JOURNAL OF CHEMICAL EDUCATION

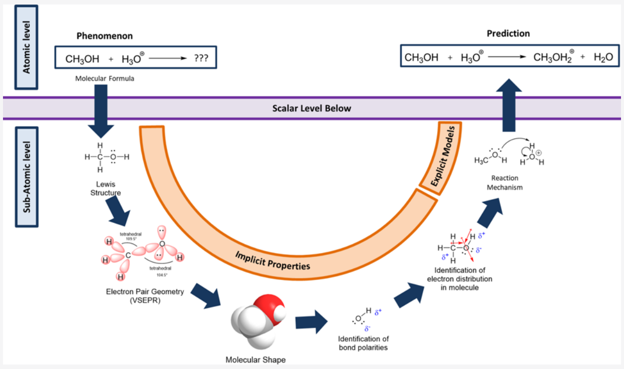
pubs.acs.org/jchemeduc Article

Mechanisms, Models, and Explanations: Analyzing the Mechanistic Paths Students Take to Reach a Product for Familiar and Unfamiliar Organic Reactions

Samantha K. Houchlei, Rosalyn R. Bloch, and Melanie M. Cooper*

Cite This: https://doi.org/10.1021/acs.jchemed.1c00099 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: This study is a follow up to two earlier studies characterizing student real-time use of mechanistic arrows. In these previous studies, students were asked to predict a product by drawing a curved arrow mechanism using an interface that allowed recording and replay of student actions. In the present study two different student cohorts responded to the same tasks as the original studies: a cohort who were enrolled in a traditional organic course, and a cohort who were part of a transformed organic course (Organic Chemistry, Life, the Universe and Everything, OCLUE). Both cohorts improved in their ability to predict an appropriate product over the two semesters, and we found little meaningful difference in the ability of students from either cohort to predict the outcome of a familiar reaction. However, students in the OCLUE cohort were more likely to draw mechanistic arrows than the students from the traditional course. In contrast, when the task involved predicting the product of an unfamiliar reaction, OCLUE students were over three times more likely to draw mechanistically reasonable steps and produce a plausible product than students from the traditional cohort. We propose that the differences between the two cohorts emerge from the following: (1) explicit attempts in the OCLUE course to link drawing reactions mechanisms using the electron pushing formalism to the scientific practice of constructing explanations. It is our contention that this approach changes the arrow pushing mechanism from a skill to the construction of a model which students can use to predict and explain outcomes; and (2) the numerous opportunities in the OCLUE course to try out ideas without penalty, leading to a willingness to try to determine outcomes in unfamiliar situations.

KEYWORDS: *Organic Chemistry, Chemical Education Research, Mechanisms of Reactions*

FEATURE: Chemical Education Research

■ INTRODUCTION

Organic chemistry is often viewed as a difficult course: a hurdle that is deemed necessary for many preprofessional schools as well as many STEM major requirements.[1] The rationales for this requirement are often given as (1) organic chemistry can provide an understanding of reactions that are biologically important, and (2) it provides students with reasoning and critical thinking skills that will prove helpful in their future careers.[2] Indeed, the ability to use structural information to predict the properties of substances, how they interact, and what possible products could form when they react should be important for any science that relies on understanding molecular level phenomena.

To support such understanding, organic chemists have developed the curved arrow notation that is intended to depict the flow of electrons from source to sinks, resulting in the formation and breaking of bonds to produce new products.[3] The use of such curved arrows can be both explanatory (how does a reaction happen), and predictive (what might happen when two substances are mixed and react). When Bhattacharyya surveyed organic chemistry instructors to ascertain what they thought were the uses of the curved arrow, the respondents indicated that "the principle uses of mechanistic reasoning. . . are to explain and predict the outcomes of chemical processes".[3] However, numerous studies have shown that students often struggle with using this notation appropriately for both explanatory or predictive purposes.[3−8] For example Bhattacharyya and Bodner interviewed both undergraduate and graduate students and found that when

Received: February 2, 2021
Revised: July 27, 2021

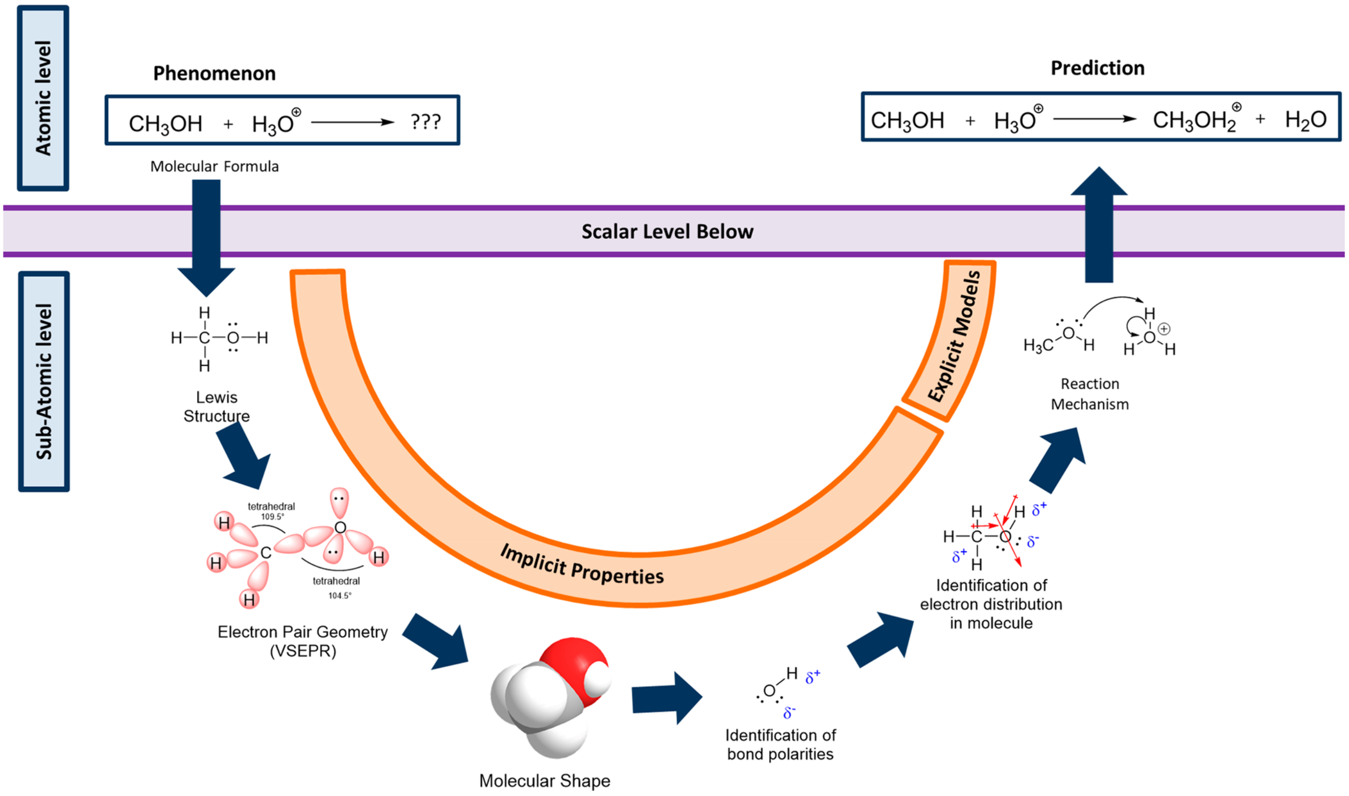


Figure 1. Sequence of inferences and connections that students must be able to make to construct a causal mechanistic explanation about acid base reactions.

given a product they tended to propose arrow pushing mechanisms that might or might not be chemically plausible, but they "get me to the product".[9] Flynn and Featherstone investigated a range of different arrow pushing tasks and found that students were more successful when they were asked to draw arrows onto a reaction scheme in which the reactant and intermediate were given, than tasks for which reactants and arrows were provided and the student had to interpret what the arrows meant and draw the product.[6] That is, students could draw appropriate arrows when shown the starting material and product for a mechanistic step, but were less successful in predicting the outcome of mechanistic arrows. While in this study, students were not asked to predict potential products by constructing appropriate mechanisms, even for the simpler tasks that were studied the authors showed that students have a great deal of difficulty interpreting the meaning of curved arrows. Such research has shown that many students struggle to use them as they were intended; as part of a model of the reacting system to predict and explain the mechanism and outcome of a reaction. A more recent review by Graulich elaborated on the nature of students' understanding and the various factors that influence undergraduate and graduate students' success in organic chemistry, including the use of mechanistic arrows. This review also emphasized the need for researchers and instructors to address the why and how of organic chemistry. Indeed, the author proposed that research should address the idea that emphasizing the connection between a structure and its underlying meaning by verbalizing the properties supports a deeper understanding.[10]

Theoretical Framing: Mechanistic Arrows and the Construction of Models and Causal Mechanistic Explanations

In this section we discuss the relationship between the use of the electron-pushing formalism (mechanistic arrows), the construction and use of a model, and causal mechanistic explanations. While drawing a mechanism using curved arrows is a construct that is usually limited to organic chemistry it can be considered, in a broader sense, as an application of the scientific practice of constructing and using models. This is one of the eight scientific practices defined by the National Academies consensus report "A Framework for K-12 Science Education" (the *Framework*), in which the authors state "Science often involves the construction and use of a wide variety of models and simulations to help develop explanations about natural phenomena. Models make it possible to go beyond observables and imagine a world not yet seen."[11] While there are other approaches to the study of models and modeling that have been used specifically in chemistry,[12,13] here we use the approach to this scientific practice described in the *Framework*. As Schwarz, Passmore, and Reiser note[14] "Models are defined by how they are used. . . scientific models are sense-making tools that help us predict and explain the world." Since arrow pushing mechanisms do allow us to predict and explain the course of reactions, we propose that they belong to this larger class of models; that is, they are representations of a system, using a defined set of components that allow us to provide mechanistic accounts of and predict outcomes of phenomena.

The components of a mechanistic arrow model of a reaction are the structures of the reactants, the concomitant implicit information about electron distribution and properties that emerge from this distribution, and the explicit mechanistic arrows which must be used in conjunction with the implicit information gleaned from the structure. Using these components, the model constructor who constructs an arrow pushing mechanism should be able to both predict *what* will happen, and also explain *why* it will happen. That is, drawing a mechanism makes it possible to characterize both how and why a reacting system produces particular products from given reactants. (We are aware that other factors will also impact why reactions occur—such as energy and entropy changes, but those factors are not the focus of this study).

Unfortunately, the successful use of mechanistic arrows is predicated on the mechanism user understanding the encoded information in both the chemical structures and the arrow symbols, and as discussed earlier there is plenty of evidence to support the idea that many students have difficulty not only with the use of mechanistic arrows,[3−8] but also with the relationship between the structure of a molecule, the electron distribution within that molecule, and how these factors impact the ways that molecules interact.[15−18] Ideally we would want students to uses arrows to construct a mechanism that is concordant with the implicit information encoded in the symbols that are being used. Furthermore, Russ and co-workers in their discussion of mechanistic reasoning propose that "... mechanisms account for observations by showing that underlying objects cause local changes in the system by acting on one another".[19] Krist and co-workers also emphasize the role of the underlying entities at a scalar level below the phenomenon.[20] For our purposes, the reaction is the phenomenon, and the scalar level below includes the electrons (in bonds and lone pairs) and their distribution, which is caused by electrostatic forces among the nuclei, electrons, and atoms themselves.

The use of curved arrows as part of a model to predict and explain requires that students are able to identify potential sources and sinks for electrons, that is the implicit information encoded in the structural representation, resulting from processes that take place at a scalar level below the reaction itself.[19,20] For most polar reactions, the arrow starts at a source of high electron density and ends at the electron "sink" where the electrons will be localized on an atom or where a new bond is formed. To use this approach effectively, students must use their cognitive resources to decide where the electron rich and deficient sites are, and sometimes decide between multiple locations of high and low electron density. Such expertise requires students to connect an extended series of inferences as shown in Figure 1. To construct an arrow-pushing mechanism students begin with the reacting structures. From there they must understand and predict how electron density differences in molecules arise and then use that information to determine how molecules might interact, and then translate this understanding into the electron pushing formalism. It is our contention that the appropriate use of mechanistic arrows to predict and explain is the organic chemistry equivalent of constructing a causal mechanistic explanation. Both require that students use cognitive resources[21] such as how and why electron distributions vary in a particular molecule, and how Coulombic interactions (attractions and repulsions) govern the ways that molecules interact, to fashion the mechanism or explanation. Certainly, using mechanistic arrows is a far more parsimonious and efficient approach to predicting outcomes of organic reactions, but just as with chemical structures themselves, the implicit information embedded within the reactants and arrows that is required to use arrows appropriately may inhibit the use of mechanistic arrows in ways that were originally intended. Indeed, much of the extant literature on how students use curved arrows indicates that students are not using their understanding of how and why electron density differs in a structure, to guide how they draw mechanisms, and instead may "decorate" their structures with arrows,[22] either before or after they have written down the product of a reaction, or simply using them to "get to the product" or "connect the dots" without an underlying reasoning process that involves predicting how and why electrons move during a reaction.[22,23] That is, some students may not be using the mechanism as a model at all and may be memorizing the pattern of electron flow (at best), or simply drawing random arrows (at worst).

Prior Work on Supporting Students Use of Mechanistic Arrows

The question then arises: How can we support students as they learn to draw mechanisms so that this act is based on their understanding of how and why electrons move during a reaction? There are a number of possible productive approaches to helping students learn the use of mechanistic arrows.[7,24−26] For example, Flynn has developed a course in which students first learn the skills associated with arrow use, before they learn how and why reactions occur.[24] The same authors have developed an online learning module "Mastering the Arrows" which showed significant learning gains from pre- to post-test, particularly for tasks in which students were asked to draw products of a reaction.[27,28] In turn, Graulich has proposed that the use of contrasting cases may support students mechanistic reasoning, in particular through the use of scaffolded activities.[29,30] Additionally, Watts et al. have found that engaging students in a scaffolded writing-to-learn prompt helps promote students mechanistic reasoning across more complex acid−base reactions.[31] In a separate study these authors also found that students who engage in a writing-to-learn prompt frequently made explicit and implicit mention of electron movement for a hydrolysis reaction.[32]

In our work and curricular development, we have taken yet another approach to the development of mechanistic expertise: we have chosen to address formal electron pushing mechanism use in the context of developing causal mechanistic explanations about reactions. The *Framework* emphasizes this connection between constructing models and explanations: "The goal for students is to construct logically coherent explanations of phenomena that incorporate their current understanding of science, or a model that represents it, and are consistent with the available evidence."[11] Indeed there is strong evidence (multiple replicated studies across diverse student populations in a range of disciplines) that constructing mechanistic explanations of phenomena, supports deeper understanding.[33,34] As we noted earlier, we believe that drawing a mechanism and constructing a causal mechanistic explanation should call upon the same resources, and therefore we might expect students who learn to construct causal mechanistic explanations should also be able to draw mechanistic arrows. To be clear, students may draw an arrow pushing mechanism without engaging in causal mechanistic reasoning, and indeed many students appear to do exactly this. However, because we believe that the practices of constructing models and causal mechanistic explanations are inextricably intertwined, we have developed curricula where these practices are emphasized and connected. It is our hypothesis that by supporting student construction of written causal mechanistic explanations, we may also enhance student use of mechanistic arrows as models that can also predict and explain reaction outcomes.

We do have some support for this hypothesis reported in earlier studies on student causal mechanistic reasoning about acid base chemistry[35] and nucleophilic substitutions.[35,36] We found that general chemistry students who were able to construct causal mechanistic explanations for how and why simple acid base reactions occurred were also more likely to draw appropriate mechanistic arrows for the same reaction.[35,36] In a study on nucleophilic substitutions we found that students' explanations typically correlated with the mechanistic arrows that they drew; that is, if students believed a reaction was an S_N2 reaction their arrows showed a simultaneous one step reaction, whereas if they described or explained an S_N1 mechanism, their arrows corresponded with this mechanism. In this same study of nucleophilic substitution we found that student written explanations included more discussion of electron movement after drawing mechanistic arrows. Students were first asked to explain why the reactants interact, then to draw the mechanistic arrows for the reaction, and then asked to explain why they drew their arrows as indicated. Before drawing the mechanism 34% of responses explicitly discussed electron movement while this percentage jumped to 51% after drawing mechanistic arrows. This further supports the idea that the resources required for constructing causal mechanistic explanations and arrow pushing models may be connected under appropriate circumstances.

At this point we also acknowledge that other researchers work on characterizing mechanistic reasoning has taken a somewhat different approach. For example, Talanquer and coworkers have written extensively on student reasoning including mechanistic reasoning,[37−39] and following on from this work, Flynn has analyzed student explanations about organic reactions using a framework based partially on the granularity of the student's discussion that also encompasses the number of causal links students make from descriptive to relational to linear causal to multicomponent causal.[40]

Prior Studies on Which This Work Is Based

Most organic chemistry courses do not focus on the construction of causal mechanistic explanations, but rather on the construction of mechanisms using mechanistic arrows, and it is likely that for many organic chemists the ability to draw mechanisms would be seen as more convincing evidence of expertise. It is for this reason that we decided to reinvestigate an earlier set of studies from our group,[22,23] in which we characterized both how students use mechanistic arrows to predict the products of a reaction,[22] and whether the use of mechanistic arrows improves students' success in predicting the correct answer.[23] In these earlier studies, students responded to a set of tasks in which they were asked to draw an arrow pushing mechanism to predict the product for a number of familiar and unfamiliar reactions at four time points across a full year of organic chemistry. These reactions are shown in Figure 2. We used a software system that allowed us to record and replay student responses[41] so that the sequence in which students drew arrows and mechanisms could be determined. The reactions that are labeled Prompt A: Electrophilic Addition of Water to Alkene; Prompt B: Electrophilic Addition of Water to Alkyne; Prompt C: Alkyne Deprotonation Followed by S_N2; and Prompt D: Nucleophilic Attack at a Carbonyl represented tasks that were similar to those presented to students in class and in their textbooks, whereas Prompt E: Unfamiliar Reaction required students to apply their knowledge of organic reactivity to a situation that, to our knowledge, they had not encountered before but should have been able to work through to produce a plausible product.

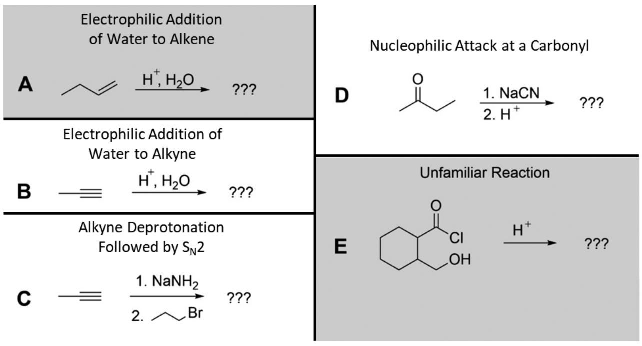


Figure 2. Summary of mechanisms administered to participants in the papers.[22,23] Highlighted prompt will be discussed in this paper.

In these earlier studies we found that many students tended to simply draw out products of known reactions A−D rather than draw mechanisms as we had asked in the prompt.[22] At each time point where data were recorded, between 30 and 60% of students did not use any curved arrows to help predict the products of these reactions and of those who did draw arrows, between 15 and 20% of those students drew the arrows onto the reaction scheme after they had predicted the products. We also found that for familiar reactions (prompts A−D), student use of arrows did not affect the chances of drawing the correct structure.[22] There are a number of possible explanations for this finding, it may be that they simply did not see the benefit of using arrows to produce a product that they already knew was the answer, or perhaps students simply memorized the answer. However, when students were faced with reactions that they had not seen before, but "should" have been able to make predictions about, the students who attempted to draw mechanisms were more likely to predict the correct product; even so, the overall percentage of students who drew a plausible product was very small (9% of the total).

Our goal in this present study is to investigate how different students, at a different university, address the same tasks that were studied earlier.[22,23] In this study we analyze responses from two demographically matched cohorts of students. One group of students were enrolled in a traditional organic chemistry course, similar to the earlier study, using a commercial text, and a lecture format. The second group of students were enrolled in a transformed course, Organic Chemistry, Life, the Universe and Everything.[42]

Research Questions

1. In what ways are responses for familiar reactions from students who are enrolled in a transformed organic course similar or different to an equivalent group of students from a traditional organic course?
2. In what ways are responses for an unfamiliar reaction from students who are enrolled in a transformed organic course similar or different to an equivalent group of students from a traditional organic course?

METHODS

Student Participants: Present Study

This study was conducted at a large midwestern research-intensive university. Students were enrolled in a two-semester organic chemistry course for nonchemistry majors. This study was designated "exempt" and all student participants were informed of their rights as research participants in accordance with our institutions' IRB. Student participants were enrolled in either a transformed organic chemistry course, Organic Chemistry, Life, the Universe, and Everything (referred to as OCLUE) or a traditional organic chemistry course (referred to as Traditional). At this institution students register far in advance, and students typically enroll in sections before the instructor is listed for the course time; students also enroll for a whole year at one time. To ensure that the two groups of students were similar we compared several academic and demographic measures using a Mann−Whitney U test and calculated the effect size for any differences using Cramer's V. While Traditional students had a slightly higher OChem 2 course GPA (Mann−Whitney: Traditional = 3.32, OCLUE = 3.01, $U = 14308.5$, $z = -2.149$, $p = 0.032$, $r = 0.11$ small effect size),[43] no differences were found when comparing ACT scores, GPA Prior to Organic, OC1 course grade. Supporting Information Tables S1−S3 provide a summary of all statistical analyses, which were performed in SPSS,[44] and a report of gender and major distributions for each cohort.

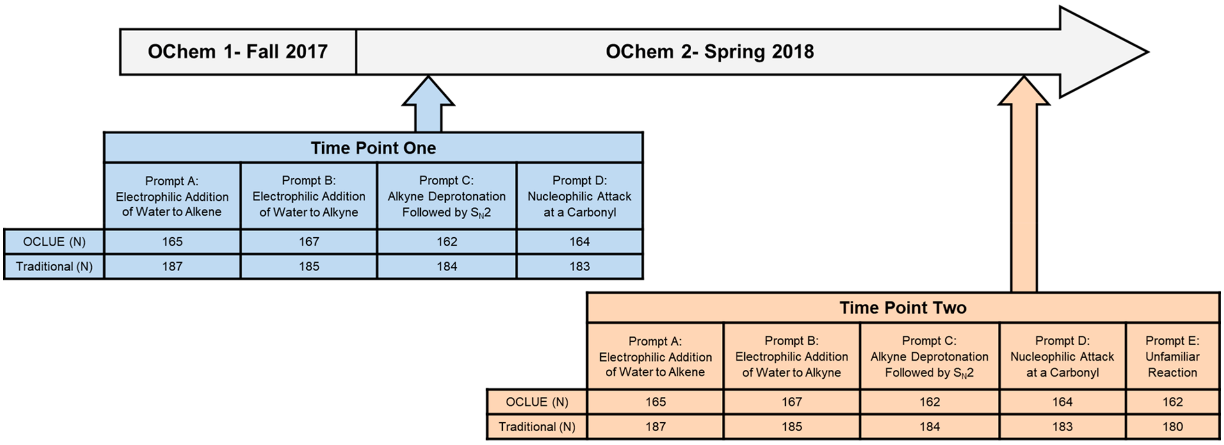
Both types of section meet for either three 50 min or two 80 min lecture classes of around 300−350 students, and a once a week 50 min recitation taught by graduate teaching assistants. The differences between OCLUE and Traditional courses are not a matter of topical coverage—indeed many students must switch sections between semesters for scheduling reasons, so it important that the same topics are "covered". Additionally, this course is a service course, and we are aware of what external expectations are for what students should know and be able to do. As we have previously discussed, the OCLUE curriculum emphasizes biologically important mechanisms, such as acid−base reactions, nucleophilic additions, and substitutions, but also includes material required by the more traditional approaches should students switch sections type.

The development of the curriculum for the transformed course OCLUE, has been reported previously,[42] but will be reviewed briefly here. OCLUE is based on what has come to be known as three-dimensional learning (3DL).[11] It is organized around four core ideas of chemistry: electrostatic forces and bonding interactions, structure property relationships, stability and change in chemical systems, and energy. These core ideas are used in the context of scientific and engineering practices (SEPs), such as analysis and interpretation of data, construction of evidence based arguments, explanations, and models.[11] The third dimension consists of crosscutting concepts that allow instructors and students to focus on a particular aspect of a phenomenon. The aspect of OCLUE relevant to this report is the emphasis on construction and use of models, construction of explanations, and the crosscutting concept of cause and effect, that combine with the core ideas of structure−property relationships and bonding and interactions.

Assessment in the OCLUE course includes biweekly formative assessments delivered via an online system, beSocratic,[45] that allows open-ended drawn and written responses, group recitation worksheets facilitated by graduate teaching assistants, in-class clicker questions, and summative exams that contain both multiple choice and open ended responses. Approximately 45% of the course grade comes from formative assessments that are designed to support and extend student learning, rather than test it. For example, in the biweekly formative assessments delivered on beSocratic students complete both retrospective tasks that are designed to consolidate ideas, and prospective tasks that introduce new material. This is done to allow students to use their knowledge and skills to predict outcomes for systems that will be discussed in the next class.[42] Additionally, all of the formative assessments are graded based on student participation rather than correctness, with the goal of providing students with a "safe space" to make mistakes without penalty. For all the formative assessments (homework and recitation) students are provided with contextual feedback, either from a teaching assistant, or in class. The summative assessments are three midterm and one final examination which typically are evenly split between multiple choice and open response questions.[42] Both the summative and formative assessments in OCLUE include 3D tasks; that is, they require the student to construct, predict, explain, argue, and analyze in the context of core ideas and crosscutting concepts. For example, students might be asked to draw a reaction energy diagram (a model) and use it to predict, explain, and draw mechanisms to show how different products are formed under different reaction conditions; that is, students use the core ideas of energy and structure property relationships, the scientific practices of modeling and explanations, and the crosscutting concept of cause and effect.

The Traditional course uses a commercial textbook[46] that is organized mostly by functional group, and the topics in the course are typically taught in the same order as the textbook. By agreement, both Traditional and OCLUE sections cover the same material in the first semester, so that students who switch sections will have been exposed to the same general content, including the familiar reactions A−D. The course is taught in a lecture format, and there is ample opportunity to ask and answer questions. The instructor provides the course material and solves a wide range of problems for students throughout the lecture. Homework using the associated publishers' online materials is suggested but not required. Recitation sections are also taught by graduate teaching assistants, and typically involve short quizzes worth 20% of their overall grade in the course, a short lecture, and/or a question-and-answer session. Eighty percent of students' grades comes from performance on three summative exams worth 40% of the overall grade and a final exam also worth 40% of their overall grade. These exams are all open response (no multiple choice) and consist of short answer predict-the-product, -reactant, or -reagent, draw a mechanism, analyze an unknown from spectra, and synthesize a given molecule. In other words, the assessments are typical organic chemistry questions that are found in many institutions, reflective of similar content assessed on the ACS organic exam.[47] Just as Stowe et al. found in their analysis of exams from elite institutions across the country,[48] these questions are not 3D. In particular, they do not elicit evidence that students have engaged with the scientific practices. All exams are hand-graded, and feedback is provided on the examination by the instructor or graduate teaching assistant.

In summary, while both are large enrollment courses with around 300−350 students per section that "cover" the same



Time Point One

	Prompt A: Electrophilic Addition of Water to Alkene	Prompt B: Electrophilic Addition of Water to Alkyne	Prompt C: Alkyne Deprotonation Followed by S_N2	Prompt D: Nucleophilic Attack at a Carbonyl
OCLUE (N)	165	167	162	164
Traditional (N)	187	185	184	183

Time Point Two

	Prompt A: Electrophilic Addition of Water to Alkene	Prompt B: Electrophilic Addition of Water to Alkyne	Prompt C: Alkyne Deprotonation Followed by S_N2	Prompt D: Nucleophilic Attack at a Carbonyl	Prompt E: Unfamiliar Reaction
OCLUE (N)	165	167	162	164	162
Traditional (N)	187	185	184	183	180

Figure 3. Summary of the data collection

material, OCLUE and Traditional organic chemistry have different approaches to both curriculum design and assessment, and these differences appear to have led to a different class culture (manuscript in review). For example, students in a traditional section are more likely to say that they are assessed on what they can memorize, whereas OCLUE students are more likely to indicate that their use of knowledge is assessed.

Although students may switch section type between the first and second semester, in this study we focus solely on students who were enrolled either in both semesters of OCLUE, or both semesters of Traditional organic chemistry. We will report on how students who "switch" sections fare in a subsequent publication.

Prompt Timing and Administration

The prompts A–E that were administered in the previous study (Figure 2) were also administered to students in the current study. Time Point One (2 weeks into OChem 2) was chosen because it should provide information about what students had learned in OChem 1. Students were asked to respond to prompts A–D because these are typical reactions students would be expected to know and be able to do in OChem 1 regardless of the students' course type. Prompts A–D are all familiar reactions, and similar reactions to these were assessed by both course types in OChem 1. These will be referred to as familiar reactions from now on.

At Time Point Two (1 week before the end of the second semester) students were asked to respond again to prompts A–D, and to prompt E which involved the reaction that, to our knowledge, students had not seen before in either the OCLUE or Traditional sections of OChem. Prompt E is referred to as an unfamiliar reaction in the earlier papers.[22,23] Figure 3 shows a summary of the administrations of the prompts, and the numbers of students who answered each one. Students at Time Point Two were selected from the group of students who completed Time Point One, so that only students who had answered both prompts appear in this data set. There was some attrition from Time Point One to Time point Two for both courses, about 10% for OCLUE and 5% for Traditional.

In the present study we used the online homework and research platform beSocratic to capture the mechanisms that students drew.[45] Using this system students can draw reactions and mechanisms, just as they would on a hand-written paper assignment. The advantage of using this system is that the responses are recorded and can be replayed to show stroke-by-stroke what students drew and in what order they drew it. Figure 4 is an example of the screen students see when working in beSocratic. In the previous study a stand-alone system,

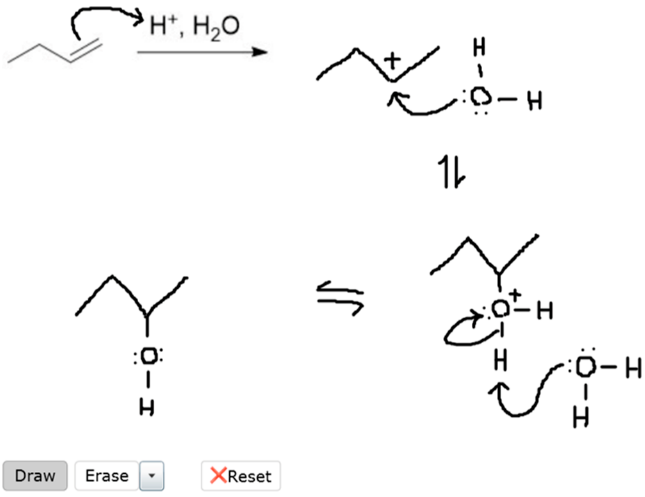


Figure 4. Example of a student's completed mechanism in beSocratic.

OrganicPad,[41] which has similar but more limited capabilities was used. This software is no longer available for use.

Potential for Bias

We, as researchers, acknowledge that there is an inherent conflict when a researcher sets out to investigate the impacts of a curriculum or intervention of their own design. Such a conflict exists here, and in our implementation of the study and the data analysis we have attempted to remove or negate sources of bias. In both sections the activity was given online as an extra credit activity. For OCLUE this extra credit would contribute 0.1% to the overall grade, in the traditional sections the credit would add 1%. Both sections were familiar with the homework system since it is also used for general chemistry. Once the data were collected, they were deidentified and coded anonymously, so that coders were not aware of the source of the data.

Data Analysis

Our original intent was to use the original coding scheme,[22] but because of the differences in the way the data were recorded, for some of the prompts we were able to expand on the original and develop a richer coding scheme. This expanded revised coding scheme encompasses the original

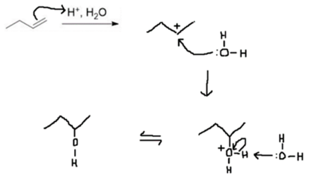
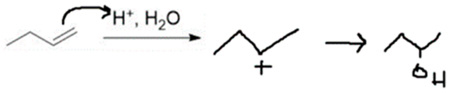
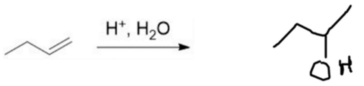
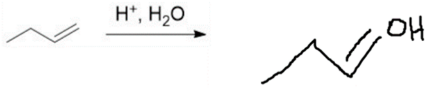
Code Grouping	Description of Code Grouping	Student Example
Plausible Product & Plausible Arrows	Student drew mechanistically reasonable steps and a plausible product	
Plausible Product & Mixture of Plausible and Incorrect Arrows	Student drew a mixture of mechanistically reasonable steps and incorrect arrows and a plausible product	
Plausible Product & Incorrect or No Arrows	Student drew no arrows or incorrect arrows and a plausible product	
Incorrect Product & Mixture of Plausible and Incorrect Arrows	Student drew mixture of mechanistically reasonable steps and incorrect arrows and an incorrect product	
Incorrect Product & Incorrect or No Arrows	Student drew no arrows or incorrect arrows and an incorrect product	

Figure 5. Overarching code groupings for all reactions. The color scheme shows the colors of the code groupings in the graphs in the Results section.

Table 1. Percent of Students Who Drew a Plausible Product[a]

Prompt	OCLUE (count/total *n*)	Traditional (count/total *n*)	χ^2 (df = 1)	*p*-value	Cramer's V
		Time Point One			
A	58% (95/165)	69% (129/187)	4.930^{β}	0.026^{β}	0.12^{β}
B	41% (68/167)	41% (76/185)	0.005	0.945	
C	14% (22/162)	4% (7/184)	10.722^{α}	0.001^{α}	0.18^{α}
D	27% (44/164)	6% (11/183)	29.466^{α}	$<0.001^{\alpha}$	0.29^{α}
		Time Point Two			
A	76% (126/165)	75% (129/187)	0.044	0.833	
B	62% (104/167)	48% (88/185)	7.658^{α}	0.006^{α}	0.15^{α}
C	31% (50/162)	9% (17/184)	25.801^{α}	$<0.001^{\alpha}$	0.27^{α}
D	68% (112/164)	55% (101/183)	6.263	0.012	

[a] For all chi-square analysis $\alpha = 0.01$. (α) students in OCLUE outperformed those in Traditional, (β) students in Traditional outperformed those in OCLUE.

coding scheme, so it is still possible to compare to the students from the previous study and we can also provide here a richer picture of the mechanistic approaches used by students in the present study. In both the original work and current study, we are interested in not only the products students drew but what steps they took to get there. A summary of the revised coding schemes for all five prompts are provided in Figures S4–S8.

Because we are interested in appropriate use of mechanistic arrows, for the revised coding scheme we generated codes for every mechanistically plausible step that students took to get to a plausible product. That is, an arrow that starts at an electron rich site and ends at an electron sink, or generates an appropriate resonance structure is considered reasonable. For example, for prompt B, we expanded the plausible products to the enol, and the geminal diol (the hydrate of the carbonyl) and the vicinal diol. In the current scheme a plausible product is not only the major product but also the minor products that make mechanistic sense for this reaction. This approach is more consistent with the intent of the original coding scheme, it simply extends it to a somewhat broader range of products.

The number of codes for each prompt A–E varied from 7 to 22 to fully capture and characterize students' mechanistic pathways to the product (whether major or minor) which are provided in Tables S9–S13. There were also a small number of students who did not engage with the prompt, for example by drawing a line through the prompt or writing "I don't know". Over all prompts and time points there were 11 OCLUE students and 16 Traditional students who did not engage with any of prompts and thus were removed from the data set.

Because there were so many codes for each reaction, we chose to determine inter-rater reliability (IRR) for each code, rather than for the mechanism as a whole. Two coders, author (S.K.H.) and a trained postbaccalaureate coder (R.B.), blind-coded the anonymized student responses so that the coders were not aware of the students' background (course type) while coding. The two coders took sets of 15 responses and coded them and then discussed differences to reach agreement and occasionally additional codes were added for other plausible ways students drew their arrows. Once agreement was reached on the codes, both coders evaluated 60 random responses from each prompt to obtain kappa values and the author (S.K.H.) coded the remaining responses. Cohen's Kappa values ranging between 0.78 and 1.0 were obtained for all prompts, indicating substantial agreement. Results are provided in Tables S9−S13.[43,49]

Since the mechanisms for each reaction often generated large numbers of codes, for the purposes of comparison the responses were assigned to one of five code groupings based on the description outlined in Figure 5. A plausible arrow refers to any arrow that starts at a source of electron density and ends at an electron sink.

■ RESULTS

Research Question 1: In What Ways Are Responses from Students Who Are Enrolled in a Transformed Organic Course Similar or Different to an Equivalent Group of Students from a Traditional Organic Course for Familiar Reactions?

Finding 1: The Frequency of Plausible Products Drawn Depends on the Reaction for Familiar Reactions. The responses for prompts A−D were classified into two bins: students who proposed a plausible product as discussed earlier, and those who did not. Table 1 shows the percentage of students who drew a plausible product for prompts A−D at Time Point One and Time Point Two. We used a series of chi-square tests of independence to compare the cohorts based on the percentage of students who drew a plausible product.[50] While there are some statistically significant differences between the cohorts, the effect size for all prompts A−D is small, indicating there is not a strong correlation between the course type and drawing a plausible product after both semesters of organic chemistry.[43] In general, after one semester Traditional students are significantly better at drawing a plausible product for reaction A, but this difference is removed after two semesters. At Time Point Two, as one might expect all cohorts produced a plausible product more frequently than they did at Time Point One. It appears that after two semesters there is little difference in the ability of students from either cohort to predict the outcome of familiar reactions.

Finding 2: OCLUE Students Tend to Use Plausible Arrows More Frequently than Traditional Students. We will now turn our attention to how students predicted the products: that is whether students used mechanistic arrows to predict the outcome of the reaction. As discussed earlier, at each Time Point, students were classified into five separate groups based on the product they produced and the way they used mechanistic arrows: (1) students who drew plausible arrows and a plausible product, (2) students who drew a mixture of mechanistically reasonable steps and incorrect arrows and a plausible product, (3) students who drew no arrows or incorrect arrows and a plausible product, (4) students who drew a mixture of mechanistically reasonable steps and incorrect arrows and an incorrect product, and (5) those who drew no arrows or incorrect arrows and an incorrect product. It should be noted that in the earlier studies both groups 1 and 2 were "counted" as using mechanistic arrows in a fruitful manner. An example of a student response for each code is shown in Figure 5.

Figure 6 shows the distribution of the five code groupings for prompt A at Time Point 1. The figures showing the percent of students who drew the various coding groups for Prompt B−D at both time points are provided in the Supporting Information, Tables S14−S20. As shown in Table 2, at Time Point One for prompts A and B it appears that students enrolled in the OCLUE course were significantly more likely (with a large effect size) to draw all mechanistic steps correctly than those who were in a Traditional course.

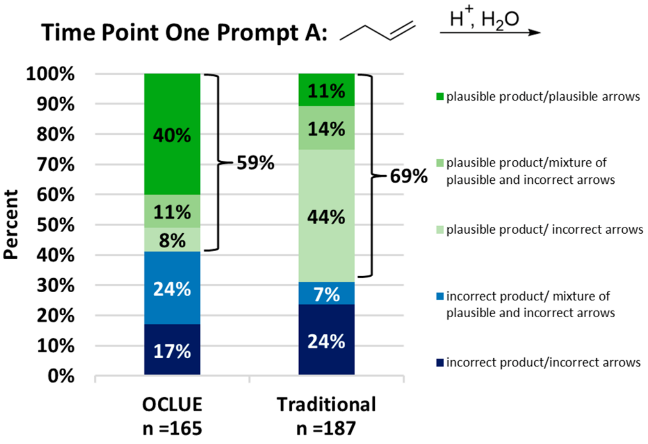


Figure 6. Percent of students who drew all mechanistic steps correctly, students who drew some mechanistic steps correctly, students who drew no mechanistic steps correctly, and students who got the incorrect product based on the courses they took.

At Time Point Two, the frequency with which both groups drew plausible mechanistic arrow increased from Time Point One. The frequency with which students draw all mechanistic steps correctly, now differs between the Traditional and OCLUE students across all prompts A−D as shown in Table 3. For example, Figure 7 shows the students who drew all mechanistic steps correctly for a simple alkene hydrolysis reaction.

Finding 3: A Higher Percentage of OCLUE Students Improve Their Responses over the Course of Two Semesters than Traditional Students. Since this is a longitudinal study, student responses at both Time Point One and Time point Two were plotted on Sankey diagrams. Figure 8 shows how the responses provided by each student changed for Prompt A the Familiar Alkene. The width of the pathways between the two time points represents the proportion of students who took that path. For example, at Time Point One 40% of OCLUE students who drew all plausible arrows and predicted a plausible product, while at time point 2 there were 64% who did this. However, only 34% of OCLUE students drew all plausible arrows and predicted a plausible product at both. Similarly for Traditional students while the overall percent increases, only 3% of students completed the prompt correctly for both time points.

We grouped the students by the change in coding groups between the two time points, increasing, decreasing, or stayed the same (tied), these will be referred to as ranks. Overall, there is a significant difference between the changes in the ranks for the OCLUE students and Traditional students with a medium effect size as shown in Table 4 (χ^2 (2) = 24.141, $p <$ 0.001, Cramer's V = 0.26). The specific differences between the ranks from Time Point One to Time Point Two show that

Table 2. Time Point One: Percent of Students Who Drew All Plausible Arrows and Predicted a Plausible Product[a]

Prompt	OCLUE (count/total *n*)	Traditional (count/total *n*)	χ^2 (df = 1)	*p*-value	Cramer's V
A	40% (64/165)	11% (20/187)	71.449$^{\alpha}$	<0.001	0.56
B	22% (37/167)	8% (15/185)	24.024$^{\alpha}$	<0.001	0.41
C	6% (10/162)	1% (2/184)			
D	17% (26/164)	3% (6/189)			

[a] For all chi-square analysis $\alpha = 0.01$; (α) students in OCLUE outperformed those in Traditional.

Table 3. Time Point Two: Percent of Students Who Drew All Plausible Arrows and Predicted a Plausible Product[a]

Prompt	OCLUE (count/total *n*)	Traditional (count/total *n*)	χ^2 (df = 1)	*p*-value	Cramer's V
A	64% (106/165)	21% (38/187)	87.479$^{\alpha}$	<0.001	0.57
B	38% (64/167)	10% (18/185)	50.479$^{\alpha}$	<0.001	0.51
C	18% (29/162)	1% (2/184)	27.991$^{\alpha}$	<0.001	0.65
D	53% (88/164)	21% (38/183)	48.835$^{\alpha}$	<0.001	0.48

[a] For all chi-square analysis $\alpha = 0.01$; (α) students in OCLUE outperformed those in Traditional.

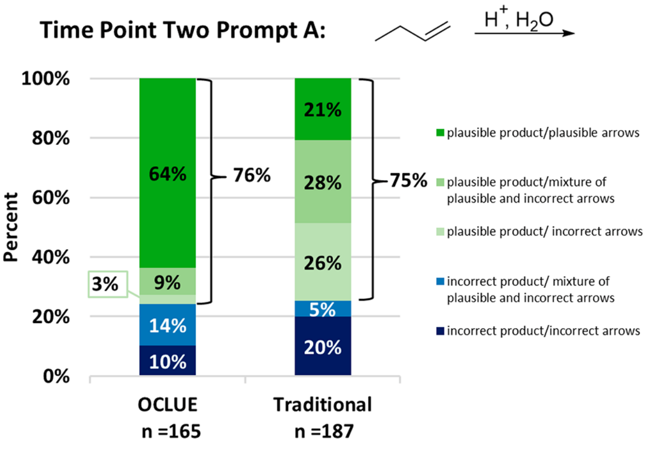


Figure 7. Prompt A: Percent of students who drew all mechanistic steps correctly, students who drew some mechanistic steps correctly, students who drew no mechanistic steps correctly, and students who got the incorrect product based on the courses they took.

very few students in OCLUE are decreasing in their rank over time (8%), while a quarter of the Traditional students fell to a lower coding group by Time Point Two.

Research Question 2: In What Ways Are Responses from Students Who Are Enrolled in a Transformed Organic Course Similar or Different to an Equivalent Group of Students from a Traditional Organic Course for an Unfamiliar Reaction?

Thus far we have explored reactions (prompt A–D) that we know are familiar to students. However, to our knowledge, prompt E is an unfamiliar reaction that students have not seen before. Nevertheless, plausible products for this reaction can be predicted if students use mechanistic arrows as a prediction tool. For example, one might expect that students would use the lone pair on the alcohol oxygen to initiate intramolecular attack at a carbonyl carbon, producing a tetrahedral intermediate, followed by loss of the best leaving group (chloride) to product a lactone.

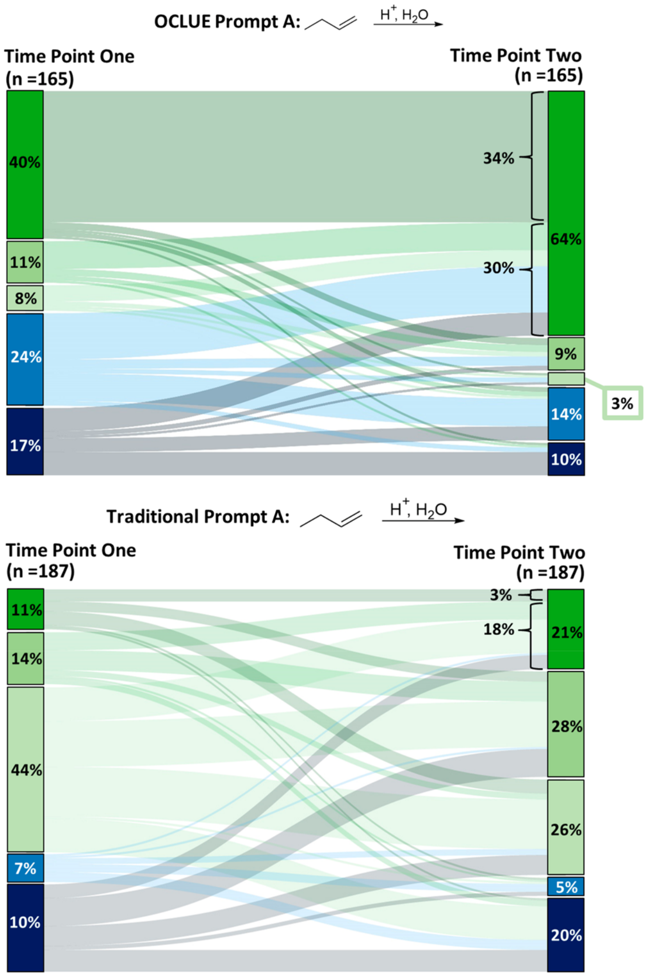


Figure 8. Prompt A: Percent of students who drew all mechanistic steps correctly, students who drew some mechanistic steps correctly, students who drew no mechanistic steps correctly, and students who got the incorrect product based in the courses they took.

Prompt E was administered only at Time Point Two near the end of the two-semester sequence to minimize the chance that students would have seen it before in an earlier administration. Analysis of the student responses indicated that there was a significant difference between the percent of students who proposed a plausible structure for the product between the students in OCLUE (45%) and those in Traditional (8%) as shown in Figure 9. This difference is significant (χ^2 (1) = 60.009, $p < 0.001$, Cramer's V = 0.42, medium-large effect size). The number of correct responses in our previous study was 9%, which is similar to the Traditional student cohort.[23]

https://doi.org/10.1021/acs.jchemed.1c00099

Table 4. Prompt A Change in Ranks for Students from Time Point One to Time Point Two[a]

Ranks	OCLUE (count/total n)	Traditional (count/total n)	χ^2 (df = 1)	p-value	Cramer's V
Positive	40% (67/165)	47% (87/187)	24.141	<0.001	0.26
Tie	52% (85/165)	30% (56/187)			
Negative	8% (13/165)	23% (44/187)			

[a] For all chi-square analysis α = 0.01.

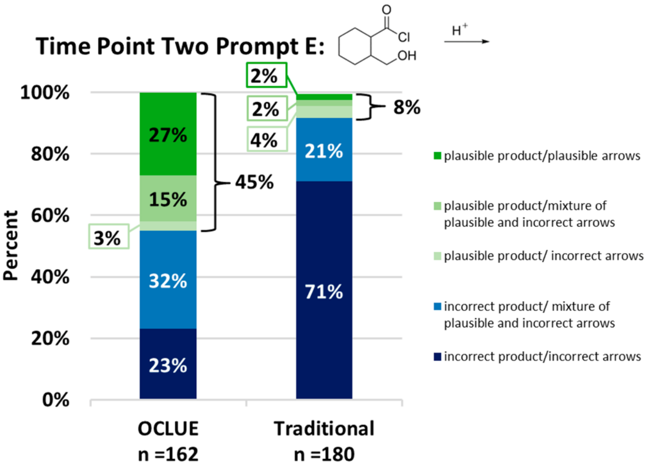


Figure 9. Percent of students who attempted drawing a mechanism but still drew an incorrect product and students who only drew an incorrect product with no arrows.

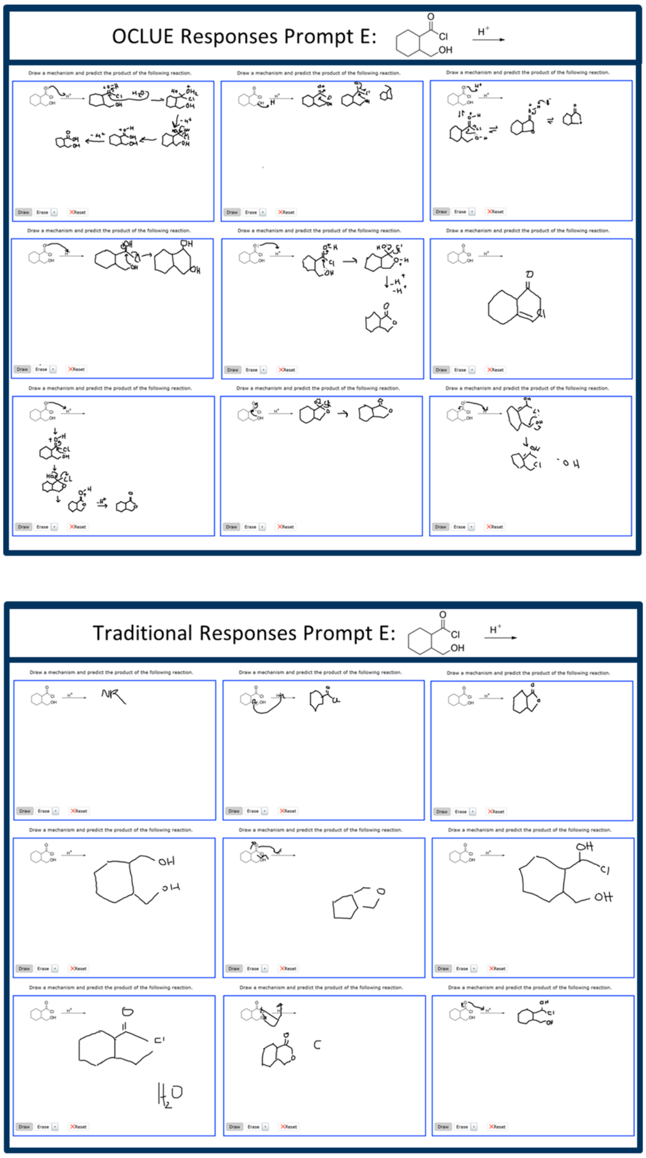


Figure 10. Screenshot from beSocratic in which student responses are shown in a grid for both OCLUE (top) and Traditional (bottom).

Just as with the more familiar prompts A–D, OCLUE students were more likely to draw all mechanistic steps correctly than Traditional students for the unfamiliar prompt. Analysis of the mechanistic approach taken by students is shown in Figure 9. The numbers of Traditional students who were able to complete this task appropriately are too small to compare statistically, but again we see a larger percent of students from OCLUE sections both drawing appropriate mechanisms and producing a plausible product.

This finding further supports the idea that students who habitually draw mechanisms as part of course expectations to predict outcomes for reactions are more likely to produce plausible products.

Further analysis of the students who produced an implausible product also showed differences between the two cohorts. For example, there were far fewer students in OCLUE who did not attempt some kind of plausible mechanism (23% OCLUE vs 71% Traditional) (Figure 9). As an illustration of this phenomenon we show Figure 10, which is a screenshot from beSocratic in which student responses are shown in a grid for both OCLUE and Traditional. These responses are representative samples of each cohort and show qualitatively the different approaches between the two student cohorts. We can see from inspection that OCLUE students are more likely to draw mechanistic arrows and intermediates than Traditional students.

How Do the Mechanism Attempts from the Earlier Studies Compare to the Current Study?

The major goal of this study is to characterize ways in which students in a transformed and a traditional course respond to the prompts from the original study; however, there is an opportunity here to briefly compare how students from the original and present studies respond. The students in the original study and the students in this study are comparable (demographic information is provided in Table S21), but we have chosen only to compare the original students and the traditional cohort because they participated in a similar course environment. In the previous study, between 30–60% of students simply wrote down a product and did not attempt to draw a mechanism.[22] Students were counted as not drawing a mechanism if they did not draw any arrows or intermediate structures when answering the prompt. Here we find a very similar pattern of responses, depending on the prompt: between 27 and 68% did not attempt to draw a mechanism, as shown in Table S22. Additionally, just as in the original study, where 15–20% of students drew their arrows after they had already drawn their product for all prompts,[22] there were students in the Traditional course who also drew a product before drawing a mechanism for between 9 and 20% of their responses. It is striking that two different groups of students, in two different universities, 10 years apart provide responses that appear quite similar.[22]

DISCUSSION

Our goal in this study was to characterize the ways in which students in two different organic chemistry courses, Traditional and OCLUE, complete a set of mechanism tasks that were first studied over 10 years ago in a Traditional course setting. Our findings indicate that after one semester there is little difference in the ability of students from either course to draw a plausible product for familiar reactions. By the end of two semesters (Time Point Two) both cohorts had improved in their prediction of products, and for prompts A, B, and D at least 50% of students were able to produce a plausible product. Both cohorts were less successful in predicting the outcome of prompt C, which involves deprotonation of an alkyne, which appeared not well remembered, and without sufficient knowledge students would not be able to predict this outcome. Again, any significant differences in product prediction were marked by small effect sizes. However, the approach that students use to reach the known product does seem to differ. At both time points, OCLUE students were significantly more likely to draw plausible arrows to produce a chemically feasible product, than were Traditional students. By the end of the second semester this difference was significant with a large effect size for all familiar tasks (A–D).

Perhaps we should not be surprised that many students opt not to draw a mechanism for a known reaction. The idea that organic chemistry requires a huge amount of memorization is well recognized both as part of student lore, and has been documented in a number of research studies.[51,52] As Graulich noted "It is evident from the current studies that students still rely heavily on rote-memorization and that traditional give-the-product exercises are frequently solved without a deeper understanding."[10] It is almost certainly not the intent of organic instructors that this should be the case, but it is clear that this idea is quite pervasive, and as such may impact how students address the work in the course. In a recent study (in review) when students were asked in an open response survey about what they were assessed on in an organic course, a majority of Traditional students responded that they were assessed on their ability to memorize a large amount of material.

Perhaps what is more interesting is the larger percentage of OCLUE students who do attempt to draw mechanisms, even for familiar reactions. It is our hypothesis that the willingness to draw mechanisms stems from the design and enactment of the course itself, and we believe that there are two particular characteristics that contribute to this. The first is that students routinely (at least two or three times a week) complete three-dimensional homework and recitation tasks, where (when relevant) reaction mechanisms and causal mechanistic explanations are a central focus. These formative assessments account for approximately 45% of the course grade and are graded based on student participation rather than correctness, with the intent of fostering a "safe space" for students to make mistakes without penalty. Such tasks are designed to support the connection of resources that serve to support both the explanation and the model. For example, students might be asked to construct an explanation for both how and why a reaction proceeds as it does (using resources such as Coulombic interactions and electron density distributions), and then asked to construct a mechanism.[35,36] That is, for students the explanation and mechanism are typically part of the same activity, which may help consolidate the connection of cognitive resources. One of the major findings of studies on learning is that the development of expertise involves supporting the connection of knowledge into a coherent framework.[53,54] We know that experts' knowledge frameworks are more likely to involve connections, so that the knowledge is contextualized and useful. Indeed, the efforts to design learning experiences around "big ideas" are intended to help students connect their knowledge in order to develop more expert-like frameworks.[11,42,55−61] These connections are established by helping students use their knowledge—for example by constructing models and explanations. Indeed, in the study discussed earlier (in review) a majority of OCLUE students reported that they were assessed on how they use their knowledge, rather than what they know.

A consequence of this design is that the connection between the resources required to construct an explanation and those to construct a mechanism in a meaningful way (i.e., from source to sink) is made quite explicit to students, and we believe, allows a greater number of students to attempt to draw a mechanism for a reaction with which they were unfamiliar.

The second characteristic is that during most of the course activities students are not penalized for making mistakes, on homework or in recitation. Indeed, for many activities we explicitly ask them to try to figure out what is happening. Students receive full points for a "good faith effort" and are therefore more likely to try to address the task at hand. Indeed, as Bodner once said, when solving problems it is important to "try something" then "try something else".[62] Often the first step is addressing any unknown task is that first step, and we know that some students are unwilling to begin a task if they do not know where it is going.[63,64]

Regardless of *why* OCLUE students are more likely to use arrow pushing mechanisms to predict outcomes of reactions, when we look at the responses for the unfamiliar reaction E, we see that OCLUE students significantly outperform their traditional peers in drawing plausible arrows and predicting a plausible product, 27% and 2%, respectively. Just as we found earlier, students who use mechanistic arrows are more likely to predict plausible products, than those who do not. The difference between this study and the earlier ones is that, in the earlier studies (and similar) to Traditional students, only 9% of students were able to predict a reasonable product, compared to 45% of OCLUE students in this study, 42% of whom used at least one plausible mechanistic arrow.

In summary, these two matched cohorts are of similar ability, when it comes to predicting a product for a familiar reaction, but there is a significant difference in the outcome for an unfamiliar reaction. We ascribe this difference to two main factors: (1) the explicit treatment of mechanisms and explanations as complementary practices, that use the same cognitive resources, and (2) the learning environment of the OCLUE course in which students are given the freedom to make errors without penalty, which may make students more willing to attempt a response in the face of a new task.

Implications for Teaching

For those who value mechanistic reasoning in organic chemistry but are unwilling or unable to completely transform a course, there are several actionable approaches that may improve outcomes.

(1) Emphasize the explanatory and predictive power of mechanistic reasoning early and often in the context of both using mechanistic arrows and constructing written explanations about how and why reactions occur in a particular way. If we want students to use mechanistic arrows and causal mechanistic reasoning, we have to ensure that students understand what their purpose is. As discussed earlier we have previously shown that students who are able to construct causal mechanistic written explanations (that is, for both how and why reactions occur) are also more likely to draw appropriate mechanistic arrow models.[35,36] The act of constructing an explanation can help link and consolidate ideas leading to a more expert-like understanding.

(2) Provide frequent opportunities for students to practice such model construction and use, with formative assessment tasks that count toward the ultimate grade, but that are rewarded for effort not correctness. Obviously, these formative tasks must come with appropriate feedback, which can be provided in various ways. In OCLUE, because of the large numbers involved, feedback cannot be provided to individual students—but rather homework responses are used to drive the next class discussion. These activities provide students with opportunities to explore their ideas, and then they can go back and reconstruct answers that need more work. Students also have the opportunity to engage with these idea during their weekly recitation and receive feedback on their work from graduate teaching assistants. It has also been shown that feedback is more likely to be addressed on nongraded assessments.[65]

(3) To ensure that students get the message such tasks should be incorporated in summative examinations. That is students should be asked to explain as well as draw products for reactions.

(We note that there are other important aspects of organic chemistry not discussed here such as constructing an argument from spectroscopic evidence, developing syntheses of desired compounds, and predicting kinetic and thermodynamic outcomes for reactions, using multicomponent reasoning, but these are not the focus of this report.)

In summary, we believe that the goal of education should be to provide students with the tools to use their knowledge in a productive manner, to think flexibly, and to take that leap into the unknown. To do this requires more than changing **how** we teach, it also requires us to think about what is important and change **what** we teach and how we assess what students know and can do. If students expect that success can be achieved by memorization, then we should not be surprised when this is what they do.

Limitations

As noted in the Results section, the OCLUE curriculum was developed by a team that includes our research group, and as such there is a potential for bias not only in the analysis of the results but in their interpretation. We have attempted to minimize this bias as discussed earlier, but readers should be aware that it exists.

This study was a repeat of a previous study with different student groups. However, the same prompts that were administered in the original study were also administered in this new study. The original study included multiple familiar prompts but only one unfamiliar prompt which limits the generalizability to more complex unfamiliar questions. Additionally, the original software (OrganicPad)[41] we used to deliver the tasks was obsolete. In the original study the students used tablet computers that were provided for this purpose in a laboratory setting that was not connected directly to the course through which the material was learned. In the new study, students completed the tasks online on the beSocratic system for homework. That being said, there is a remarkable similarity in the patterns of responses for Traditional students. Additionally, in the present study OCLUE students do use the beSocratic system for homework on a weekly basis, which may mean that they were more familiar with the system. However, most (over 95%) of these students in both OCLUE and Traditional had taken other courses in which the beSocratic system is used for homework, and they had also previously completed several other extra credit tasks on this system.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.1c00099.

Study participant demographics and inter-rater reliability (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Melanie M. Cooper − *Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States*; orcid.org/0000-0002-7050-8649; Email: mmc@msu.edu

Authors

Samantha K. Houchlei − *Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States*

Rosalyn R. Bloch − *Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jchemed.1c00099

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 1043707 (1420005), DUE 1122472 (1341987), and 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) Seymour, E.; Hunter, A.-B. *Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education*; Springer International Publishing: New York, 2019.

(2) Stowe, R. L.; Cooper, M. M. Practicing What We Preach: Assessing "Critical Thinking" in Organic Chemistry. *J. Chem. Educ.* **2017**, *94* (12), 1852−1859.

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

(4) Gerlach, K.; Trate, J.; Blecking, A.; Geissinger, P.; Murphy, K. Valid and Reliable Assessments To Measure Scale Literacy of Students in Introductory College Chemistry Courses. *J. Chem. Educ.* **2014**, *91* (10), 1538−1545.

(5) 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.

(6) 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* (1), 64−77.

(7) 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.

(8) 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.

(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) Graulich, N. The Tip of the Iceberg in Organic Chemistry Classes: How Do Students Deal with the Invisible? *Chem. Educ. Res. Pract.* **2015**, *16* (1), 9−21.

(11) *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*; National Academies Press: Washington, D.C., 2012.

(12) *Visualization in Science Education*; Gilbert, J., Ed. Models and modeling in science education; Springer: Dordrecht, 2005.

(13) Finkenstaedt-Quinn, S. A.; Watts, F. M.; Petterson, M. N.; Archer, S. R.; Snyder-White, E. P.; Shultz, G. V. Exploring Student Thinking about Addition Reactions. *J. Chem. Educ.* **2020**, *97* (7), 1852−1862.

(14) *Helping Students Make Sense of the World Using Next Generation Science and Engineering Practices*; National Science Teachers Association, 2017. DOI: 10.2505/9781938946042.

(15) 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.

(16) Cooper, M. M.; Corley, L. M.; Underwood, S. M. An Investigation of College Chemistry Students' Understanding of Structure-Property Relationships. *J. Res. Sci. Teach.* **2013**, *50* (6), 699−721.

(17) Becker, N. M.; Noyes, K.; Cooper, M. M. Characterizing Students' Mechanistic Reasoning about London Dispersion Forces. *J. Chem. Educ.* **2016**, *93* (10), 1713−1724.

(18) Noyes, K.; Cooper, M. M. Investigating Student Understanding of London Dispersion Forces: A Longitudinal Study. *J. Chem. Educ.* **2019**, *96* (9), 1821−1832.

(19) 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.

(20) Krist, C.; Schwarz, C. V.; Reiser, B. J. Identifying Essential Epistemic Heuristics for Guiding Mechanistic Reasoning in Science Learning. *Journal of the Learning Sciences* **2019**, *28* (2), 160−205.

(21) Hammer, D. Student Resources for Learning Introductory Physics. *Am. J. Phys.* **2000**, *68* (S1), S52−S59.

(22) 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.

(23) Grove, N. P.; Cooper, M. M.; Cox, E. L. Does Mechanistic Thinking Improve Student Success in Organic Chemistry? *J. Chem. Educ.* **2012**, *89* (7), 850−853.

(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) Dood, A. J.; Dood, J. C.; Cruz-Ramírez de Arellano, D.; Fields, K. B.; Raker, J. R. Using the Research Literature to Develop an Adaptive Intervention to Improve Student Explanations of an S_N1 Reaction Mechanism. *J. Chem. Educ.* **2020**, *97* (10), 3551−3562.

(26) Mistry, N.; Nicholson, S. Investigating Students Understanding of Reaction Mechanisms from Performing Chemistry Experiments. *New Dir. Teach. Phys. Sci.* **2020**, *14*, 1.

(27) Carle, M. S.; Visser, R.; Flynn, A. B. Evaluating Students' Learning Gains, Strategies, and Errors Using OrgChem101's Module: Organic Mechanisms—Mastering the Arrows. *Chem. Educ. Res. Pract.* **2020**, *21* (2), 582−596.

(28) Visser, R.; Flynn, A. B. What Are Students' Learning and Experiences in an Online Learning Tool Designed for Cognitive and Metacognitive Skill Development? *CELT* **2019**, *11*, 11.

(29) 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.

(30) Graulich, N.; Caspari, I. Designing a Scaffold for Mechanistic Reasoning in Organic Chemistry. *Chemistry Teacher International* **2021**, *3* (1), 19−30.

(31) Watts, F. M.; Schmidt-McCormack, J. A.; Wilhelm, C. A.; Karlin, A.; Sattar, A.; Thompson, B. C.; Gere, A. R.; Shultz, G. V. What Students Write about When Students Write about Mechanisms: Analysis of Features Present in Students' Written Descriptions of an Organic Reaction Mechanism. *Chem. Educ. Res. Pract.* **2020**, *21* (4), 1148−1172.

(32) Schmidt-McCormack, J. A.; Judge, J. A.; Spahr, K.; Yang, E.; Pugh, R.; Karlin, A.; Sattar, A.; Thompson, B. C.; Gere, A. R.; Shultz, G. V. Analysis of the Role of a Writing-to-Learn Assignment in Student Understanding of Organic Acid-Base Concepts. *Chem. Educ. Res. Pract.* **2019**, *20* (2), 383−398.

(33) Pashler, H.; Bain, P. M.; Bottge, B. A.; Graesser, A.; Koedinger, K.; McDaniel, M.; Metcalfe, J. Organizing Instruction and Study to Improve Student Learning. *Amer. Psychol. Assoc.* **2007**, 1.

(34) Cooper, M. M. Why Ask Why? *J. Chem. Educ.* **2015**, *92* (8), 1273−1279.

(35) Cooper, M. M.; Kouyoumdjian, H.; Underwood, S. M. Investigating Students' Reasoning about Acid-Base Reactions. *J. Chem. Educ.* **2016**, *93* (10), 1703−1712.

(36) 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.

(37) Talanquer, V. Importance of Understanding Fundamental Chemical Mechanisms. *J. Chem. Educ.* **2018**, *95* (11), 1905−1911.

(38) Weinrich, M. L.; Talanquer, V. Mapping Students' Modes of Reasoning When Thinking about Chemical Reactions Used to Make a Desired Product. *Chem. Educ. Res. Pract.* **2016**, *17* (2), 394−406.

(39) Talanquer, V. Progressions in Reasoning about Structure-Property Relationships. *Chem. Educ. Res. Pract.* **2018**, *19* (4), 998−1009.

(40) 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.

(41) Cooper, M. M.; Grove, N. P.; Pargas, R.; Bryfczynski, S. P.; Gatlin, T. OrganicPad: An Interactive Freehand Drawing Application for Drawing Lewis Structures and the Development of Skills in Organic Chemistry. *Chem. Educ. Res. Pract.* **2009**, *10*, 296−301.

(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) Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20* (1), 37−46.
(44) IBM Corp. *SPSS Statistics for Windows*; IBM Crop., Armonk, NY, USA, 2017.
(45) Bryfczynski, S. P. *BeSocratic: An Intelligent Tutoring System for the Recognition, Evaluation, and Analysis of Free-Form Student Input.* Ph.D. Dissertation, Clemson University, South Carolina, 2010.
(46) Wade, L. G.; Simek, J. W. *Organic Chemistry*, 9th ed.; Pearson: New York City, NY, 2017.
(47) Raker, J.; Holme, T.; Murphy, K. The ACS Exams Institute Undergraduate Chemistry Anchoring Concepts Content Map II: Organic Chemistry. *J. Chem. Educ.* **2013**, *90* (11), 1443−1445.
(48) Stowe, R. L.; Cooper, M. M. Practicing What We Preach: Assessing "Critical Thinking" in Organic Chemistry. *J. Chem. Educ.* **2017**, *94*, 8.
(49) Landis, J. R.; Koch, G. G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33* (1), 159.
(50) Green, S.; Salkind, N. *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data*; Pearson Education Inc: Boston, MA, USA, 2010.
(51) Grove, N. P.; Bretz, S. L. Perry's Scheme of Intellectual and Epistemological Development as a Framework for Describing Student Difficulties in Learning Organic Chemistry. *Chem. Educ. Res. Pract.* **2010**, *11* (3), 207−211.
(52) Grove, N. P.; Lowery Bretz, S. A Continuum of Learning: From Rote Memorization to Meaningful Learning in Organic Chemistry. *Chem. Educ. Res. Pract.* **2012**, *13* (3), 201−208.
(53) *How People Learn: Brain, Mind, Experience, and School*, Expanded ed.; National Academies Press: Washington, D.C., 2000; p 9853. DOI: 10.17226/9853.
(54) Committee on How People Learn II: The Science and Practice of Learning; Board on Behavioral, Cognitive, and Sensory Sciences; Board on Science Education; Division of Behavioral and Social Sciences and Education; National Academies of Sciences, Engineering, and Medicine. *How People Learn II: Learners, Contexts, and Cultures*; National Academies Press: Washington, D.C., 2018; p 24783. DOI: 10.17226/24783.
(55) 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.
(56) Sevian, H.; Talanquer, V. Rethinking Chemistry: A Learning Progression on Chemical Thinking. *Chem. Educ. Res. Pract.* **2014**, *15* (1), 10−23.
(57) College Board. *Advances in AP: AP Chemistry*; College Board, 2021.
(58) College Board. *College Board Standards for College Success: Science*; College Board, 2009.
(59) Holme, T.; Luxford, C.; Murphy, K. Updating the General Chemistry Anchoring Concepts Content Map. *J. Chem. Educ.* **2015**, *92* (6), 1115−1116.
(60) Holme, T.; Murphy, K. The ACS Exams Institute Undergraduate Chemistry Anchoring Concepts Content Map I: General Chemistry. *J. Chem. Educ.* **2012**, *89* (6), 721−723.
(61) 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.
(62) Bodner, G. M.; Herron, J. D. Problem-Solving in Chemistry. In *Chemical Education: Towards Research-based Practice*; Gilbert, J. K., Jong, O., Justi, R., Treagust, D. F., Driel, J. H., Eds.; Kluwer Academic Publishers: Dordrecht, 2003; Vol. *17*, pp 235−266. DOI: 10.1007/0-306-47977-X_11.
(63) Sandi-Urena, S.; Cooper, M. M.; Gatlin, T. A.; Bhattacharyya, G. Students' Experience in a General Chemistry Cooperative Problem Based Laboratory. *Chem. Educ. Res. Pract.* **2011**, *12* (4), 434−442.
(64) Schoenfeld, A. H. Explorations of Students' Mathematical Beliefs and Behavior. *J. Res. Math. Educ.* **1989**, *20* (4), 338−355.
(65) Winstone, N. E.; Nash, R. A.; Rowntree, J.; Parker, M. 'It'd Be Useful, but I Wouldn't Use It': Barriers to University Students' Feedback Seeking and Recipience. *Stud. High Educ.* **2017**, *42* (11), 2026−2041.