accordingly, students require practice calling to mind and connecting the relevant knowledge elements to construct reasonable explanations for observable events in terms of atomic/molecular behavior.[13] We claim that assessment of 3D learning, whether online or in-person, is only feasible if all parts of the course engage students in making atomic/molecular sense of phenomena. To characterize the degree to which our learning environment coherently emphasized sensemaking, we examined the intellectual work prompted for on both high stakes assessments (i.e., exams) and on low-stakes homework and discussion activities using the 3D-LAP.[31] As examinations given in prior years are provided to students as practice, our analysis included exams from the last 5 years. Two authors (K.S.D. and C.E.S.) independently coded the potential of each task to elicit evidence of student engagement in scientific practices, core ideas, and/or crosscutting concepts. They subsequently met to discuss any discrepancies in coding and reached consensus on the assignment of all codes. Consensus codes describing the potential of each assessment prompt to elicit evidence of 3D learning have been provided in the Supporting Information. All assessment items that were coded have been likewise appended. Importantly, assessment items that meet the 3D-LAP criteria have the *potential* to elicit evidence of 3D learning, but, without further evidence, we cannot infer the degree to which this potential is realized. Additionally, sets of tasks linked to a diagram, context, question stem, or similar construct are coded as a single unit for the purposes of this analysis. Coding clusters of related items with a single set of descriptors has the potential to overstate the amount of points an assessment dedicates to 3D tasks. For example, an eight-part prompt in which only three parts meet 3D-LAP criteria for potentially eliciting evidence of 3D learning would be coded as a “3D prompt”.

RQ2: To What Extent Did the Sudden Shift to Remote Assessment Impact the Predicted Outcomes of Students?

Two approaches were used to determine whether significant differences existed between the achievement measures and grade trajectories of students taught online and those enrolled in in-person coursework. First, we regressed students’ standardized exam 2 (administered in-person) performance onto predictions of standardized final exam scores. Then, we divided the residuals into a top, middle, and bottom third to examine how many students performed above, at, or below what was predicted. Students’ exam 2 score was the most predictive of their final exam score of all achievement measures collected prior to the pivot to remote instruction. A Pearson’s χ^2 test was used to analyze the relationship between enrolling in organic chemistry II during a given semester and the distribution of students who scored “above”, “at”, or “below” their predicted final exam score. The effect size of this association was calculated using Cramer’s V and interpreted using guidelines published by Cohen.[46] These guidelines stipulate that a small effect would have a Cramer’s V of 0.1, a medium effect would have a Cramer’s V of 0.3, and a large effect would have a Cramer’s V of 0.5. As a significant association emerged from this χ^2 test, a posthoc analysis of the resulting contingency table was conducted in order to support inferences as to the driver(s) of that significance. This analysis consisted of comparing the standardized residual for each cell to the critical value, which was 2.58 for this study. Standardized residuals provide a means of evaluating how different the observed value reported in a cell is from what would be expected.[49] The sign of these residuals indicate whether an observed value is greater than expected by chance (in which case it will be positive), or less than expected by chance (in which case it will be negative). Residuals greater in magnitude than the critical value were deemed primary drivers for the significant relationship denoted by the results of the χ^2 test.

We also conducted a series of Mann−Whitney *U* tests comparing, in a pairwise fashion, the distribution of earned points on exam 3 and the distribution of earned points on the final exam. Recall that exams 3 and 4 were held after the move to remote instruction during the Spring 2020 semester. Our threshold for significance in this work was a $p \leq 0.01$. To reduce the risk of a type I error in our analysis of four paired Mann−Whitney *U* tests, we applied a Bonferroni correction—thus, a $p \leq 0.0025$ was our criterion for significance.

■ RESULTS AND DISCUSSION

Three-Dimensional Learning in an Online Organic Chemistry Learning Environment Is Feasible

In Figure 1, we report the percentage of points on organic chemistry II examinations dedicated to tasks capable of eliciting evidence of 3D learning. The number of students who engaged with each assessment is represented by the size of each dot. Earlier versions of the course were taught by a single instructor (n = 215 in 2016, n = 171 in 2017, and n = 206 in 2018). For the last 2 years, organic chemistry II was taught collaboratively by a team of two (2019, n = 398) or three (2020, n = 556) instructors.

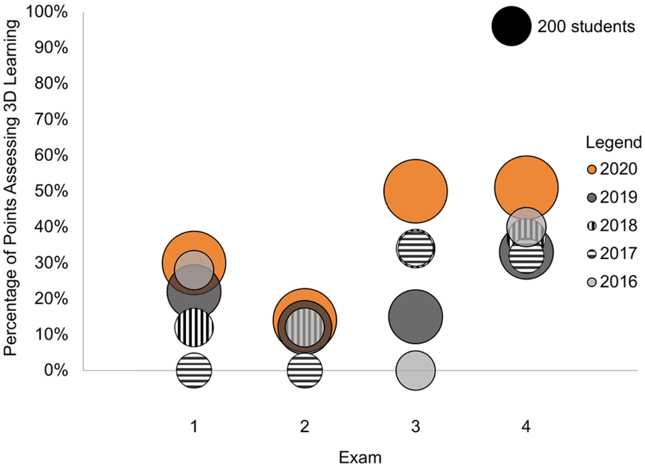


Figure 1. Percentage of points on organic chemistry II examinations dedicated to tasks that have the potential to elicit evidence of 3D learning over a 5 year period. Exams 1−3 represent midterm examinations, and exam 4 represents the final exam for the course. Each data point is scaled by the number of students who took each exam.

Examination of Figure 1 reveals several interesting trends that merit unpacking. During the 5 year period examined, the final exam has always dedicated a substantial amount of points (33−51%) to tasks that require students to explain how and why chemical phenomena occur. As this exam is meant to embody the expectations for the course as a whole, consistent emphasis on 3D prompts is encouraging. It is important to note that a significant amount of points on the final exam given this past spring were dedicated to a large multipart question which, when considered as a single unit of analysis, had the potential to elicit evidence of 3D learning. Although this sequence of prompts related to an overarching narrative, many individual tasks did not meet 3D-LAP criteria for potentially engaging students in making sense of chemical phenomena. Consequently, the high percentage of points we describe as dedicated to 3D assessments on the Spring 2020 final exam is due in part to coding clusters of related questions together. Exam 2 typically places little emphasis on assessing 3D learning (0−14% of points dedicated to 3D tasks). The second midterm assesses a portion of the course that focuses on (1) phenomena deemed too complex to be readily explained by students (e.g., palladium-catalyzed cross-coupling reactions, oxidation of phenols) and (2) slight variations on phenomena grappled with on previous assessments (e.g., reactions of allylic and benzylic systems, electrophilic aromatic substitution of phenol). Accordingly, it is not too surprising that there are

Table 2. Correlations Suggest Students' Performance on Exam 2 Is Most Predictive (among the Measures Collected before the Shift to Online Learning) of Their Final Exam Scores

Year	Measure	$ACT_{Composite}$	$GPA_{Cumulative}$	Z_Exam1	Z_Exam2
2016	R_{Final}	0.125	0.754[a]	0.739[a]	0.825[a]
2017	R_{Final}	0.131	0.758[a]	0.753[a]	0.833[a]
2018	R_{Final}	0.302[a]	0.765[a]	0.750[a]	0.789[a]
2019	R_{Final}	0.295[a]	0.783[a]	0.758[a]	0.825[a]
2020	R_{Final}	0.272[a]	0.611[a]	0.694[a]	0.755[a]

[a] Correlation is significant at the 0.01 level (2-tailed).

fewer prompts on this exam asking students to make sense of observable events. In future iterations of this course, we would like to increasingly emphasize sensemaking during the middle third of the semester. The degree to which exam 1 and exam 3 emphasize making sense of phenomena has varied substantially during the 5 year period examined. However, the first and third midterms given during the spring of 2020 emphasize 3D learning to a greater extent than analogous test variants given in earlier years. Notably, the two high-stakes assessments given after the shift to remote instruction this spring (i.e., exam 3 and exam 4) place substantial emphasis on students making sense of chemical phenomena. This demonstrates that it is possible to continue the focus of assessments on 3D learning in an online learning environment. That is, we were logistically able to create, administer and grade assessments which placed *substantial* emphasis on students constructing explanations and explanatory models detailing how and why phenomena occurred. Additionally, as noted previously, this was accomplished without substantially increasing the burden on our instructional team.

No points are allotted to either discussion activities or problem sets, so it is not possible to report the percentage of points dedicated to 3D tasks on these assessments. Accordingly, we report how many discussion and homework activities had at least one 3D item and how many had two or more such items. Recall that 3D items very often consist of multiple parts and so "one 3D item" may in fact engage students in several related tasks. Out of a total of 14 problem sets, 12 had at least one question that met the 3D-LAP criteria for potentially eliciting evidence of 3D learning. Eight problem sets had two or more 3D assessment prompts. Out of a total of 13 discussion activities for the semester, 7 had at least one 3D question and 3 had two or more 3D questions. Early in the course, there is substantial emphasis on analysis and interpretation of spectroscopic data to support claims as to the structure of an unknown. While this sort of activity constitutes a pairing of two scientific practices (analysis and interpretation of data and argumentation from evidence),[12] students need not understand the conceptual basis of spectroscopic techniques to construct evidence-based claims. Thus, few prompts in the first three problem sets and discussion activities are "three-dimensional". Most 3D tasks given as homework or in discussion meetings required students to construct explanations and/or explanatory models. This is consistent with our stated goal of students explaining how and why chemical phenomena occur. Given that virtually all of the problem sets and well over half of the discussion activities contained at least one item that had the potential to elicit evidence of 3D learning, we would argue that our course consistently emphasized making sense of chemical phenomena. These resources, coupled with practice exams derived from prior years' assessments, were designed to convey the message that weaving together core ideas to explain how and why things happened "counted" in our course. As problem sets and discussion activities, together with corresponding keys, were uploaded to Canvas for students to complete at their leisure, providing students with low-stakes 3D assessments incurred no additional burden on our instructional team after the pivot to remote instruction. As per usual, students were able to obtain feedback on their assessment responses, as well as offer feedback to their peers, via posting on Piazza. Codes describing the emphasis of each problem set and discussion item on scientific practices, core ideas, and/or cross-cutting concepts are appended in the Supporting Information.

Sudden Shift to Remote Assessment Had No Considerable Impact on the Predicted Outcomes of Students

At the conclusion of the spring semester, we sought to assess how the outcomes of students enrolled in our online environment compared to those of students enrolled in more typical enactments. As shown previously, organic chemistry II assessments administered at UW—Madison have long placed substantial emphasis on students explaining how and why chemical phenomena occur. This consistent emphasis on sensemaking, coupled with the similar achievement characteristics of cohorts enrolled in the course from 2016 to 2020, suggests it may be reasonable to compare the performance of these groups of students. Across all years examined, students' performance on exam 2 is strongly correlated with their performance on the final exam (Table 2). Indeed, exam 2 performance is the most strongly correlated of all predictors for student achievement that were collected prior to the shift to remote instruction (e.g., ACT score, composite GPA, exam 1 score). In order to account for differences in exam difficulty, exam scores were standardized by year for this analysis.

We can gain some insight as to whether outcomes this past spring align with historic norms by examining the percentage of students who score considerably higher or lower on the final exam than what would be predicted by their exam 2 score. If a substantial percentage of students score above their predicted final exam score relative to what is typical, widespread collaboration may have inflated scores on Internet-mediated exams. If, conversely, a substantial percentage of students score below their predicted final exam score relative to the norm, the demands of virtual assessment and stressors of the pandemic may have been too great a burden for a subset of students. To conduct this analysis, we regressed the exam 2 scores of the 1546 student participants described in the Methods (see Table 1) onto predictions of achievement on the final exam for all 5 years under study (see Table 3). With such a sample size, error

Table 3. Regressions (by Year) of Students' Standardized Exam 2 Scores into Predicted Final Exam Scores Were Statistically and Substantively Different

Year	F	df	p	η^2	β	t	p	R^2
2016	453.6	1	< 0.001	0.68	0.83	21.3	< 0.001	0.680
2017	382.9	1	< 0.001	0.69	0.83	19.6	< 0.001	0.694
2018	337.1	1	< 0.001	0.62	0.79	18.4	< 0.001	0.623
2019	845.4	1	< 0.001	0.68	0.83	29.1	< 0.001	0.681
2020	743.2	1	< 0.001	0.57	0.76	27.1	< 0.001	0.570

Table 4. Number and Percentages of Students Whose Scores on the Final Exam Were above, at, or below the Predicted Score Extrapolated from Their Exam 2 Performance[a]

Outcome | Year	2016	2017	2018	2019	2020
Above	38	58	119	130	170
At	72	53	72	134	183
Below	105	59	15	133	203
Outcome | Year (%)	**2016**	**2017**	**2018**	**2019**	**2020**
Above	17.7%	34.1%	57.8%	32.7%	30.6%
At	33.5%	31.2%	35.0%	33.8%	32.9%
Below	48.8%	34.7%	7.3%	33.5%	36.5%

[a] Cross-tabulations suggest students in the spring of 2020 did not perform unusually better or worse than predicted.

terms were expected to be independent and approximate normality. As indicated by correlation values listed in Table 2, the relationship between students' exam 2 scores and mean final exam scores is linear (0.755 < R < 0.833). No systematic relationship between standardized residuals and predicted values by the model was observed (see "Evaluating Homoscedasticity" in the Supporting Information). As such, no violations of assumptions of simple linear regression were detected.[50]

Residuals, or the difference between students' actual and predicted scores, were used to characterize whether a student's performance on the final exam was above (top third of all residuals), at (middle third), or below (bottom third) what was predicted from their exam 2 score (see Table 4 and Figure 2). If students in Spring 2020 more often performed above or below what was predicted than their peers in prior (in-person) semesters, then concerns as to the differences between online and in-person assessments may be justified, as mentioned above.

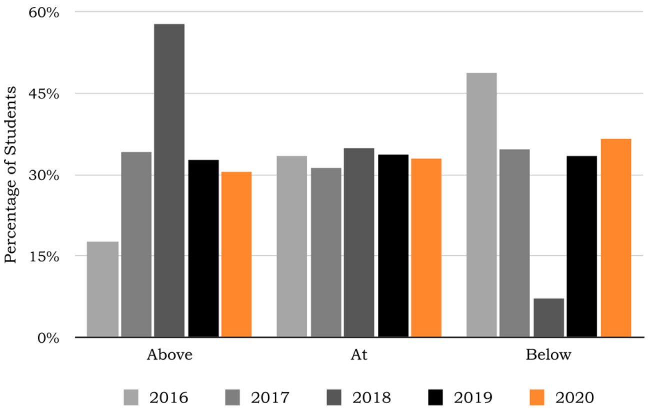
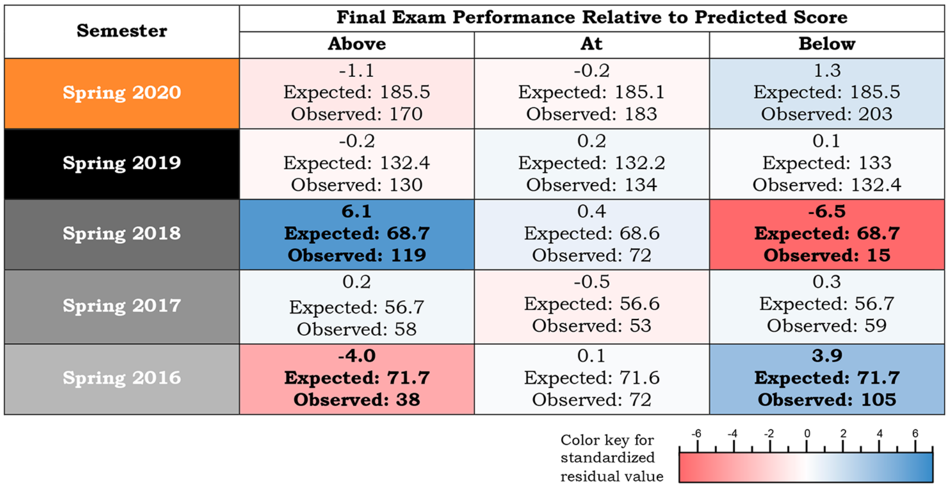


Figure 2. Bar graphs depicting the percentage of students from each year whose scores on the final exam were above, at, or below the predicted score extrapolated from their exam 2 performance.

As is evident from Figure 2, students who enrolled in organic chemistry II during the spring of 2016 more often performed below what was predicted, whereas students who enrolled in the course during the spring of 2018 more often performed above what was predicted. Although the method applied to these data is sensitive to differences in performance by semester, students enrolled in organic chemistry II during the Spring 2020 semester do not appear to score dramatically above or below their predicted final exam score relative to the historical norms.

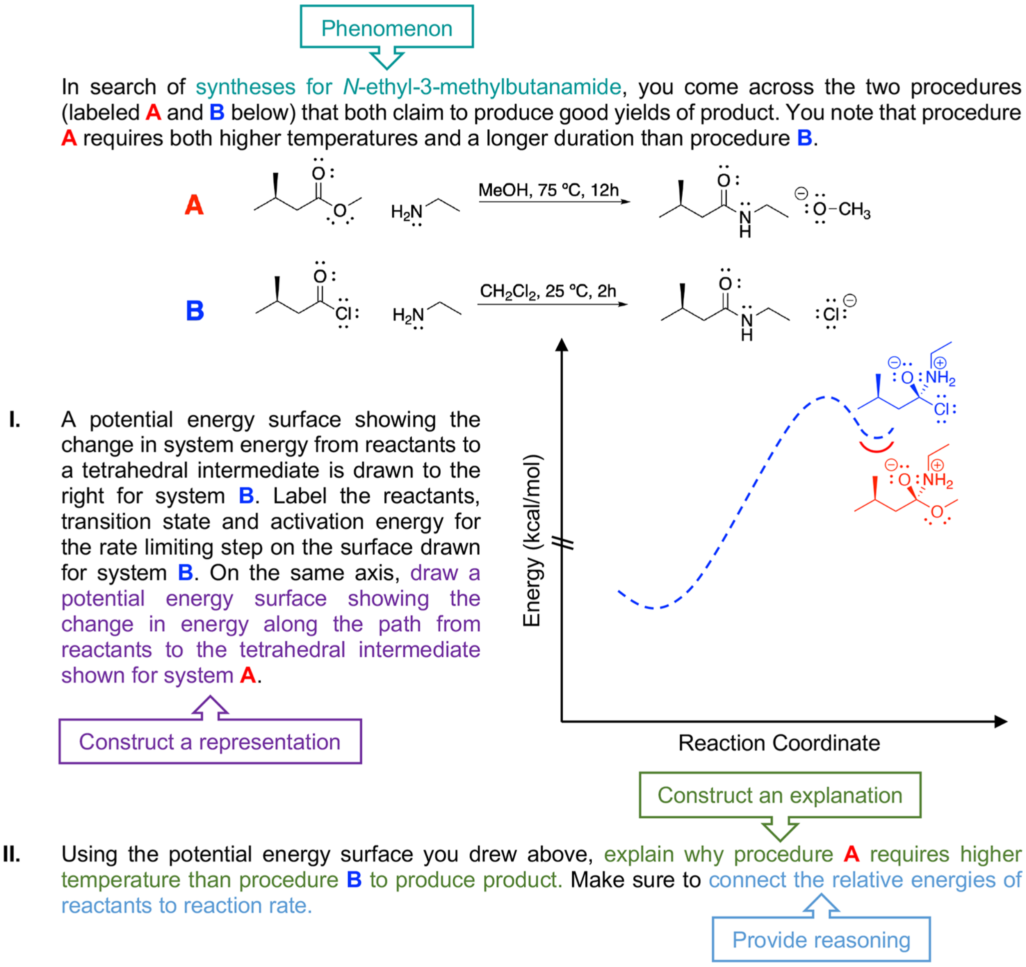
The relationship between enrolling in organic chemistry II in a given semester and the distribution of students who scored above, at, or below their predicted final exam score was examined via a Pearson's χ^2 test. A significant association between the semester enrolled and distribution of student final exam scores relative to their predicted score was indicated by the results of our χ^2 test, $\chi^2(8) = 113.66$, $p < 0.001$, Cramer's V = 0.192. A subsequent posthoc analysis of the results of this test (Figure 3) showed that enrollment in organic chemistry II during the spring of 2018 was strongly associated with students scoring better on the final exam than their exam 2 score would predict. Additionally, enrollment in the course during the spring of 2016 was strongly associated with scoring lower on the final exam than students' exam 2 grade would predict. These associations were the primary drivers of significance for the overall significant association. Consistent with the bar graphs depicted in Figure 2, our posthoc analysis suggests that the distribution of students who scored "above", "at", or "below" their predicted final exam score during the spring of 2020 was in-line with historic norms.

To determine whether significant differences existed between the achievement measures of students taught online



Semester	Final Exam Performance Relative to Predicted Score		
	Above	**At**	**Below**
Spring 2020	-1.1 Expected: 185.5 Observed: 170	-0.2 Expected: 185.1 Observed: 183	1.3 Expected: 185.5 Observed: 203
Spring 2019	-0.2 Expected: 132.4 Observed: 130	0.2 Expected: 132.2 Observed: 134	0.1 Expected: 133 Observed: 132.4
Spring 2018	**6.1** **Expected: 68.7** **Observed: 119**	0.4 Expected: 68.6 Observed: 72	**-6.5** **Expected: 68.7** **Observed: 15**
Spring 2017	0.2 Expected: 56.7 Observed: 58	-0.5 Expected: 56.6 Observed: 53	0.3 Expected: 56.7 Observed: 59
Spring 2016	**-4.0** **Expected: 71.7** **Observed: 38**	0.1 Expected: 71.6 Observed: 72	**3.9** **Expected: 71.7** **Observed: 105**

Figure 3. Contingency table for the χ^2 test examining association between enrolling in organic chemistry II during a given semester and the distribution of students whose final exam scores were "above", "at", or "below" the final exam score predicted by their exam 2 performance. In each cell, the standardized residual value is reported along with the observed and expected values. Standardized residuals larger than the critical value (±2.58) are in bold. To visualize the sign and magnitude of the standardized residuals, the cells are color coded from dark blue (most positive) to dark red (most negative).



Phenomenon

In search of syntheses for *N*-ethyl-3-methylbutanamide, you come across the two procedures (labeled **A** and **B** below) that both claim to produce good yields of product. You note that procedure **A** requires both higher temperatures and a longer duration than procedure **B**.

I. A potential energy surface showing the change in system energy from reactants to a tetrahedral intermediate is drawn to the right for system **B**. Label the reactants, transition state and activation energy for the rate limiting step on the surface drawn for system **B**. On the same axis, draw a potential energy surface showing the change in energy along the path from reactants to the tetrahedral intermediate shown for system **A**.

Construct a representation

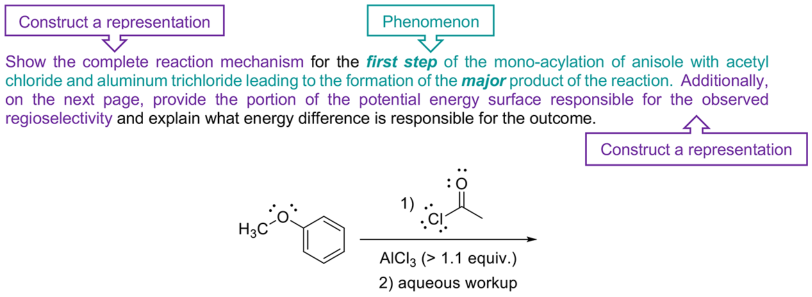
Construct an explanation

II. Using the potential energy surface you drew above, explain why procedure **A** requires higher temperature than procedure **B** to produce product. Make sure to connect the relative energies of reactants to reaction rate.

Provide reasoning

Figure 4. Assessment item with the potential to elicit evidence of student engagement in using models to explain a phenomenon.

and those enrolled in in-person coursework, we also conducted a series of Mann−Whitney *U* tests comparing, in a pairwise fashion, the distribution of earned points on exam 3 and the distribution of earned points on the final exam for students enrolled from 2016 to 2020. Outputs of all statistical analyses of score distributions are reported in Table S2 in the Supporting Information. In summary, no statistically significant differences existed between the distribution of final exam



Show the complete reaction mechanism for the ***first step*** of the mono-acylation of anisole with acetyl chloride and aluminum trichloride leading to the formation of the ***major*** product of the reaction. Additionally, on the next page, provide the portion of the potential energy surface responsible for the observed regioselectivity and explain what energy difference is responsible for the outcome.

Figure 5. Assessment task that includes the word "explain" but does not have the potential to elicit evidence of engagement in the practice of constructing explanations.

scores earned in Spring 2020 and the distribution of final exam scores earned in earlier years. The distribution of scores on exam 3 differed significantly in three of our four pairwise Mann−Whitney *U* tests. Indeed, the difference between the distribution of points earned on the third midterm given during Spring 2020 and the points earned on the analogous exam given during the spring of 2019 is substantial (means of 68 vs 46, $U = 177290.5$, $z = 15.883$, $p < 0.001$, $r = 0.51$; large effect). This is not surprising, given that student scores on the exam 3 form given during the spring of 2019 tended to be much lower than those on any other analogous midterm in our data set. The difference between the Spring 2020 exam 3 score distribution and the distribution of scores on the third midterms given during spring of 2018 met our criteria for "significance" but did not rise to the level of practical significance given its small effect sizes ($U = 43199.0$, $z = -5.215$, $p < 0.001$, $r = -0.19$; small effect). The distribution of scores earned on the exam 3 during Spring 2016 differed substantively from the distribution of scores earned on exam 3 this past spring (means of 68 vs 56, $U = 81620.5$, $z = 7.881$, $p < 0.001$, $r = 0.28$; medium effect).

When taken in aggregate, the analyses we performed to address RQ2 suggest that students enrolled in organic chemistry II this past spring maintained their ability to make sense of chemical phenomena on exams after the pivot to remote instruction. This may, in part, be due to the fact that 3D tasks were consistently emphasized on homework and exams throughout the entire semester. Additionally, the structure of our course and the resources available to students did not change dramatically as we moved operations online. Stated differently, students knew they would be asked to explain how and why phenomena happen on the final exam and likely prepared accordingly.

■ IMPLICATIONS FOR INSTRUCTION

We are hopeful that the analyses presented here suggest to instructors that it is possible to assess authentic intellectual "heavy lifting" as part of online learning environments. Indeed, we would argue that to do otherwise redirects the focus of the course to competencies not reflective of scientific practice. Instructors who desire to incorporate more 3D tasks into their organic chemistry assessments could use the scientific practices criteria given in the 3D-LAP[31] as a guide when writing prompts. These criteria suggest task scaffolding that has the potential to elicit evidence of student engagement in a practice. For example, the criteria for "developing and using models" note that items should (1) give a phenomenon, (2) give or ask the student to construct a representation that relates to that phenomenon, (3) ask student to explain or make a prediction about the focal phenomenon, and (4) provide the reasoning linking their explanation or prediction to their representation. Several papers exist on adapting existing general[33] and organic chemistry[16] tasks to assess 3D learning, as guided by the 3D-LAP. For an item to have the potential to elicit evidence of 3D learning, it must fulfill all of the 3D-LAP criteria for one or more scientific practices as well as the criteria for one or more core ideas and cross-cutting concepts.

An organic chemistry prompt with the potential of eliciting evidence of 3D learning can be seen in Figure 4. This assessment item was given on one of the exams we administered online this past spring. This item, like many that populate organic chemistry assessments, is situated in the context of a reaction system. Here, the phenomenon is the relative rate of two reactions that result in formation of the same product. In part I, students are to represent the changes in system energy that occur on the path from reactants to products in system **A**. From this representation alone, it is impossible to know whether students have vested their drawn potential energy surface with meaning. Indeed, it has been reported that students very often struggle to connect these representations to phenomena in a meaningful way.[51,52] Part II is an attempt to elicit evidence that students understand the conceptual basis for the drawing they constructed in part I. Here, students are asked to use their potential energy surface to inform construction of an explanation for the observed rate differential between system **A** and system **B**. Thus, this task has the potential to engage students in the scientific practice of "developing and using models" as they weave together the knowledge elements clustered under the core ideas of "energy" and "electrostatic and bonding interactions". It is important to point out that items which fulfill 3D-LAP criteria for "potentially eliciting evidence of 3D learning" (such as the item given in Figure 4) are not guaranteed to elicit this evidence. The 3D-LAP provides a minimum bar an item must surmount to have any hope of prompting students to leverage big ideas to make sense of phenomena. However, as past studies have shown,[11] student responses are powerfully affected by prompt structure. Thus, realizing the potential of 3D assessment items may require one to carefully calibrate item scaffolding.

It is tempting to reduce authoring 3D assessments down to adding the word "explain" to existing items. However, one can easily write an item including the word "explain" that does not have the potential to elicit evidence of students clarifying the cause of a phenomenon using core ideas. For example, in the assessment item shown in Figure 5 (taken from exam 1 given in Spring 2017), students are tasked with drawing two representations related to a phenomenon, a curved arrow mechanism and a potential energy surface. They are asked to subsequently use this PE surface to "explain what energy difference is responsible for this outcome". The answer to this question is straightforward: the relative magnitude of the activation energies required to form *para*-, *ortho*-, and *meta*-arenium cations is responsible for the product distribution observed under the reaction conditions given. From this answer, one cannot know anything about why the *para*-arenium cation is lower in energy than the other alternatives. Indeed, the prompt in Figure 5 does not request an explicit link between the claim and any reasoning based upon core ideas. Thus, this item is unlikely to elicit evidence that students understand the origin of the energy differences in arenium cation intermediates or why the relative energy difference controls the reaction outcome. When authoring a task intended to be 3D, it is worth considering what sort of response would earn full credit on that task. If a description of what is happening without invoking a core idea would address the expectations of a prompt, that prompt cannot be 3D.

We suspect that our students' ability to successfully make sense of chemical phenomena on remotely administered assessments may be due, in part, to the consistent message they received about "what counts" from all course components. Accordingly, if instructors wish to prioritize 3D learning in their online organic chemistry course, we recommend coherent integration of many opportunities for molecular-level sense-making on both high- and low-stakes assessments. These opportunities should build on each other in order that students are figuring out the cause for ever more complex phenomena as they progress throughout the semester. To assist instructors with constructing opportunities for students to engage in 3D performances, we provide all homework, discussion, and exam prompts given this past spring in the Supporting Information of this article. If readers would value the analogous resources for the first-semester course, they are encouraged to contact the corresponding author. We are also happy to share recorded lectures if they would be of use to the reader.

IMPLICATIONS FOR RESEARCH

This study examined how the emergency pivot to remote instruction affected student outcomes by comparing the distribution of students who scored above, at, or below their predicted score on the online final exam to historic norms. This sort of analysis should be tractable for many instructors and, so long as exams have similar emphases year-to-year, supports inferences as to how changing instructional and assessment modality impacted achievement. However, our study does not enable comparison between our assessment approach and other alternatives (e.g., administration of more frequent, lower stakes assessments emphasizing 3D tasks). Future work should examine how assessment strategy affects student achievement, with particular emphasis on the achievement of marginalized groups. Relatedly, little is known about the characteristics of equitable 3D assessment prompts as well as whether those characteristics change depending on assessment modality. Future work should investigate how instructors can promote equitable engagement of learners in both online and in-person enactments of organic chemistry.

LIMITATIONS

As is the case with all studies, this work has limitations that merit mentioning. Our data corpus relates only to student achievement on course exams—we can say nothing about how student anxiety and other affective constructs varied as a consequence of the ongoing pandemic or our particular assessment strategy. Likewise, student responses to a prompt represent activation of knowledge elements in that moment and may or may not signify a stable pattern of resource activation. Student participants in this work were enrolled in a second-semester organic chemistry course. Virtually all had previously enrolled in an in-person enactment of organic chemistry 1, and all experienced over half of organic chemistry 2 in-person. Thus, the experience of these students may differ substantially from students who enroll in an organic chemistry course taught wholly online. We cannot predict how (or whether) these differences will impact student engagement with 3D tasks. Finally, all of our data derive from a single institution which is, by most metrics, extremely privileged. We are fortunate to have a dedicated educational technologist and many highly qualified individuals that may be enlisted to help grade. Further, our course was structured in a manner that made the pivot to remote instruction fairly straightforward—recorded lectures, problem sets, discussion activities, and practice assessments were hosted on the course webpage from the beginning of the semester. Additionally, drafts of all high-stakes assessments were written before the start of the spring semester. It is likely that compiling these resources on-the-fly during an emergency pivot online would require significant effort. Accordingly, we have no evidence that our findings as to the feasibility of assessing 3D learning online and student response to those assessments will generalize beyond our local context.

CONCLUSIONS

In this contribution, we have demonstrated that it is possible to focus assessment in an online organic chemistry learning environment on making sense of phenomena. Assessing student engagement in authentic intellectual "heavy lifting" did not meaningfully increase the burden on instructors and TAs relative to the norms of in-person enactments in our institutional context. That is, the time required to prepare, administer and grade exams was essentially unchanged after in-person classes were shuttered. This would likely not be true had the instructors been required to generate more course resources de novo upon the switch to remote instruction. Students' scores on the final exam this past spring did not differ considerably from what would be predicted by their success on exams administered in-person. Additionally, the distribution of final exam scores earned this spring are remarkably consistent with historic norms. Significant differences in exam 3 score distributions were found in our data set, but these were likely a function of anomalously difficult exams (e.g., exam 3 given during Spring 2019) rather than the pivot to remote instruction. In summary, it appears assessing 3D learning is feasible in an online learning environment that consistently emphasizes sensemaking on homework and exams, and that this emphasis need not negatively affect student outcomes or the burden on the course instructional team.

Given that assessing engagement in making sense of chemical phenomena is possible in online learning environments, we would argue that a *substantial* amount of points in chemistry coursework taught remotely should be dedicated to 3D performances. The use of atomic/molecular models to explain how and why things happen is the intellectual core of chemistry and, we would argue, should be central to chemistry coursework. Focusing instruction and assessment solely (or largely) on recall of facts and performance of algorithms should not be acceptable in any chemistry learning environment, whether that environment be enacted in-person or online. Students engaged in disconnected skills will doubtless emerge from their chemistry courses thinking that the discipline consists of a jumble of seemingly random tasks. We prefer a paradigm in which students have the potential to emerge from chemistry courses with an appreciation of the prodigious predictive and explanatory model of atomic/molecular ways of thinking.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.0c00757.

Overview of Supporting Information contents (PDF, DOCX)
Problem sets (PDF)
Key for problem sets (PDF)
Discussion activities (PDF)
Key for discussion activities (PDF)
Spring 2020 exam forms (PDF)
Key for Spring 2020 exams (PDF)
Spring 2019 exam forms (PDF)
Spring 2018 exam forms (PDF)
Spring 2017 exam forms (PDF)
Spring 2016 exam forms (PDF)
3D LAP codes for exam items (XLSX)
3D LAP codes for problem sets (XLSX)
3D LAP codes for discussion sets (XLSX)
Grading spreadsheet used to generate grade reports and email students exam pdfs (ZIP)
Grading spreadsheet user guide (PDF)
Data file containing all deidentified student data analyzed for this article (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Ryan L. Stowe − *Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States;* orcid.org/0000-0002-5548-495X; Email: rstowe@chem.wisc.edu

Authors

Brian J. Esselman − *Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States;* orcid.org/0000-0002-9385-8078

Vanessa R. Ralph − *Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States*

Aubrey J. Ellison − *Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States*

Jeffrey D. Martell − *Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States*

Kimberly S. DeGlopper − *Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States*

Cara E. Schwarz − *Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jchemed.0c00757

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Support for this work was provided by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin—Madison with funding from the Wisconsin Alumni Research Foundation.

■ REFERENCES

(1) Tallent-Runnels, M. K.; Thomas, J. A.; Lan, W. Y.; Cooper, S.; Ahern, T. C.; Shaw, S. M.; Liu, X. Teaching Courses Online: A Review of the Research. *Review of Educational Research* **2006**, *76* (1), 93−135.

(2) Gaytan, J.; McEwen, B. C. Effective Online Instructional and Assessment Strategies. *American Journal of Distance Education* **2007**, *21* (3), 117−132.

(3) Momsen, J.; Offerdahl, E.; Kryjevskaia, M.; Montplaisir, L.; Anderson, E.; Grosz, N. Using Assessments to Investigate and Compare the Nature of Learning in Undergraduate Science Courses. *CBE Life Sci. Educ.* **2013**, *12* (2), 239−249.

(4) Scouller, K. The Influence of Assessment Method on Students' Learning Approaches: Multiple Choice Question Examination versus Assignment Essay. *High. Educ.* **1998**, *35* (4), 453−472.

(5) Scouller, K. M.; Prosser, M. Students' Experiences in Studying for Multiple Choice Question Examinations. *Stud. High. Educ.* **1994**, *19* (3), 267−279.

(6) Snyder, M. D. B. R. *The Hidden Curriculum*, 1st ed.; Knopf: New York, 1971.

(7) Entwistle, N. J. Approaches to Learning and Perceptions of the Learning Environment. *High. Educ.* **1991**, *22* (3), 201−204.

(8) Crooks, T. J. The Impact of Classroom Evaluation Practices on Students. *Rev. Educ. Res.* **1988**, *58* (4), 438−481.

(9) The National Research Council. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*; National Academies Press: Washington, DC, 2012.

(10) National Research Council. *Knowing What Students Know: The Science and Design of Educational Assessment*; National Academies Press: Washington, DC, 2001.

(11) Cooper, M. M.; Kouyoumdjian, H.; Underwood, S. M. Investigating Students' Reasoning about Acid−Base Reactions. *J. Chem. Educ.* **2016**, *93* (10), 1703−1712.

(12) Stowe, R. L.; Cooper, M. M. Arguing from Spectroscopic Evidence. *J. Chem. Educ.* **2019**, *96* (10), 2072−2085.

(13) Stowe, R. L.; Herrington, D. G.; McKay, R. L.; Cooper, M. M. The Impact of Core-Idea Centered Instruction on High School Students' Understanding of Structure−Property Relationships. *J. Chem. Educ.* **2019**, *96* (7), 1327−1340.

(14) Stowe, R. L.; Cooper, M. M. Assessment in Chemistry Education. *Isr. J. Chem.* **2019**, *59*, 598−607.

(15) Cooper, M. M.; Posey, L. A.; Underwood, S. M. Core Ideas and Topics: Building Up or Drilling Down? *J. Chem. Educ.* **2017**, *94* (5), 541−548.

(16) Stowe, R. L.; Cooper, M. M. Practice What We Preach: Assessing "Critical Thinking" in Organic Chemistry. *J. Chem. Educ.* **2017**, *94* (12), 1852−1859.

(17) Cooper, M. M. The Crosscutting Concepts: Critical Component or "Third Wheel" of Three-Dimensional Learning? *J. Chem. Educ.* **2020**, *97* (4), 903−909.

(18) diSessa, A. A. Toward an Epistemology of Physics. *Cognition Instruct.* **1993**, *10* (2-3), 105−225.

(19) Hammer, D. Student Resources for Learning Introductory Physics. *Am. J. Phys.* **2000**, *68* (S1), S52−S59.

(20) Hammer, D.; Elby, A. On the Form of a Personal Epistemology. In *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing*; Hofer, B. K., Pintrich, P. R., Eds.; Lawrence Erlbaum Associates: Mahwah, NJ, 2002; pp 169−190.

(21) Russ, R. S.; Scherr, R. E.; Hammer, D.; Mikeska, J. Recognizing Mechanistic Reasoning in Student Scientific Inquiry: A Framework for Discourse Analysis Developed from Philosophy of Science. *Sci. Educ.* **2008**, *92* (3), 499−525.

(22) Hammer, D.; Elby, A.; Scherr, R. E.; Redish, E. F. Resources, Framing, and Transfer. In *Transfer of Learning from a Modern Multidisciplinary Perspective*; Mestre, J. P., Ed.; IAP: Greenwich, CT, 2005; pp 89−119.

(23) Bing, T. J.; Redish, E. F. Analyzing Problem Solving Using Math in Physics: Epistemological Framing via Warrants. *Phys. Rev. ST Phys. Educ. Res.* **2009**, *5* (2), 020108.

(24) Hutchison, P.; Hammer, D. Attending to Student Epistemological Framing in a Science Classroom. *Sci. Educ.* **2009**, *94* (3), 506−524.

(25) Scherr, R. E.; Hammer, D. Student Behavior and Epistemological Framing: Examples from Collaborative Active-Learning Activities in Physics. *Cognition and Instruction* **2009**, *27* (2), 147−174.

(26) Berland, L. K.; Reiser, B. J. Making Sense of Argumentation and Explanation. *Sci. Educ.* **2009**, *93* (1), 26−55.

(27) Crowder, E. M. Gestures at Work in Sense-Making Science Talk. *Journal of the Learning Sciences* **1996**, *5* (3), 173−208.

(28) Ford, M. J. A Dialogic Account of Sense-Making in Scientific Argumentation and Reasoning. *Cognition Instruct.* **2012**, *30* (3), 207−245.

(29) Goffman, E. *Frame Analysis: An Essay on the Organization of Experience*; Harvard University Press: Cambridge, MA, 1974.

(30) Odden, T. O. B.; Russ, R. S. Vexing Questions That Sustain Sensemaking. *International Journal of Science Education* **2019**, *41* (8), 1052−1070.

(31) Laverty, J. T.; Underwood, S. M.; Matz, R. L.; Posey, L. A.; Carmel, J. H.; Caballero, M. D.; Fata-Hartley, C. L.; Ebert-May, D.; Jardeleza, S. E.; Cooper, M. M. Characterizing College Science Assessments: The Three-Dimensional Learning Assessment Protocol. *PLoS One* **2016**, *11* (9), e0162333.

(32) Matz, R. L.; Fata-Hartley, C. L.; Posey, L. A.; Laverty, J. T.; Underwood, S. M.; Carmel, J. H.; Herrington, D. G.; Stowe, R. L.; Caballero, M. D.; Ebert-May, D.; Cooper, M. M. Evaluating the Extent of a Large-Scale Transformation in Gateway Science Courses. *Science Advances* **2018**, *4* (10), eaau0554.

(33) Underwood, S. M.; Posey, L. A.; Herrington, D. G.; Carmel, J. H.; Cooper, M. M. Adapting Assessment Tasks To Support Three-Dimensional Learning. *J. Chem. Educ.* **2018**, *95*, 207−217.

(34) Grove, N. P.; Cooper, M. M.; Rush, K. M. Decorating with Arrows: Toward the Development of Representational Competence in Organic Chemistry. *J. Chem. Educ.* **2012**, *89* (7), 844−849.

(35) Bhattacharyya, G.; Bodner, G. M. It Gets Me to the Product": How Students Propose Organic Mechanisms. *J. Chem. Educ.* **2005**, *82* (9), 1402−1407.

(36) Kraft, A.; Strickland, A. M.; Bhattacharyya, G. Reasonable Reasoning: Multi-Variate Problem-Solving in Organic Chemistry. *Chem. Educ. Res. Pract.* **2010**, *11* (4), 281−292.

(37) KewalRamani, A.; Zhang, J.; Wang, X.; Rathbun, A.; Corcoran, L.; Diliberti, M.; Zhang, J. *Student Access to Digital Learning Resources Outside of the Classroom*; NCES 2017-098; National Center for Education Statistics: Washington, DC, 2018.

(38) National Research Council. *Developing Assessments for the Next Generation Science Standards*; National Academies Press: Washington, DC, 2013.

(39) Bryfczynski, S. P. BeSocratic: An Intelligent Tutoring System for the Recognition, Evaluation, and Analysis of Free-Form Student Input. Doctoral Dissertation, Clemson University, 2012.

(40) Cooper, M. M.; Underwood, S. M.; Bryfczynski, S. P.; Klymkowsky, M. W. A Short History of the Use of Technology To Model and Analyze Student Data for Teaching and Research. *Tools of Chemistry Education Research*; ACS Symposium Series; American Chemical Society: Washington, DC, 2014; Vol. *1166*, pp 219−239.

(41) Loudon, M.; Parise, J. *Organic Chemistry*, 6th ed.; W.H. Freeman and Company: New York, 2016.

(42) Cooper, M. M.; Stowe, R. L.; Crandell, O. M.; Klymkowsky, M. W. Organic Chemistry, Life, the Universe and Everything (OCLUE): A Transformed Organic Chemistry Curriculum. *J. Chem. Educ.* **2019**, *96* (9), 1858−1872.

(43) Cooper, M.; Klymkowsky, M. Chemistry, Life, the Universe, and Everything: A New Approach to General Chemistry, and a Model for Curriculum Reform. *J. Chem. Educ.* **2013**, *90* (9), 1116−1122.

(44) Sevian, H.; Talanquer, V. Rethinking Chemistry: A Learning Progression on Chemical Thinking. *Chem. Educ. Res. Pract.* **2014**, *15* (1), 10−23.

(45) IBM Corp. *SPSS Statistics for Mac*; IBM Corp.: Armonk, NY, 2017.

(46) Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2 ed..; Routledge: Hillsdale, NJ, 1988.

(47) Cooper, M. M.; Underwood, S. M.; Hilley, C. Z.; Klymkowsky, M. W. Development and Assessment of a Molecular Structure and Properties Learning Progression. *J. Chem. Educ.* **2012**, *89* (11), 1351−1357.

(48) Stowe, R. L.; Herrington, D. G.; McKay, R. L.; Cooper, M. M. Adapting a Core-Idea Centered Undergraduate General Chemistry Curriculum for Use in High School. *J. Chem. Educ.* **2019**, *96* (7), 1318−1326.

(49) Agresti, A. Inference for Two-Way Contingeny Tables. *Categorical Data Analysis*; Wiley: Hoboken, NJ, 2013; pp 69−112.

(50) Field, A. *Discovering Statistics Using IBM SPSS Statistics: North American Edition*; SAGE, 2017.

(51) Popova, M.; Bretz, S. L. Organic Chemistry Students' Challenges with Coherence Formation between Reactions and Reaction Coordinate Diagrams. *Chem. Educ. Res. Pract.* **2018**, *19* (3), 732−745.

(52) Popova, M.; Bretz, S. L. It's Only the Major Product That We Care About in Organic Chemistry": An Analysis of Students' Annotations of Reaction Coordinate Diagrams. *J. Chem. Educ.* **2018**, *95* (7), 1086−1093.

https://dx.doi.org/10.1021/acs.jchemed.0c00757