JOURNAL OF CHEMICAL EDUCATION

pubs.acs.org/jchemeduc Article

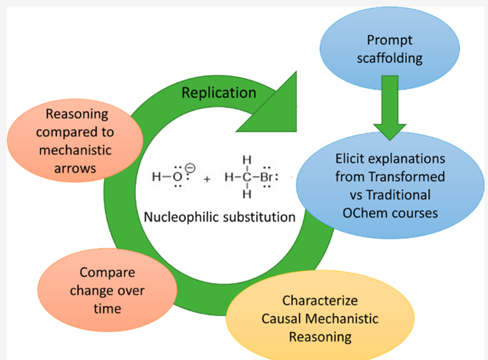
Arrows on the Page Are Not a Good Gauge: Evidence for the Importance of Causal Mechanistic Explanations about Nucleophilic Substitution in Organic Chemistry

Olivia M. Crandell,* Macy A. Lockhart, and Melanie M. Cooper

Cite This: *J. Chem. Educ.* 2020, 97, 313−327 Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: This three-year study builds on prior work analyzing students' causal mechanistic explanations about acid−base reactions. Here we extend that work to characterize and investigate how students construct causal mechanistic explanations of simple nucleophilic substitution reactions. After an initial pilot study, we adopted a modified version of the original task prompt which was used in two subsequent years to compare responses from students enrolled in a transformed organic chemistry course (Organic Chemistry, Life, the Universe and Everything) and a traditional organic chemistry course. Student responses were sampled in the middle of the first semester organic chemistry, just after they had learned the material, and once again at the end of the course, to identify how responses changed over time. Our findings from this study include (1) eliciting causal mechanistic explanations requires careful scaffolding to activate productive elements of a response, and (2) students from a transformed course are more likely to construct a causal mechanistic explanation at the end of Organic Chemistry 2 than students from the control group, suggesting that if students are to retain the use of valuable knowledge, this must be supported by instruction and course expectations.



KEYWORDS: *Organic Chemistry, Chemical Education Research, Mechanisms of Reactions, Second-Year Undergraduate*

FEATURE: Chemical Education Research

■ INTRODUCTION

Since the release of Morrison and Boyd's Organic Chemistry in 1959,[1] the use of curved arrows to denote electron flow (that is to show the electron pushing mechanism) has been emphasized in most organic chemistry courses. In a national survey of organic faculty, organic chemistry experts agreed that mechanistic reasoning using the electron-pushing formalism should "conform to patterns established by known mechanisms and reflect an understanding of partial or formal charges that may exist among the reactants and intermediates."[2] There are numerous studies that have identified undergraduate and graduate student difficulties using the electron-pushing formalism in this expert-like way.[3−7] However, students may not be demonstrating an understanding of structure−property relationships or electrostatic attractions when they use the electron-pushing formalism but rather they are drawing arrows to "get them to the product"[6] or "decorating with arrows"[3] after drawing a memorized product.

For example, in a study on how students draw mechanisms, Grove et al. determined that only about 50% of students used mechanistic arrows to predict products, and of the students who did draw mechanistic arrows, 20% of them drew the arrows after predicting a product instead of using mechanistic arrows as a tool to guide their prediction.[3] In a study on how graduate students use mechanisms, Bhattacharyya found that some students struggled to explain their mechanistic arrow use in terms of electrostatic attractions and often resorted to memorized patterns to draw the mechanism.[6] Additionally, Flynn et al. have shown that some students tend to use surface features to predict patterns of reactivity rather than thinking about mechanistic processes,[8] and Graulich et al. found that many students can still be successful at matching reagents to a given transformation even when relying on surface features.[9] In another study on students' understanding of alkene mechanisms, Graulich et al. found that many novices group reactions by the surface features of the starting material, reagents, or functional group in the product rather than by the type of mechanism.[10] These studies investigated pattern recognition and problem-solving strategies via qualitative interviews leading

Received: August 29, 2019
Revised: December 30, 2019
Published: January 17, 2020

to rich descriptions about the understandings for a small set of students.[8−10]

Just as multiple choice assessment items have been shown to overestimate what students know,[11] reproducing a reaction mechanism may also overestimate student understanding of what an organic reaction mechanism actually denotes.[3,6] If we are to make assertions about what students know and are able to do in organic chemistry, we believe it is important to elicit more robust evidence of student thinking beyond asking students to draw arrow-pushing mechanisms or simply to draw a predicted product−in other words, we must elicit student reasoning about how and why reactions happen.

Importance of Scaffolding to Activate Resources

Research on student reasoning in chemistry is typically conducted via student interviews at which the researcher can engage with the student by asking follow-up questions to expand or clarify student thinking. These robust qualitative data sets offer rich insights into student thinking for a small subset of participants but lack power to make broad generalizations about the larger population, nor do they allow for comparisons between student populations. Our goal in this study is to collect and analyze evidence of student reasoning from large numbers of students. To accomplish this requires that we develop appropriately scaffolded task prompts that signal to students the type of reasoning for which we are looking.

There is a fairly extensive literature base of theories and studies on how to better scaffold learning environments.[12−15] However, the term "scaffolding" has come to mean many things including but not limited to student−teacher interactions,[12] written instructional supports,[14] and interactive technology environments.[14,15] Pioneering work by Wood et al. defined scaffolding as "...[a] process that enables a child or novice to solve a problem, carry out a task or achieve a goal which would be beyond his unassisted efforts."[16] Wood et al. identified functions of scaffolding for the interactions between a young child and an adult tutor when the child was tasked with solving a puzzle of wooden blocks. While Wood et al.'s study may seem distant from eliciting chemical explanations from young adults, some of these functions of scaffolding apply to our context, namely (1) reduction of degrees of freedom and (2) marking critical features.[16]

Reducing the degrees of freedom simply means simplifying the task so the learner can recognize the expectations of the task. Wood et al. placed this burden on the tutor but in our case, the written instructions from the task prompt must accomplish this without any additional encouragement from the researcher. Reiser's work with scaffolding in educational technologies and software leveraged this idea when he concluded "...if reasoning is difficult due to complexity or the open-ended nature of the task, then one way to help learners is to use the tool to reduce complexity and choice by providing additional structure to the task."[13]

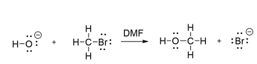
In our previous work, we developed a scaffolded explanation prompt designed to elicit student reasoning about a simple acid−base reaction and have shown that this approach elicits richer responses than simpler, less targeted prompts.[17] In this task we asked students to explain both "what is happening" and then asked them to "explain why" the reaction is happening, and found that this task structure elicited more causal mechanistic responses (defined in the next section) than asking students to explain what and why in the same response box. By separating the prompt into two sections—"what" and "why", students are cued into the fact that "what" and "why" are different.[17]

In another study that takes place in the context of structure−property relationships, Underwood et al. found that asking students to construct an explanation for why one substance had a higher boiling point elicited evidence of more sophisticated understanding than asking students to construct an argument for which substance has a higher relative boiling point.[18] That is, by reducing the degrees of freedom (telling students which substance has the highest boiling point), students were able to produce more robust reasoning.

Marking critical features means that relevant features of the task are emphasized and made clear for the learner.[16] The prompt directs students to discuss at the molecular level the role of each reactant and to explain why the reactants form the products shown. We saw in our previous work that each of these pieces appears to cue students appropriately so that they draw on their knowledge of the molecular level, the activity of each reactant (rather than just discussing one reactant), and why the reaction occurs (rather than just restating that products form). Scaffolds such as these can "[provide] structured work spaces to help learners recognize important goals to pursue."[13]

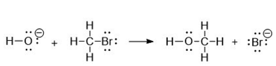
It is particularly necessary to guide student thinking in organic chemistry because not only do they have more knowledge at their disposal, but also organic chemistry students have been found to hold incorrect ideas about acid strength as measured by a concept inventory[19] and interviews[20] and fragmented ideas about structure−property relationships of nucleophiles and electrophiles as elicited from interviews.[21] Therefore, we have chosen to experiment with different prompt wording to better understand the possible ways that organic students' ideas (or intellectual resources) may be sensitive to (or activated by) different prompt wording.[22] The final goal being a prompt that elicits (or activates) causal and mechanistic elements when engaging in explanation (defined in the next section) and to offer students multiple opportunities within the prompt to articulate the desired causal mechanistic explanation (Figure 1). Kraft et al. found student reasoning to be sensitive to the task used to elicit mechanistic problem-solving.[23] Two different tasks were used to elicit students' modes of reasoning: complete a mechanism that was already started or predict products.[23] Students' success varied on these tasks with more successful students being cued to invoke specific prior knowledge about a similar case (case-based reasoning) rather

(A) Original S_N2 Prompt



i) How would you classify this reaction? Please explain why you chose that classification.
ii) Can you describe in full detail what you think is happening on the molecular level for this reaction? Specifically, discuss the role of each reactant.
iii) Using a molecular level explanation, please explain why this reaction occurs? Specifically, why the reactants form the products shown.
iv) For the following reaction, please draw arrows in the BLUE box to indicate how this reaction occurs.
v) Now please explain why you drew your arrows as indicated.

(B) Modified S_N2 Prompt



i) How would you classify this reaction? Please explain why you chose that classification.
ii) Please describe the sequence of events that occur at the molecular level during the reaction shown above.
iii) Please explain why these reactants interact.
iv) For the following reaction, please draw arrows in the BLUE box to indicate how this reaction occurs.
v) Now please explain why you drew your arrows as indicated.

Figure 1. (A) Original S_N2 Prompt structure administered using beSocratic. (B) Modified S_N2 Prompt administered using beSocratic.[33]

https://dx.doi.org/10.1021/acs.jchemed.9b00815
J. Chem. Educ. 2020, 97, 313−327

than being cued to invoke sets of memorized rules (rule-based reasoning).[23]

In summary, if we want students to construct explanations that contain certain components, we must clearly communicate these expectations in the prompt by providing appropriate scaffolding to elicit students' reasoning. We have used the above-mentioned literature on scaffolding to inform the design of our prompts to elicit reasoning. We have alluded to causal mechanistic explanation as our desired type of explanation. In the next section we define causal mechanistic reasoning as it is utilized in our work as well other investigations of reasoning in organic chemistry more broadly.

Causal Mechanistic Reasoning in Organic Chemistry

Explanation is a central practice to the understanding of science.[24,25] *The Framework for K-12 Science Education* explicitly states "the goal of science is the construction of theories that can provide explanatory accounts of features of the world."[25] The IES report titled *Organizing Instruction and Study to Improve Student Learning: IES Practice Guide* cites the construction of deep explanations of phenomena as one of its chief pedagogical recommendations with strong evidence to support the importance of "[asking] questions that elicit explanations, such as those with the following question stems: why, what caused X, how did X occur..."[24] There is also evidence to support the importance of asking students to ask deep-level questions to themselves and construct explanations.[26] Asking students to construct deep explanations about *why* and *how* is the impetus on which causal mechanistic reasoning is built.

To define causal mechanistic reasoning, we draw from Russ et al.'s review of mechanistic reasoning where they conclude "...that mechanistic reasoning involves describing how the particular components of a system give rise to its behavior."[27] These authors modify a framework originally posed by Machamer, Darden, and Craver[28] and apply it to analysis of student mechanistic reasoning in the context of classroom discussion. Russ et al.'s framework identifies several components of reasoning for which the most basic components are (1) identifying entities and (2) identifying activities of those entities. These two components are foundational for more complicated reasoning to take place (e.g., reasoning about a string of events that occurred to bring a phenomenon about or making a prediction about what will happen next).

For our work in the context of a chemical reaction, we have defined causal mechanistic reasoning as (1) a discussion of the electrostatic attraction between electron-rich and electron-deficient regions (the underlying causal factors) and (2) a step-by-step account of the activities of the underlying entities responsible for the mechanism: that is the movement of electrons during bond breaking and formation. We acknowledge that there are other causal factors that we have not specifically prompted for at this time. Although Russ et al. argue that mechanistic reasoning is inherently causal[27] and therefore use the term mechanistic reasoning as an encompassing definition for the process of a cause bringing about an effect, we emphasize both of these elements because we have evidence that students can engage in one aspect of causal mechanistic reasoning without engaging in the other.[17,29] That is, a response can be mechanistic only or causal only. For example, organic chemistry students tended to provide explanations of the reaction of NH_3 with the Lewis acid, BF_3, in which they discussed electron movement but did not include a discussion of electrostatic attraction for the reaction—that is, why the electrons move in this way.[17] Ideally, we want students to construct explanations for chemical reactions that include both an electrostatic cause *and* an account of electron movement. We believe that this emphasis on causality is important; for example, it may support students as they draw appropriate electron pushing mechanisms. In our previous work, general chemistry students who engaged in causal mechanistic reasoning about a simple acid−base reaction were more successful at drawing correct mechanistic arrows.[17]

In the context of organic chemistry, other researchers have used somewhat different approaches to defining mechanistic reasoning. Sevian and Talanquer have posed four modes of reasoning that they have used to characterize various levels of complexity in student responses.[30] The lowest level modes being descriptive in instances for which there is no mechanism or cause and then relational in instances for which no mechanisms are discussed but "properties and behaviors are established but not explained or justified." The next mode of reasoning being linear causal in which "relevant direct interactions between entities are invoked. . .but phenomena [are] reduced to the result of actions of a single entity." Finally, multicomponent reasoning being the most sophisticated in which the mechanism and cause are included but also "...effects of several variables are considered and weighed."

Caspari et al. have utilized these modes of complexity to analyze student responses.[31] Caspari et al. define mechanistic reasoning to mean "comparative reasoning about cause−effect relationships between explicit structural differences and structural and energic changes occurring in a mechanistic step." Their definition of mechanistic reasoning incorporates a comparison between contrasting cases rather than reasoning about a single phenomenon in isolation.[31] Flynn et al. have similarly incorporated these modes of reasoning in their analysis of causal mechanistic arguments.[32] They define a descriptive mechanism as "the elementary steps that comprise an overall reaction" and the causal mechanism as "the reasons or cause behind a phenomenon or process."[32] Our goal in this work is to characterize the nature of an explanation in terms of the causal elements (electrostatic interactions) and mechanistic elements (explicit electron movement) invoked to discuss a simple S_N2 reaction and relate this reasoning to their mechanistic arrow use.

Research Questions

This study is guided by these research questions:

1. How does the nature of the prompt affect student responses about a simple nucleophilic substitution reaction?
2. How does the type of organic chemistry course affect student ability to engage in causal mechanistic reasoning?
3. How does the reasoning about a reaction change over the course of two semesters?
4. How do student written explanations of reaction type compare to their mechanistic arrow drawings?

■ METHODS

Design of Assessment Tasks

The design of the assessment tasks evolved over the first year of this three-year study. In the first year of the study (Year 1), we piloted two different prompt structures (called the Original S_N2

Prompt and Modified S_N2 Prompt). In years two and three, only the Modified S_N2 Prompt was administered.

Original S_N2 Prompt

Because a simple S_N2 reaction can be considered a type of Lewis acid—base reaction, we began by adopting the same prompt structure as the one used in our acid—base work[17,29] in which students were asked to (i) classify the reaction and explain their reasoning, (ii) describe what is happening on the molecular level, (iii) please explain why the reaction occurs using a molecular level explanation, and (iv) draw arrows onto predrawn Lewis structures to afford given products (Figure 1A). We also added a space for students to (v) explain why they drew their arrows as indicated to give even more opportunities to activate the relevant resources.[22] The original acid—base reaction was replaced with the reaction of methyl bromide (CH_3Br) with hydroxide (OH^-). A "textbook" example of an S_N2 reaction—a methyl halide substrate with a good leaving group, and a strong unhindered nucleophile that only undergoes S_N2 reactions. Lewis structures of the reactants and products were included as shown because we wanted students to explain a simple, clear-cut reaction where the products are clearly given rather than engage in argumentation about whether the reaction should be an S_N2 or an S_N1 by giving them a set of ambiguous reaction conditions that could be argued to be an S_N2, S_N1, E2, or E2. This prompt structure will be referred to as the Original S_N2 Prompt and was administered at the end of Spring 2017 to students completing organic chemistry 2 (OC2).

Modified S_N2 Prompt

We modified the Original S_N2 Prompt slightly to investigate how the wording of the prompt might be activating different resources in organic chemistry students. We thought that the original wording of "Specifically, why the reactants form the products shown" might not activate resources related to electrostatic interaction and rather might activate reasons for why the products are more stable or other teleological reasons for why the products "want" to form. Thus, the "describe what" phrasing was modified to "Please describe the sequence of events that occur at the molecular level during the reaction shown above" and "explain why" was changed to "Please explain why these reactants interact" (Figure 1B). We recognize that there are other causal factors; however, we did not specifically prompt for them at this time. We made this specific choice to prompt for students' resources concerning the core idea of electrostatic interactions.[34] This prompt structure will be referred to as the Modified S_N2 Prompt.

Both versions (original and modified S_N2 prompts) were administered at the end of Spring 2017 to students completing OC2 (Year 1—Time Point 2) (see Figures 1 and 2). On the basis of our analysis (shown in Results and Discussion below), the Modified S_N2 Prompt shown in Figure 1B seemed to elicit richer responses. Therefore, we continued with the modified S_N2 Prompt which was administered twice the next year: in the middle of OC1 in Fall 2017 (Year 2, Time Point 1) just after students learned about nucleophilic substitution and then again at the end of OC2 in Spring 2018 (Year 2, Time Point 2). It was administered again at the end of OC2 in Spring 2019 (Year 3, Time Point 2).

Student Participants

Course Contexts. Students in this study were selected on the basis of their enrollment in two types of organic chemistry

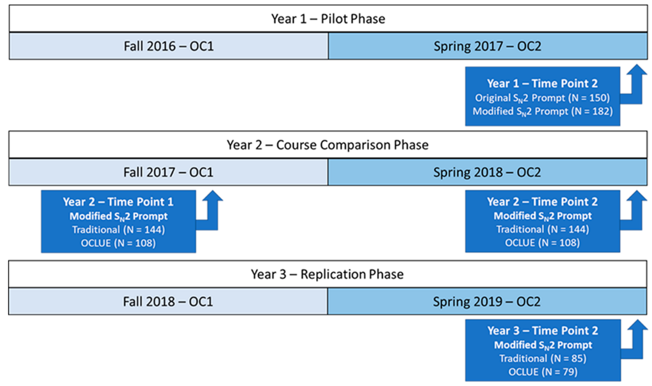


Figure 2. Summary of data collections over the three years of this study.

courses: Traditional OC (referred to as Traditional students) and Transformed OC (referred to as OCLUE students). Both courses were taught at a large, research-intensive Midwestern university. We have previously reported on a transformed organic chemistry course Organic Chemistry, Life, the Universe, and Everything (OCLUE).[35] The course emphasizes connecting student knowledge of reactions and topics to core ideas of chemistry (structure property relationships, electrostatic forces and bonding interactions, stability and change in chemical systems, and energy) in the context of scientific practices.[34,36] In contrast, the traditional organic course curriculum is organized by functional group and topic. OCLUE also requires students to use their knowledge to make predictions about phenomena and OCLUE formative and summative assessments are designed to elicit evidence of student reasoning as well as the more traditional tasks such as mechanism construction or predicting products.[35,37] OCLUE students are challenged to construct explanations for phenomena such as relative nucleophile strength, relative proton acidity, and kinetic and thermodynamic control in terms of atomic and molecular structure/properties and other core ideas, whereas the assessments in the traditional course ask students to provide missing reactants, products or reagents or draw a mechanism for a given reaction without any explanation.[37] Each instructor wrote their own exams for their course; there were no common exams.

Assessments send strong messages about what is most important to know and what students should be able to do, and we know that students tend to value (in terms of studying) what will appear on their exams.[38] OCLUE students practice constructing explanations for phenomena on their weekly homework, in weekly TA led recitation group work sessions and in lectures so they are prepared to do so on their exams. Homework and recitation work are included in the OCLUE syllabus as 30% of their course grade. Traditional organic students are not given any credit for completing homework or practice problems. Rather, the traditional course grade is composed of summative assessments in the form of exams and quizzes. In an analysis of three years of OCLUE exams using the three-dimensional learning assessment protocol instrument,[36] between 25 and 50% of exam points were dedicated to questions that required students to use their knowledge of core ideas to construct an explanation or argument, reason about a model, or analyze and interpret data. Traditional organic exams did not offer any opportunities for students to engage in such scientific practices.

During this study, three lecture sections of OC1 were taught each fall and three lecture sections of OC2 were taught each spring by various instructors and by varying course type (i.e., Traditional OC or OCLUE OC). Each section has a maximum enrollment of 360 students. At the time of course enrollment, students do not know the instructor or type of course because enrollment often occurs a year in advance before teaching assignments have been decided. It is also possible, because of scheduling restraints, for students to take the first semester of Traditional OC1 followed by the second semester of OCLUE OC2 or vice versa. Effects of students switching from one course type to the other are not discussed in this paper but will be reported on later. All the instructors from the different course types agreed on the "topics" (e.g., reactivity of certain functional groups, skills, and content) that would be covered in the first semester so students who switched between course types would have been exposed to the same content and skills albeit with different emphases and organization.[35] All student participants were informed of their rights as research participants in accordance with our institutions' IRB. Data for this study was collected over six academic semesters starting in Fall 2016 and ending in Spring 2019. The Fall 2016 to Spring 2017 semesters will be referred to as Year 1: Pilot Phase. The Fall 2017 to Spring 2018 semesters will be referred to as Year 2: Comparing Course Types. Data was collected in Spring 2019 as a replication study of Year 2.

Year 1—Pilot Phase. In Spring 2017, two sections of Traditional OC2 were taught by two organic professors both possessing 10 or more years of teaching experience. One section of OCLUE OC2 was taught in Spring 2017 by the coauthor of the OCLUE curriculum (M.M.C.). Students who took either Traditional OC2 or OCLUE OC2 (~960 students) in Spring 2017 were randomly assigned to one of the nucleophilic substitution activity prompts (see Figure 1) as discussed earlier. Since the purpose of piloting these versions was to identify which wording and prompt structure elicited the most causal mechanistic explanations for the reaction of CH_3Br and OH^-, we did not separate students based on their organic course type nor will we make any claims about course enrollment for Year 1.

The "Original S_N2 Prompt" was administered to 298 students across both course types with an 87% response rate. The Modified S_N2 Prompt was administered to 317 students across both course types with a 91% response rate. Once the data were collected, we removed students for whom we could not obtain the following information: a reported general chemistry 1 course grade, a general chemistry 2 course grade, an organic chemistry 1 course grade, an organic chemistry 2 course grade, and an ACT score or SAT score. This left 150 responses for the Original S_N2 Prompt and 182 responses for the Modified S_N2 Prompt (Figure 1).

To ensure that we could reasonably compare responses across both versions, we compared the students in each group on various academic and demographic measures. A series of Mann−Whitney U-tests were performed and effect sizes reported[39] comparing one cohort to the other on ACT or SAT score, GC1 course grade, GC2 course grade, GPA prior to spring 2017, OC1 grade, and OC2 grade. The only observed difference was in OC2 course grade (mean of 3.51 for OC2 course grade for the Original S_N2 Prompt compared to 3.32 for the Modified S_N2 Prompt $U = 11970.0$, $z = -2.118$, $p = 0.034$, $r = 0.116$ small effect size). The cohorts were also compared on gender and major and no differences were found (see full statistical output in the Supporting Information, Tables S1−S4).

Year 2−Course Comparison Phase. In Fall 2017, one section of Traditional OC1 was taught by a professor who had over 10 years of teaching experience. Two sections of OCLUE OC1 were taught in Fall 2017. The second author (M.M.C.) taught one and oversaw a postdoctoral researcher with no teaching experience who taught the other. Both OCLUE sections used the same instructional materials and assessments.

In Spring 2018, two sections of Traditional OC2 and one section of OCLUE OC2 were offered. Therefore, there were many students who changed course type as they moved from OC1 to OC2. Considering only those students who did not switch course type, we identified students who had completed both Year 2 data collections (Time Point 1 and Time Point 2), and for whom we had obtained the same academic and demographic measures used in Year 1 comparisons. Students who met these criteria were retained for our Year 2 sample. Table S5 provides a summary of this selection process for Year 2 participants. There were 144 students who took a Traditional course for OC1 and OC2 in Year 2 and will be referred to as Year 2−Traditional. Similarly, there were 108 students who took OCLUE for OC1 and OC2. These students will be referred to as Year 2−OCLUE. Using a Mann−Whitney comparison, we found a difference in OC2 course grade between Year 2−Traditional and Year 2−OCLUE. The mean OC2 course grade was slightly higher for Year 2−Traditional than for Year 2−OCLUE (Traditional = 3.43, OCLUE = 3.23, $U = 6587.5$, $z = -2.205$, $p = 0.027$, $r = 0.139$ small effect size). No differences were found in the comparison of ACT, general chemistry course grades, OC1 course grade, gender distribution, or major (see Supporting Information S5−S8).

Year 3−Replication Phase. A replication study was performed by administering the Modified S_N2 Prompt at the end of OC2. We identified those students who had the same course type for OC1 and OC2 and completed the Year 3−Time Point 2 data collection for our Year 3 sample. The Year 3−Traditional ($N = 85$) cohort averaged a slightly lower OC1 course grade (small effect size) and averaged a higher OC2 course grade (medium effect size) than the Year 3−OCLUE cohort ($N = 79$). Finally, we compared these cohorts across years and found Year 2−OCLUE to have a higher ACT (small effect size), Year 2−Traditional to have a higher OC1 course grade (medium effect size), and Year 3−Traditional to have a higher OC2 course grade (medium effect size). All statistical analyses were performed in SPSS and are provided in the Supporting Information (S9−S14).

Data Collection

The data reported in this study were collected on an online homework platform beSocratic[33] and take the form of students' typed explanations and drawn mechanistic arrows. The beSocratic system allows students to type responses to questions and draw mechanistic arrows, chemical structures, and drawings using a mouse, trackpad, or pen and touchscreen. This system also allows us to replay student responses so that we can determine the sequence of arrows drawn by students.

Year 1. At the end of Spring 2017, all three sections of OC2 participated in the study. Students in Traditional OC2 and OCLUE OC2 were randomly assigned one of the two beSocratic prompts shown in Figure 1. Both versions of the activity were administered in the 14th week (that is, near the end) of OC2. This round of data collection is referred to as Year

Table 1. Causal Mechanistic Characterization Scheme

	Mutually Exclusive Code with Description	Examples of Related Student Responses
No response (NR)	Student does not provide an answer.	Jessica: "I don't know what to say."
	Explanations are unreadable or incomprehensible.	
	Student does not even attempt to answer.	
Non-normative (NN)	Student provides a non-normative or unrelated explanation.	Emmy: "While oxygen accepted a proton and formed O−H bond and Br leaves."
	Student attributes the mechanism to other types of reactions or other types of macroscopic observations.	Mason: "The Br is attracted to the H."
	The response is discussing incorrect entities and/or incorrect processes.	
Descriptive general (DG)	Student provides a scientifically simplistic description of bond formation and bond breaking.	Calvin: "The nucleophile attacks the electrophile which makes the leaving group leave."
		Phyllis: "First the OH attacks the carbon center and the Br leaves (carbon−bromine bond breaks) this happens in one step."
Descriptive causal (DC)	Student discusses the electrostatic attraction between the species.	Barbara: "These reactants interact because the OH group has a negative charge and is therefore nucleophilic. It wants to attack a carbon center (or something with a positive charge even if its partial) the Br is a good leaving group (better than OH) so OH is able to come in and take its place."
	Student gives evidence that they understand that there is an attraction between the OH^- and the partial positive carbon atom.	
	Students do not need to justify why the carbon is partial positive. They just need to demonstrate an understanding of the intermolecular electrostatic attraction.	Ryan: "The carbon is slightly positive because the bromine is pulling the electrons away from the carbon. The negative oxygen attracts the partially positive carbon and the bromine is pushed off and a new bond is made between the carbon and the oxygen."
Descriptive mechanistic (DM)	Student identifies electrons as the entities responsible for the reaction mechanism and explains their activities that lead to bond formation/bond breaking.	Wanda: "The electrons from the negatively charged OH are going to attack the carbon. This will push off the bromine and the bromine will get the electrons from the bond between the carbon−bromine bond."
	Student gives evidence that they understand that electron movement from source to sink is how the reaction occurs.	Morgan: "The lone pair on OH is forming a bond with carbon."
	Response may only explicitly discuss the movement of the lone pair of electrons on the OH^- or the electrons in the C−Br bond. This is still considered mechanistic.	
Causal mechanistic (CM)	Student provides both the causal and the mechanistic account of the reaction.	Megan: "The carbon has a partial positive on it due to the Br and so the negatively charged O attacks positive carbon with its lone pair breaking the bond of C−Br and those electrons go to the Br."
	Evidence that the student understands that the lone pair of electrons on OH^- is attracted to the carbon on methyl bromide and the electrons in the C−Br bond go to the Br to become Br^-.	Travis: "The bromine leaves and takes the C−Br electrons with it. This leaves a carbocation which then attracts the lone pair on the oxygen to make the bond."

https://dx.doi.org/10.1021/acs.jchemed.9b00815
J. Chem. Educ. 2020, 97, 313−327

1—Time Point 2 (Figure 2). For Year 1—OCLUE students, this beSocratic activity was included as part of their final homework assignment for the course. Students in OCLUE completed an average of two beSocratic homework assignments per week which counted for 15% of their total OCLUE course grade. In other words, the assessment was part of one out of ~25 homework assignments. These homework assignments are typically graded for participation not correctness. Year 1—Traditional students completed this beSocratic activity for 5% of their total course grade. That is, the overall contribution to the grade was higher for the traditional students. Most students were familiar with the platform because it is used for general chemistry at this institution.

Year 2 and Year 3. The Modified S_N2 Prompt was administered in Years 2 and 3. In Year 2, at the 10th week of Fall 2017 to both Year 2—Traditional students and Year 2—OCLUE students. At this time in the semester, students in both groups had discussed nucleophilic substitution including S_N2 and S_N1 and eliminations. Students had not seen these prompts before nor were they provided with the "desired" response that could have been memorized. This data collection is referred to as Year 2—Time Point 1 (Figure 2). Completion of this activity was part of the OCLUE student's regular homework assignments. Traditional OC1 students were offered a small amount of extra credit for completing the activity (approximately 2% of their final course grade). The same Modified S_N2 Prompt was given to Traditional OC2 and OCLUE OC students at the end of OC2 in Spring 2018. These data will be referred to as the Year 2—Time Point 2 (Figure 2). The Modified S_N2 Prompt was once again given at the end of OC2 in Spring 2019 to both Traditional and OCLUE students. This data collection is referred to as Year 3—Time Point 2 to attempt to replicate findings from Year 2.

Data Analysis

Characterization of Causal Mechanistic Reasoning. The coding scheme reported below (Table 1) is modified from the published acid—base coding schemes previously applied to the reaction of $HCl + H_2O$ and $NH_3 + BF_3$.[29] Explanations that describe bond formation are characterized as descriptive general (DG). Phyllis's response is a description of the reaction given to her in the prompt: "First the OH attacks the carbon center and the Br leaves (carbon—bromine bond breaks) this happens in one step." Some organic students use more advanced vocabulary to discuss a nucleophilic substitution reaction than we found with responses to simple acid—base reactions.[17,29] For example, students might use the terms nucleophile and electrophile but simply using appropriate vocabulary does not serve as evidence of understanding without further explanation. A response that identifies the nucleophile and says that it attacks the electrophile is still characterized as a description and does not demonstrate evidence of understanding causality or the mechanism, and as such was also classified as DG. For example, Calvin's response "The nucleophile attacks the electrophile which makes the leaving group leave" was coded as DG.

Responses that demonstrate evidence of understanding that the reaction occurs because of an electrostatic attraction between a negatively charged species and a positively charged species are classified as descriptive causal (DC). Ryan demonstrates this type of reasoning when he said "The carbon is slightly positive because the bromine is pulling the electrons away from the carbon. The negative oxygen attracts the partially positive carbon and the bromine is pushed off and a new bond is made between the carbon and the oxygen." Descriptive causal responses do not discuss the role of electrons in bond formation and bond breaking, and so Ryan's response does not meet our criteria for causal mechanistic reasoning. On the other hand, descriptive mechanistic (DM) responses do discuss bond breaking and formation in terms of electrons and their activities but do not discuss the causal factors that brought about the reaction. Wanda's response shows how it is possible to reason about the mechanism without evidence of understanding electrostatic interactions. Wanda said "The electrons from the negatively charged OH are going to attack the carbon. This will push off the bromine and the bromine will get the electrons from the bond between the carbon—bromine bond."

A causal mechanistic (CM) response involves both the role of electrons and electrostatic interactions in the mechanism. Megan demonstrates causal mechanistic reasoning about an S_N2 process when she says "The carbon has a partial positive on it due to the Br and so the negatively charged O attacks positive carbon with its lone pair breaking the bond of C—Br and those electrons go to the Br." For the purposes of this coding activity, reasoning about an S_N1 mechanism in a causal mechanistic way is also coded as CM, as demonstrated by Travis "The bromine leaves and takes the C—Br electrons with it. This leaves a carbocation which then attracts the lone pair on the oxygen to make the bond." A comprehensive codebook is provided in Table 1.

There were instances in which student responses were non-normative. For example, Emmy's response "the oxygen accepted a proton and formed O—H bond and Br leaves" was explaining a process other than a nucleophilic substitution. We also observed instances in which students explained an S_N1 reaction instead of an S_N2. We still analyzed these responses using the causal mechanistic codes in Table 1. We expand further on this analysis in the next section.

We have used these coding schemes because they are based on the ways that most organic texts discuss such reactions. Ideally, we might want students to also incorporate the idea that reactions begin with collisions between the reacting entities, that the collisions must have enough energy to surmount the activation energy barrier, and that they must be in the correct orientation. However, the prompt does not specifically probe for this kind of reasoning, and little evidence of this emerged. We also use the term "attack" because most instructors describe reactions in this way, and we did not want to privilege any particular group of students with the coding scheme.

While the characterizations in Table 1 are mutually exclusive, we did also code other response characteristics that could be applied in addition of the reasoning codes. For example, when students described an S_N1 process with a carbocation, the response was tagged with an S_N1 Tag. Only those students whose explanation clearly identified the formation of a carbocation or clearly implied the leaving group leaving before the oxygen approached were assigned an S_N1 tag. Conversely, we also assigned an S_N2 Tag for those students who correctly described a simultaneous process or one in which the approach of the nucleophile initiates the reaction.

Students would occasionally justify why the carbon is partially positive by discussing the electronegativity of the bromine atom and/or discussing the polarity of the carbon—bromine bond. Justifying why the carbon is partially positive is not required to be characterized descriptive causal or causal mechanistic but we did denote responses that discussed

Table 2. Summary of Tags Assigned to a Response When Warranted

Tags[a]	Tag Description	Example from Student Responses
S_N1 Tag	The student describes an S_N1 process. It explicitly describes the Br^- leaving and then the OH^- bonding to the carbocation.	"The electrons jump from the C−Br bond to the Br and it becomes negative, since then the C will have a positive charge and then the nucleophile OH^-comes into bond."
S_N2 Tag	The student describes an S_N2 process. It explicitly describes the OH^- approaching and the Br^- leaving in that order or simultaneously.	"The oxygen approaches the carbon and the bromine leaves taking the electrons in the C−Br bond with it."
Polarity Tag	The student explains why the carbon is partially positive or explains bond polarity at the electronic level.	"I think that the C−Br is polar, so the Br is hogging electrons by induction, so that carbon has a partial positive charge. The OH^- nucleophile attacks it, and Br is a good leaving group that can support the electron pair, so it leaves and has a −1 charge."
Terminology Tag	The student uses the terms "nucleophile" and/or "electrophile" in their reasoning.	"The nucleophile attacked the electrophile and formed the C−O H bond."

[a]These tags are applied on top of the characterizations in Table 1.

electronegativity or polarity with a Polarity Tag. Additionally, student use of the terms "nucleophile" and/or "electrophile" were characterized by a Terminology Tag. These tags are in addition to a general descriptive, descriptive causal, descriptive mechanistic, or causal mechanistic characterization. We used their frequency counts to make decisions about modifying the prompt going into Year 2.

The written explanations were exported into a spreadsheet so the four pieces of a student's explanation (i.e., Classify..., Describe what..., Explain why..., Explain your arrows...) could be analyzed together. We found it important to consider all four pieces of students' written work together (see Results and Discussion). For example, students would often include part of their explanation in the "Classify" space. In other cases, the student's response became more sophisticated when we consider their "Explain your arrows" piece of the response (Finding 1b below). Therefore, we decided to analyze "Classify", "Describe what", "Explain why" and "Explain your arrows" together and assigned a code from the causal mechanistic coding scheme shown in Table 1 and any necessary tags shown in Table 2.

All explanations and drawings were coded by the first author with the assistance of two undergraduate coders. The undergraduate coders worked with the authors to refine the previously published acid−base causal mechanistic reasoning coding scheme[17,29] to encompass new complexities in vocabulary that arose with the new reaction. The first author worked with the undergraduate coders to obtain Cohen's Kappa values of 0.72−0.88 using 20% of the explanation data. The mechanistic arrow drawings were coded as correct or incorrect (Figure 3) by the first author and the undergraduate coders.

Comparing Explanations to Arrow Drawings. As previously discussed, responses from Years 2 and 3 were coded both for causal mechanistic reasoning and whether the reaction was discussed as an S_N2 or an S_N1 reaction, regardless of the type of reasoning they were using. Additionally, because we can also replay student arrow drawing, we can determine the order in which the arrows were drawn, and thus determine whether the arrows portray an S_N2 or an S_N1 reaction. To draw a mechanism for an S_N2 reaction, we would expect that the first arrow would begin at the lone pair on the oxygen atom (electron source) and end at the carbon atom in the methyl bromide species (electron sink). The second arrow would be from the carbon−bromine bond (electron source) to the bromine atom (electron sink). We recognize that, by definition, an S_N2 reaction proceeds by simultaneous bond breaking and bond forming. Attempting to model this presents a limitation as it is impossible to draw both arrows simultaneously. However, we have chosen to characterize arrows that start from the oxygen lone pair as the necessary first arrow as this is the first "source" of electrons in the chain of electron movement. Students who displayed use of mechanistic arrows in this way were identified as an S_N2 Arrow user (see Table 3).

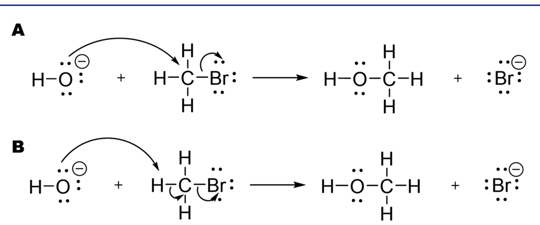


Figure 3. Example of a correct mechanistic arrow drawing (top) and an incorrect mechanistic arrow drawing (B).

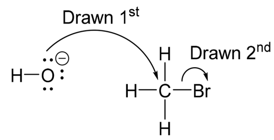
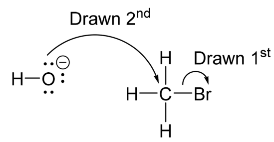
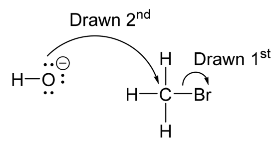
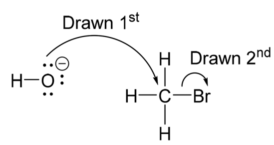
However, in some instances we observed students drawing the arrow from the carbon−bromine bond to the bromine first followed by the second arrow from the oxygen lone pair to the carbon atom second. Using arrows in this order is identified as S_N1 Arrow user because drawing the arrow from the carbon−bromine bond first indicates that the reaction is initiated by breaking of the carbon−bromine bond rather than the approach of the nucleophile which serves as the initial source of electrons. This gives us several possible combinations of explanations and mechanistic arrow use as shown in Table 3.

RESULTS AND DISCUSSION

RQ 1: How Does the Nature of the Prompt Affect Organic Student Responses about a Simple Nucleophilic Substitution Reaction?

Finding 1a: More Causal Mechanistic Responses Were Elicited by the Modified S_N2 Prompt. The goal in first year of this study was to investigate how to best elicit causal mechanistic responses for this nucleophilic substitution reaction using two different prompts. Comparing student causal mechanistic reasoning across both versions of the prompt (Figure 3a,b), we found the distributions to be similar when comparing the proportion of causal mechanistic responses to non-causal mechanistic ($\chi^2(1) = 0.626$, $p = 0.429$). While there is no difference between the proportion of causal mechanistic responses to non-causal mechanistic responses, there is a significant difference between the proportion of students who explicitly discussed electrostatic attraction and were coded as DC or CM compared to all other characterizations: 52% of students gave a causal explanation for the Original S_N2 Prompt versus 66% for the Modified S_N2

Table 3. Classifications Used To Compare Explanations to Arrow Drawings

Explanation with Student Quote	Mechanistic Arrow Use	Code
S_N2 Explanation: *"The oxygen attacks the carbon and then the bromine leaves."*	S_N2 Arrows 	S_N2 Explanation and S_N2 Arrows
S_N1 Explanation: *"The bromine leaves and then the oxygen comes in and bonds."*	S_N1 Arrows 	S_N1 Explanation and S_N1 Arrows
S_N2 Explanation: *"The oxygen attacks the carbon and then the bromine leaves."*	S_N1 Arrows 	S_N2 Explanation and S_N1 Arrows
S_N1 Explanation: *"The bromine leaves and then the oxygen comes in and bonds."*	S_N2 Arrows 	S_N1 Explanation and S_N2 Arrows
	Incorrect Arrows	Incorrect Arrows

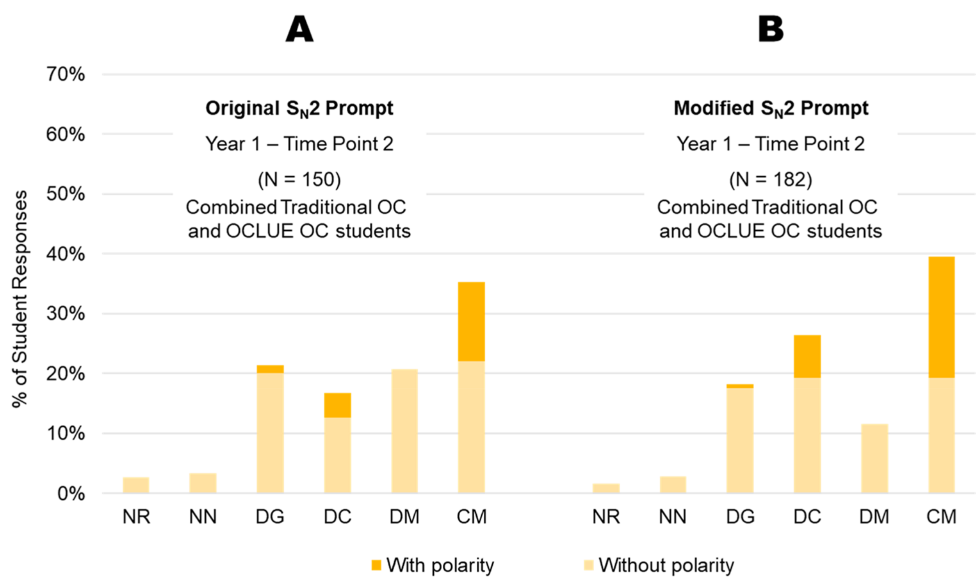


Figure 4. Comparison of causal mechanistic reasoning between the Original S_N2 Prompt (A) and the Modified S_N2 Prompt (B) at the end of Year 1−Time Point 2. These characterizations for each prompt type are further separated by student use of polarity: no response (NR), non-normative (NN), descriptive general (DG), descriptive causal (DC), descriptive mechanistic (DM), causal mechanistic (CM).

Prompt (χ^2 (1) = 6.633, p = 0.010, Cramer's V = 0.140, small effect size).

We also investigated the number of students whose response justified why the carbon in methyl bromide is partially positive by discussing the polarity of the carbon−bromine bond. Even though we did not specifically ask for that level of depth in their explanations,19% of Original S_N2 Prompt responses (28 of the 150 responses) and 28% (51 of the 182 responses) of Modified S_N2 Prompt responses included reasoning about why the carbon is partially positive ($\chi^2(1)$ = 3.969, p = 0.046, Cramer's

https://dx.doi.org/10.1021/acs.jchemed.9b00815
J. Chem. Educ. 2020, 97, 313−327

V = 0.109, small effect size). We were interested to see if discussing polarity had any relationship to student's causal mechanistic reasoning and found that half of the causal mechanistic responses in the Modified S_N2 Prompt (Figure 4a) discussed polarity compared to a third of the causal mechanistic responses in the Original S_N2 Prompt (Figure 4b). It was for these reasons that we decided to move forward with the Modified S_N2 Prompt in Year 2.

Finding 1b: Student Responses Improve after Drawing Mechanistic Arrows. Responses were assigned a code twice: the first code was based on the student responses to the questions before they drew the mechanism, and the second from any additional response students gave after they drew the mechanism. We found that student responses often became mechanistic after they drew mechanistic arrows. For example, a response that was descriptive causal based only on their "Classify...", "Describe what..." and "Explain why..." could have become casual mechanistic after taking their "Explain why you drew your arrows as drawn" into account. In the case of the Modified S_N2 Prompt, only 34% of responses explicitly discussed electron movement while this percentage jumped to 51% after drawing mechanistic arrows as shown in Table 4

Table 4. Comparative Percentage of Students Whose Response Became Mechanistic after Explaining Their Drawn Mechanistic Arrows

	Students with Mechanistic Explanations, %	
Response Codes[a]	Before Drawing Mechanistic Arrows	After Drawing Mechanistic Arrows
NN, DG, DC	66	49
DM, CM	34	51

[a]Notation: NN, non-normative; DG, descriptive general; DC, descriptive causal; DM, descriptive mechanistic; CM, causal mechanistic.

($\chi^2(1) = 9.211$, $p = 0.002$, Cramer's $V = 0.175$, small effect size). Being asked to model electron movement (i.e., draw mechanistic arrows) and then explicitly explain their model influenced students' explanations to be more mechanistic. This observation aligns with *The Framework's* goals for student engagement in modeling specifically that "science often involves the construction and use of a wide variety of models and simulations to help develop explanations about natural phenomena."[25] As such, organic chemistry students should be using their mechanistic arrows to represent their thinking of how reactions occur. We were best able to elicit this understanding by engaging students in explanation and modeling together further suggesting that student understanding in organic chemistry should be carefully elicited by activating appropriate causal and mechanistic resources.[22]

RQ 2: How Does the Type of Organic Chemistry Course Affect Student Ability to Engage in Causal Mechanistic Reasoning?

Finding 2a: Students in Both Types of Courses Provide Similar Distributions of Responses to the Prompt Immediately after Learning the Construct. At Year 2—Time Point 1, 10 weeks into the semester, all instructors agreed that their students would be prepared to answer questions about S_N2 reactions. After data analysis, we observed that students reasoned similarly, regardless of course type ($\chi^2(1) = 0.021$, $p = 0.884$). As shown in Figure 5A, over 50% of students in both courses constructed causal mechanistic

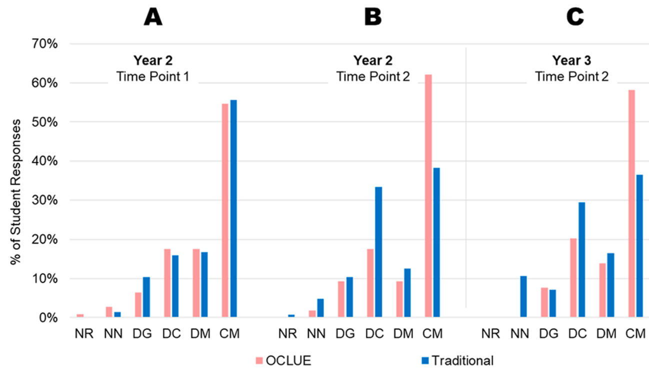


Figure 5. Distribution of causal mechanistic reasoning characterizations for OCLUE and Traditional cohorts for Year 2—Time Point 1 (A), Year 2—Time Point 2 (B), and Year 3—Time Point 2 (C). NR = no response, NN = non-normative, DG = descriptive general, DC = descriptive causal, DM = descriptive mechanistic, CM = causal mechanistic.

explanations. This is evidence that students in both courses (1) were taught about the electron movement and electrostatic attractions in their course and, (2) interpreted the prompt similarly. Combining the proportion of descriptive causal responses and causal mechanistic responses, we observed that the majority of students in both courses included a discussion of why the reaction occurred. This contrasts with Anderson and Bodner's finding, who found that organic instructors do teach why reactions occur but not all students pick up on it.[40] Anderson et al. asserted that the fast pace of the course hindered students from incorporating "the whys" into their understanding.[40] It should be noted that the Year 2—Time Point 1 assessment was given right after students learned about nucleophilic substitution and therefore, discussion of polarity and explicit discussion of electron movement would have been fresh in their minds and those resources readily available.[22]

Finding 2b: Student Use of Organic Specific Terminology Was Not an Indication of the Type of Response Provided. The causal mechanistic coding scheme is not dependent on the use of organic specific vocabulary. That is, the use of organic chemistry terminology such as nucleophile and electrophile is not necessarily accompanied by appropriate causal mechanistic reasoning. Indeed, the frequency with which students used such terminology (as noted by the Terminology Tag) for causal mechanistic and noncausal mechanistic characterizations was not statistically different between the aggregated causal mechanistic responses and the aggregated noncausal mechanistic responses for each cohort at a given time point (Figure 6) (full statistical output reported in Supporting Information, S15). For example, for the Year 2—OCLUE cohort 72% of noncausal mechanistic responses invoked terminology such as nucleophile and electrophile but did not demonstrate understanding of electrostatics or explicit electron movement. Similarly, 57% of causal mechanistic responses invoked use of this terminology. This emphasizes the importance of analyzing student responses not only by the sophistication of vocabulary but by the sophistication of the ideas invoked in their reasoning. We recognize that careful analyses such as these may be difficult in large enrollment course settings. Lexical analysis models have been invoked as a possible solution for analyzing large volume open-ended responses.[41,42]

https://dx.doi.org/10.1021/acs.jchemed.9b00815
J. Chem. Educ. 2020, 97, 313−327

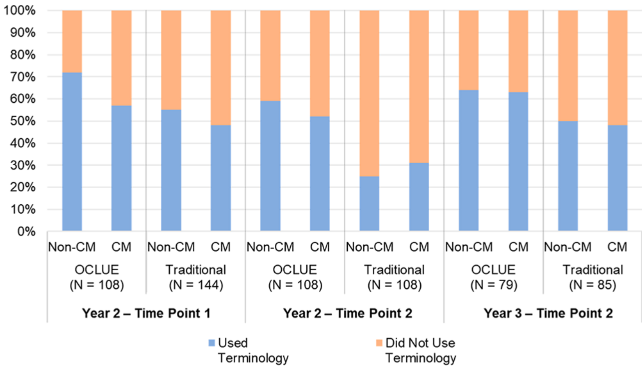


Figure 6. Distribution of organic terminology use for non-causal mechanistic and causal mechanistic responses.

RQ3: How Does the Reasoning about a Reaction Change over the Course of Two Semesters?

Finding 3a: OCLUE Students Improve over Two Semesters while Traditional Students Regress. By the end of OC2, differences in the pattern of responses between the two cohorts emerged, (χ^2 (1) = 14.047, $p < 0.001$, Cramer's V = 0.236, medium effect size) as shown in Figure 5B. At this time point, the percent of OCLUE students engaging in causal mechanistic reasoning had increased from 55% to 62%, while Traditional students had decreased from 56% to 38%, with an accompanying increase in descriptive causal responses. By the end of OC2, it appears that Traditional students were less likely to explicitly discuss electron movement (the how), but they were still likely to reason about the why (Figure 5B). It is not clear why this decrease occurs, since students were presumably more familiar with electron pushing mechanisms by this time point. There is disparity in the literature surrounding findings of this nature. There are a number of studies that show students do not connect arrow pushing with the interaction of charged species. For example, Bodner and Bhattacharyya have shown students using arrows to "get them to the product"[6] and we have found that some students draw mechanistic arrows as an afterthought.[3] However, Webber and Flynn found that students did use arrows as a tool in their problem solving process with many mentioning partial charges.[43]

It may be that the regression in causal mechanistic reasoning by Traditional students was the result of different expectations for student use of knowledge and reasoning in the two courses. In OCLUE, students are provided multiple opportunities to explain and reflect on how and why they are constructing mechanisms. Students are required to construct explanations on weekly homework and in weekly recitation sessions. In both of these formative assessment settings, students are given constructive feedback so they can iteratively practice and improve constructing explanations of various phenomena. About half of OCLUE course lecture time is spent reviewing homework explanations from the previous class so students can see what constitutes a thorough explanation and what is merely a description (although it should be noted that no feedback is provided for this particular homework assignment). In contrast, in the Traditional sections, the homework is not reviewed, and the examinations do not require the construction of explanations, models or arguments. The drop in causal mechanistic reasoning for Traditional students suggests that the expectations in the class can affect how students respond to particular prompts.

Finding 3b: Results Are Replicated from Year 2 to Year 3. Year 2—Time Point 2 was intended as a delayed post-test measure to measure longitudinal effects of each course experience. After finding such a striking difference between the course types at the end of OC2, we wanted to verify that this phenomenon was replicable. The National Research Council's report on *Discipline-Based Education Research*[44] indicated that replicated studies provide a moderate level of evidence of a given phenomenon. Indeed, the pattern of data observed in Year 2 was replicated in Year 3 (Figure 5C) and no statistical differences were found between Year 2 and 3 (Table 5).

RQ 4: How Do Student Written Explanations of Reaction Type Compare to Their Mechanistic Arrow Drawings?

Finding 4a: All Students Tend to Be Consistent in Their Explanations and Their Mechanistic Drawings. At Year 2—Time Point 1, the majority of OCLUE students (71%) discussed an S_N2 process as defined by the S_N2 Tag in Table 2 and also drew their arrows in the order of oxygen lone pair to carbon and then carbon—bromine bond to bromine atom. At that same time point, 41% of Traditional students gave an S_N2 explanation and drew arrows consistent with the explanation, while 30% discussed an S_N1 process and drew S_N1-like arrows (i.e., an arrow from the carbon—bromine bond to bromine and then drew a second arrow from the oxygen lone pair to the carbon). The majority of all students drew arrows that were consistent with their written mechanisms, and a few students discussed an S_N2 mechanistic process and then drew arrows in an S_N1-like order and vice versa (Table 6). It should be noted that we were only able to determine whether students were intending to portray an S_N1 mechanism because we could watch the replay of the arrow drawing. Students who described an S_N1 process by discussing a carbocation formation also typically drew their first arrow denoting the cleavage of the C—

Table 5. Chi-Square Comparisons of the Proportions of Non-causal Mechanistic Responses versus Causal Mechanistic Responses

OCLUE Parameters				Traditional Parameters						
Year	Time Point	Cohort	N	Year	Time Point	Cohort	N	χ^{2a}	p-Value	Cramer's V
2	1	OCLUE	108	2	1	Traditional	144	0.021	0.884	
2	2	OCLUE	108	2	2	Traditional	144	14.047	<0.001	0.236
3	2	OCLUE	79	3	2	Traditional	85	7.782	0.005	0.218
2	2	OCLUE	108	3	2	OCLUE	79	0.068	0.795	
2	2	Traditional	144	3	2	Traditional	85	0.277	0.599	
2	1	OCLUE	108	2	2	OCLUE	108	1.021	0.312	
2	1	Traditional	144	2	2	Traditional	144	10.473	0.001	

[a]All analyses were performed in SPSS.

Table 6. Comparison between Reaction Process Explanation and Mechanistic Arrow Use as Characterized in Table 3

			Students Deploying These Explanations and Mechanistic Arrows, %				
Year	Time Point	Cohort (N)	S_N2 Explanation and S_N2 Drawing	S_N1 Explanation and S_N1 Drawing	S_N2 Explanation and S_N1 Drawing	S_N1 Explanation and S_N2 Drawing	Incorrect Arrows
2	1	OCLUE (108)	71	7	17	0	5
		Traditional (144)	41	30	14	3	12
2	2	OCLUE (108)	88	5	2	1	4
		Traditional (144)	65	15	5	3	12
2	2	OCLUE (79)	86	6	6	0	2
		Traditional (85)	75	6	9	3	7

Table 7. Chi-Square Comparison of OCLUE and Traditional Cohorts

OCLUE Parameters[a]				Traditional Parameters[a]						
Year	Time Point	Cohort	N	Year	Time Point	Cohort	N	χ^{2b}	p-Value	Cramer's V
2	1	OCLUE	108	2	1	Traditional	144	22.844	<0.001	0.301
2	2	OCLUE	108	2	2	Traditional	144	17.804	<0.001	0.266
3	2	OCLUE	79	3	2	Traditional	85	3.031	0.082	
2	2	OCLUE	108	3	2	OCLUE	79	0.145	0.703	
2	2	Traditional	144	3	2	Traditional	85	2.845	0.092	

[a]The proportion of S_N2 explanation and S_N2 arrows was compared to the proportions of all other characterizations shown in Table 6. [b]Two cohorts being compared in each given chi-square analysis.

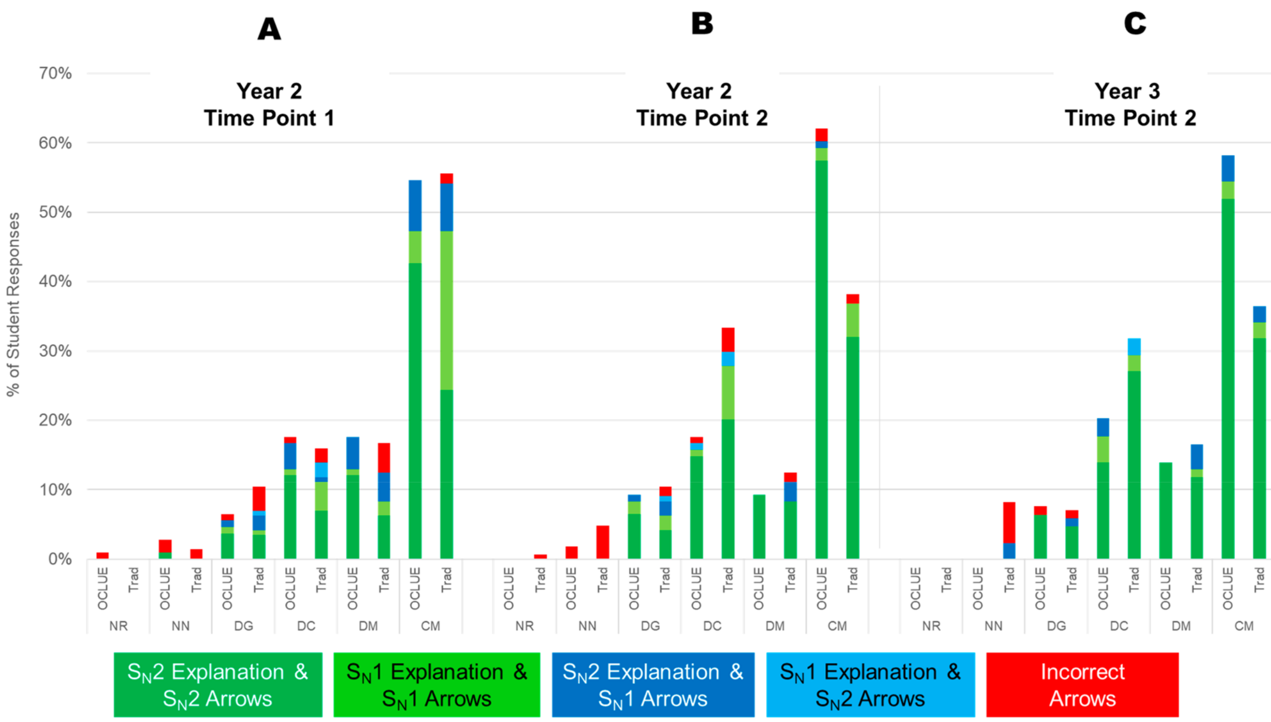


Figure 7. Explanations and arrow drawing comparison compared to causal mechanistic reasoning for Year 2—Time Point 1 (A), Year 2—Time Point 2 (B), and Year 3—Time Point 2 (C). NR = no response, NN = non-normative, DG = descriptive general, DC = descriptive causal, DM = descriptive mechanistic, CM = causal mechanistic.

Br bond. Had this been a more traditional pencil and paper task we would not have known the order of arrows drawn and could not have made this connection. This finding supports the idea that the order in which students draw mechanistic arrows is important.

It is clear that the only way to know if a student has a coherent understanding about this reaction mechanism is to watch the drawing replay and read the accompanying reasoning. Traditional organic assessments where students are asked to draw mechanistic arrows or merely predict products without demanding a mechanism rarely elicit such student reasoning[37] nor can they provide evidence of the order of their arrow use. We believe that the evidence provided here indicates that the traditional organic task of drawing mechanistic arrows does not necessarily provide strong evidence that a student understands how and why reactions occur.

https://dx.doi.org/10.1021/acs.jchemed.9b00815
J. Chem. Educ. 2020, 97, 313−327

By the end of OC2, 88% of OCLUE students and 65% of Traditional students correctly identified the reaction as an S_N2 process and drew canonical arrows. The differences between Traditional and OCLUE, both at the Year 2−Time Point 1 and Year 2−Time Point 2, have small to medium effect sizes (see Table 7) as determined by the chi-square test comparing the proportions of students who had an S_N2 explanation and canonical arrow versus all the other incorrect explanation/arrow categories.

Finding 4b: S_N2 versus S_N1 Is Difficult for Students, Even if They Do Reason in a Causal Mechanistic Way. Figure 7 shows the mechanistic reasoning codes coupled with the results from the S_N2/S_N1 Explanation/Arrows coding scheme. Recall that 55% of OCLUE and Traditional students gave a causal mechanistic response in Year 2−Time Point 2. Here we see that not all causal mechanistic responses yielded a canonically correct explanation and drawing of an S_N2 process.

At Time Point 1 (Figure 7A) over 40% of OCLUE students and 25% of Traditional students provided a causal mechanistic explanation about S_N2 reactions with appropriately drawn arrows. Even for some students who engage in causal mechanistic reasoning, distinguishing between S_N2 and S_N1 proves to be difficult, and this issue persists to a lesser degree at the end of OC2. By the end of OC2, we see that 57% of OCLUE students and 32% of Traditional students provided a causal mechanistic explanation about the canonically correct process (Figure 7B). Again, the results are replicated for Year 3 (Figure 7C).

SUMMARY

In this three-year study, we have extended our prior studies[17,29] on student explanations of acid−base reactions to simple S_N2 reactions: we characterized student written responses and their corresponding mechanistic arrows to draw the following conclusions.

1. The task prompts matter—careful construction of the task can elicit more complete and appropriate answers from students. Therefore, we slightly modified the prompt so that students were not "sidelined" by none-productive ideas (in the context of this study). This resulted in a larger proportion of students providing causal mechanistic explanations. We also found that building in opportunities to reflect and revise answers improved student responses.
2. Students in organic chemistry may use technical terminology without a concomitant understanding of the meaning of the words. For example: when students use terms such as electrophile and nucleophile this does not necessarily mean that they are able to provide the causal reasoning that underlies those terms.
3. The task prompt elicited comparable types of responses from both OCLUE and traditional students immediately after they learned the reaction. This supports both the construct validity of the prompt, since students in both courses were interpreting the task similarly, and the fact that both courses taught the relevant material from which to construct a causal mechanistic explanation.
4. Students who were in a transformed course tended to provide more causal mechanistic explanations than students in the traditional course by the end of the second semester, and these results were replicated in the following academic year.
5. Students who were in the transformed course were more likely to choose the correct sequence of events for this simple S_N2 reaction and represent the process with canonically correct arrow use and an appropriate reasoning. Especially in the first semester, even for the simplest reactions, many students have difficulty determining from the reactants whether the mechanism would be S_N1 or S_N2. However, the majority of all students were able to draw mechanistic arrows that corresponded with the mechanism that they described.

IMPLICATIONS FOR INSTRUCTION

The structure of the curriculum and the tasks that students are asked to complete matter. Tasks that only require surface understanding or that can be answered by memorization and/or pattern recognition communicate a strong message about what instructors want students to know and be able to do with their knowledge. Exams send a particularly strong message to students about what is most important.[38] Ideally, we would like to design learning environments, formative assessments, and approaches to feedback that elicit ideas, support reflection, and help students make the connections that are the hallmark of deep and useful knowledge. Typical organic chemistry curricula and associated tasks are not designed to support this kind of reasoning and the connections required to support it[37] and therefore students cannot be expected to sustain and recall it at a later date.

As a broader implication, we uncovered confusion about even the simplest nucleophilic substitution reaction, where many students were not able to determine whether an S_N1 or S_N2 mechanism was appropriate. Perhaps this should spark conversation about the purpose of emphasizing unimolecular nuclear substitutions in undergraduate organic chemistry. Holliday et al. found that the S_N1 reaction mechanism was minimally important in their analyses of the MACiE database of enzyme reaction mechanisms.[45] Rather, proton transfers, bimolecular nucleophilic addition (Ad_N2), unimolecular heterolytic elimination (E1cb), and bimolecular nucleophilic substitutions (S_N2) were the most common enzyme reaction mechanisms. At our institution, over 90% of students enrolled in organic chemistry (both Traditional and OCLUE) are preprofessional majors (premedical, preveterinary, predental, etc.) and so perhaps valuable instruction time could be spent on reactions that are more important for biological processes.

IMPLICATIONS FOR RESEARCH

This study looks at only the simplest type of organic reaction, and it will be important to extend such studies to other reactions. For example, our future work will build on this study to investigate consistency in reasoning across different reactions over time, specifically a simple, textbook S_N1 reaction and a more complex, intramolecular S_N2 reaction. As noted earlier, the present study only discusses data from students who were consistently enrolled in both semesters of OCLUE or Traditional OC. However, because of scheduling issues it will be important to investigate the effects of switching between course types on student reasoning, overall performance, and grades.

The coding schemes we have used in all of our work on mechanistic reasoning have centered around the role of interactions (electrostatic forces) as causal elements in reaction mechanisms. In future work we will also capture the role of energy and entropy.

Finally, as noted these prompts are highly scaffolded in order to elicit what students know and can do. This brings up the question of what happens when these scaffolds are removed? Do students revert to a simpler explanation, or will "habits of mind" developed over the course of time prevail? And if so, what "dosage" is effective to provide students with the tools to construct causal mechanistic explanations without prompting. There is little research in this area,[14] but in our work on mechanistic reasoning and London Dispersion Forces we found that while the sophistication of response did drop when scaffolding was removed, student responses were still higher than if the scaffolding had never been used in the first place.[46]

■ LIMITATIONS

Data for this study were collected from low-stakes homework assignments. We have not attempted to elicit reasoning to these specific prompts in a summative assessment environment, and therefore it is possible that these responses may not represent students' best efforts. However, we attempted to mediate this limitation by replicating our study and found that reasoning trends were similar from Year 2 to Year 3. We have also found in other studies that there is little difference between student responses for homework and on summative assessments.[46] Second, the reaction used in this study is a simple nucleophilic substitution with the structures fully expanded and products provided. This design was intentional so as to eliminate confusion about the representations and guide students toward explaining the given phenomenon; however, it may not be representative of the various representations students encounter throughout the course (e.g., line structures, wedge-dash representations, Newman projections, etc). This study did not provide any evidence for how student reasoning would change with more complex reactions or more complex representations. Finally, our causal mechanistic coding scheme characterized students' productive ideas about how and why the reaction occurred. Students responses did include incorrect ideas and incorrect use of vocabulary such as conflating the terms electronegative and formal negative charge. Rather than cataloging these occurrences, we have chosen to focus on making sense of students' productive ideas and triangulating their responses with their mechanistic arrow use and to study their knowledge overtime via delayed post-test.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.9b00815.

Additional information about the participant demographics and statistical comparisons (PDF, DOCX)

■ AUTHOR INFORMATION

Corresponding Author

Olivia M. Crandell – *Michigan State University, East Lansing, Michigan;* orcid.org/0000-0003-2748-8942; Email: crande48@msu.edu

Other Authors

Macy A. Lockhart – *Michigan State University, East Lansing, Michigan*

Melanie M. Cooper – *Michigan State University, East Lansing, Michigan;* orcid.org/0000-0002-7050-8649

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jchemed.9b00815

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the organic instructors and students who participated in our study. This work is supported by the National Science Foundation under DUE 1502552. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors and do not necessarily reflect the view of the National Science Foundation.

■ REFERENCES

(1) Morrison, R.; Boyd, R. *Organic Chemistry*, 3rd ed.; Allyn and Bacon: Boston, MA, 1976.

(2) Bhattacharyya, G. From Source to Sink: Mechanistic Reasoning Using the Electron-Pushing Formalism. *J. Chem. Educ.* **2013**, *90* (10), 1282–1289.

(3) 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.

(4) Bhattacharyya, G. Trials and Tribulations: Student Approaches and Difficulties with Proposing Mechanisms Using the Electron-Pushing Formalism. *Chem. Educ. Res. Pract.* **2014**, *15*, 594–609.

(5) Flynn, A. B.; Featherstone, R. B. Language of Mechanisms: Exam Analysis Reveals Students' Strengths, Strategies, and Errors When Using the Electron-Pushing Formalism (Curved Arrows) in New Reactions. *Chem. Educ. Res. Pract.* **2017**, *18*, 64–77.

(6) Bhattacharyya, G.; Bodner, G. M. It Gets Me to the Product": How Students Propose Organic Mechanisms. *J. Chem. Educ.* **2005**, *82* (9), 1402.

(7) Galloway, K. R.; Stoyanovich, C.; Flynn, A. B. Students' Interpretations of Mechanistic Language in Organic Chemistry before Learning Reactions. *Chem. Educ. Res. Pract.* **2017**, *18* (2), 353–374.

(8) Galloway, K. R.; Leung, M. W.; Flynn, A. B. Patterns of Reactions: A Card Sort Task to Investigate Students' Organization of Organic Chemistry Reactions. *Chem. Educ. Res. Pract.* **2019**, *20* (1), 30–52.

(9) Graulich, N. Intuitive Judgments Govern Students' Answering Patterns in Multiple-Choice Exercises in Organic Chemistry. *J. Chem. Educ.* **2015**, *92* (2), 205–211.

(10) Graulich, N.; Bhattacharyya, G. Investigating Students' Similarity Judgements in Organic Chemistry. *Chem. Educ. Res. Pract.* **2017**, *18*, 774–784.

(11) Lee, H.-S.; Liu, O. L.; Linn, M. C. Validating Measurement of Knowledge Integration in Science Using Multiple-Choice and Explanation Items. *Appl. Meas. Educ.* **2011**, *24*, 115–136.

(12) van de Pol, J.; Volman, M.; Beishuizen, J. Scaffolding in Teacher-Student Interaction: A Decade of Research. *Educ Psychol Rev.* **2010**, *22*, 271–296.

(13) Reiser, B. Scaffolding Complex Learning: The Mechanisms of Structuring and Problematizing Student Work. *J. Learn. Sci.* **2004**, *13* (3), 273–304.

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

(15) Guzdial, M. Software-Realized Scaffolding to Facilitate Programming for Science Learning. *Interactive learning environments* **1994**, *4* (1), 1–44.

(16) Wood, D.; Bruner, J. S.; Ross, G. The Role of Tutoring in Problem Solving. *J. Child Psychol. Psychiatry* **1976**, *17* (2), 89–100.

(17) Cooper, M. M.; Kouyoumdjian, H.; Underwood, S. M. Investigating Students' Reasoning about Acid-Base Reactions. *J. Chem. Educ.* **2016**, *93* (10), 1703−1712.
(18) Kararo, A. T.; Colvin, R. A.; Cooper, M. M.; Underwood, S. M. Predictions and Constructing Explanations: An Investigation into Introductory Chemistry Students' Understanding of Structure-Property Relationships. *Chem. Educ. Res. Pract.* **2019**, *20*, 316−328.
(19) McClary, L.; Bretz, S. L. Development and Assessment of A Diagnostic Tool to Identify Organic Chemistry Students' Alternative Conceptions Related to Acid Strength. *Int. J. Sci. Educ.* **2012**, *34* (15), 2317−2341.
(20) Bhattacharyya, G. Practitioner Development in Organic Chemistry: How Graduate Students Conceptualize Organic Acids. *Chem. Educ. Res. Pract.* **2006**, *7* (4), 240−247.
(21) Anzovino, M. E.; Bretz, S. L. Organic Chemistry Students' Fragmented Ideas about the Structure and Function of Nucleophiles and Electrophiles: A Concept Map Analysis. *Chem. Educ. Res. Pract.* **2016**, *17*, 1019−1029.
(22) Hammer, D. Student Resources for Learning Introductory Physics. *Am. J. Phys.* **2000**, *68* (S1), S52−S59.
(23) Kraft, A.; Strickland, A. M.; Bhattacharyya, G. Reasonable Reasoning: Multi-Variate Problem-Solving in Organic Chemistry. *Chem. Educ. Res. Pract.* **2010**, *11*, 281−292.
(24) Pashler, H.; Bain, P. M.; Bottge, B. A.; Graesser, A.; Koedinger, K.; McDaniel, M.; Metcalfe, J. *Organizing Instruction and Study to Improve Student Learning*; National Center for Education Research, Insitute of Education Sciences, U.S. Department of Education: Washington, D.C., 2007.
(25) *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*; National Academies Press: Washington, D.C., 2012.
(26) Rosenshine, B.; Meister, C.; Chapman, S. Teaching Students to Generate Questions: A Review of the Intervetion Studies. *Review of Educational Research* **1996**, *66* (2), 181−221.
(27) 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.
(28) Machamer, P.; Darden, L.; Craver, C. F. Thinking about Mechanisms. *Philosophy of Science* **2000**, *67* (1), 1−25.
(29) 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.
(30) Sevian, H.; Talanquer, V. Rethinking Chemistry: A Learning Progression on Chemical Thinking. *Chem. Educ. Res. Pract.* **2014**, *15* (1), 10−23.
(31) 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*, 1117.
(32) Bode, N. E.; Deng, J. M.; Flynn, A. B. Getting Past the Rules and to the WHY: Cusal Mechanistic Arguments When Judging the Plausibility of Organic Reaction Mechanisms. *J. Chem. Educ.* **2019**, *96* (6), 1068−1082.
(33) Bryfczynski, S. *BeSocratic: An Intelligent Tutoring System for the Recognition, Evaluation, and Analysis of Free-Form Student Input*. Ph.D. Dissertation, Clemson University, Clemson, SC, 2012.
(34) 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.
(35) 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*, 1858−1872.
(36) 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.
(37) Stowe, R. L.; Cooper, M. M. Practicing What We Preach: Assessing "Critical Thinking" in Organic Chemistry. *J. Chem. Educ.* **2017**, *94* (12), 1852−1859.
(38) 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.
(39) Cohen, J. A Power Primer. *Psychological Bulletin* **1992**, *112* (1), 155−159.
(40) Anderson, T. L.; Bodner, G. M. What Can We Do about "Parker"? A Case Study of a Good Student Who Didn't "get" Organic Chemistry. *Chem. Educ. Res. Pract.* **2008**, *9*, 93−101.
(41) Dood, A. J.; Fields, K. B.; Raker, J. R. Using Lexical Analysis To Predict Lewis Acid-Base Model Use in Responses to an Acid-Base Proton-Transfer Reaction. *J. Chem. Educ.* **2018**, *95* (8), 1267−1275.
(42) Dood, A. J.; Fields, K. B.; Cruz-Ramirez de Arellano, D.; Raker, J. R. Development and Evaluation of a Lewis Acid-Base Tutorial for Use in Postsecondary Organic Chemistry Courses. *Can. J. Chem.* **2019**, *97* (10), 711−721.
(43) Webber, D. M.; Flynn, A. B. How Are Students Solving Familar and Unfamiliar Organic Chemistry Mechanism Questions in a New Curriculum? *J. Chem. Educ.* **2018**, *95* (9), 1451−1467.
(44) National Research Council. *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*; Singer, S. R., Nielsen, N. R., Schweingruber, H. A., Eds.; National Academies Press: Washington, D.C., 2012.
(45) Holliday, G. L.; Almonacid, D. E.; Mitchell, J. B. O.; Thornton, J. M. The Chemistry of Protein Catalysis. *J. Mol. Biol.* **2007**, *372*, 1261−1277.
(46) Noyes, K.; Cooper, M. M. Investigating Student Understanding of London Dispersion Forces: A Longitudinal Study. *J. Chem. Educ.* **2019**, *96* (9), 1821−1832.