



Cite this: DOI: 10.1039/d0rp00298d

Investigating students' reasoning over time for case comparisons of acyl transfer reaction mechanisms

Field M. Watts, ^a Ina Zaimi, ^a David Kranz, ^b Nicole Graulich ^b and Ginger V. Shultz ^{*a}

Reasoning about organic chemistry reaction mechanisms requires engagement with multiple concepts and necessitates balancing the relative influence of different chemical properties. A goal of organic chemistry instruction is to support students with engaging in this type of reasoning. In this study, we describe our use of case comparison problems to elicit students' reasoning about acyl transfer reaction mechanisms across a semester. Using an instrumental case study methodology, we analysed three students' reasoning across three time points: in a pre-interview at the beginning of the semester, on their written responses to one implementation of an in-class scaffold activity, and in a post-interview near the middle of the semester. Through the theoretical lens of Hammer's resources framework, we analysed the resources that students activated when approaching the case comparison problems. We characterized how students used each resource to support their reasoning, alongside characterizing how students weighed the different resources they activated. Our findings indicate that the case comparison problems activated a number of resources for each student across the time points by encouraging students to relate the surface-feature differences between reactions with the associated underlying properties. Students generally used resources, such as resonance and steric effects, in similar ways to support their reasoning across the time points. The study also illustrates the range in students' abilities to weigh multiple conceptual influences and how this ability might change across the semester. This case study has implications for future research exploring how students reason with multiple concepts and for instructors seeking to implement activities that support students' reasoning with case comparison problems.

Received 1st October 2020,
Accepted 19th December 2020

DOI: 10.1039/d0rp00298d

rsc.li/cerp

Introduction

Learning organic chemistry requires students to engage with core conceptual ideas that connect a large number of different reaction types and mechanisms. Hence, students need to engage with learning strategies that promote process-oriented reasoning and problem-solving skills over product-oriented, rote learning (Grove and Bretz, 2012; Graulich, 2015). However, research shows that students often approach learning in organic chemistry by systematically memorizing specific conceptual relationships, reactions, and mechanisms rather than using process-oriented understandings of conceptual ideas (Anderson and Bodner, 2008; Kraft *et al.*, 2010; Grove and Bretz, 2012; Anzovino and Bretz, 2015; Webber and Flynn, 2018).

Therefore, it is necessary for researchers and instructors to understand and promote students' reasoning about mechanisms in organic chemistry. In this work, we describe a case study to explore students' reasoning for case comparisons, which are problems that involve posing a question alongside two similar mechanisms that have purposefully designed contrasting features. Herein, we present our analysis of students' abilities to consider and weigh different concepts for case comparison problems about acyl transfer reactions at three time points in a second-semester organic chemistry laboratory course.

Reasoning in organic chemistry

Practicing organic chemists use mechanisms as explanatory and predictive tools for describing how and why reactions occur (Goodwin, 2012). Instruction regarding reaction mechanisms typically involves presenting the electron-pushing formalism for this purpose. However, research shows that many students do not necessarily use mechanisms as intended or understand

^a Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA. E-mail: gshultz@umich.edu

^b Justus-Liebig-University Giessen, Institute of Chemistry Education, Heinrich-Buff-Ring 17, 35392 Giessen, Germany

the physical meaning associated with the electron-pushing formalism (Bhattacharyya and Bodner, 2005; Ferguson and Bodner, 2008; Grove *et al.*, 2012a; Grove *et al.*, 2012b). Furthermore, research demonstrates that students have challenges interpreting the underlying properties that are communicated by other representations, such as molecular structures, with a tendency to focus on their surface features (Domin *et al.*, 2008; Kraft *et al.*, 2010; McClary and Talanquer, 2011; Cruz-Ramírez De Arellano and Towns, 2014; Anzovino and Bretz, 2015; Galloway *et al.*, 2017, 2019; Graulich and Bhattacharyya, 2017).

Alongside the evidence suggesting students' limited understanding of the electron-pushing formalism and other representations in organic chemistry, several studies provide evidence of students' conceptual understanding related to organic reaction mechanisms (Graulich, 2015). Recent studies focus on specific reaction types typically taught within organic chemistry, including acid-base reactions (Cooper *et al.*, 2016; Dood *et al.*, 2018; Schmidt-McCormack *et al.*, 2019; Petterson *et al.*, 2020), addition reactions (Finkenstaedt-Quinn *et al.*, 2020), substitution reactions (Cruz-Ramírez De Arellano and Towns, 2014; Caspari *et al.*, 2018a; Bodé *et al.*, 2019; Caspari and Graulich, 2019; Crandell *et al.*, 2020; Dood *et al.*, 2020), and elimination reactions (Cruz-Ramírez De Arellano and Towns, 2014; Caspari and Graulich, 2019), among other reaction types (Galloway *et al.*, 2017; Bhattacharyya and Harris, 2018; Caspari *et al.*, 2018b; Watts *et al.*, 2020). These researchers describe how students apply conceptual understanding to different tasks related to organic reaction mechanisms. In particular, these studies demonstrate the range in how students can apply their conceptual understanding of key concepts in organic chemistry (*e.g.*, acid-base chemistry, resonance, nucleophilicity, *etc.*) differently across reaction types. Research of students' understanding of specific concepts, including charge, resonance, and nucleophilicity, specifically demonstrates that students tend to focus on the structural features of molecules over their function (Ferguson and Bodner, 2008; Cruz-Ramírez De Arellano and Towns, 2014; Anzovino and Bretz, 2015; Petterson *et al.*, 2020; Xue and Stains, 2020). While many studies explore the range of concepts and understandings students consider, it is necessary to explicitly understand how students weigh multiple conceptual considerations in their reasoning about organic reaction mechanisms.

The existing research demonstrates that students have a range of understanding of core concepts in organic chemistry and that students often treat mechanism tasks as product-oriented exercises. Hence, there is an ongoing effort in the chemistry education research community to suggest better ways to teach organic reaction mechanisms. In particular, it is important to know how students use multiple concepts when considering problems involving mechanisms. Therefore, it is necessary to research approaches that elicit students' reasoning and support students as they engage with connecting underlying conceptual principles to mechanistic steps occurring during a reaction. It is particularly valuable to research approaches beyond traditional mechanisms tasks, such as predicting products or drawing mechanistic arrows, since such problems are not as effective at eliciting students' reasoning

(DeCocq and Bhattacharyya, 2019). In this study, we use case comparison problems designed to support and encourage mechanistic reasoning and explore the development of students' abilities to consider and weigh multiple concepts during their reasoning.

Case comparisons to elicit students' reasoning

Case comparison problems better elicit students' reasoning as compared to problems involving a single case (Alfieri *et al.*, 2013; Caspari and Graulich, 2019). In recent work, Graulich and Schween (2018) describe the relationships between case comparisons and the epistemic practices of organic chemists and discuss how case comparison tasks can support students' abilities to develop, apply, and expand upon their conceptual understanding. Caspari *et al.* (2018a) demonstrate the usefulness of case comparisons for eliciting students' reasoning in an interview setting, with scaffolding questions that support students' construction of more complex explanations. In Bodé *et al.*'s (2019) research, students demonstrate their reasoning on a case comparison exam question. Notably, Bodé *et al.* (2019) found that students with more sophisticated reasoning made direct comparisons between structures. However, many students did not make comparisons for all explicit and implicit features in the reactions.

Similarly, in a study using eye-tracking, Rodemer *et al.* (2020) found that students tended to focus their attention on the reactants instead of the products when solving case comparisons. Rodemer *et al.* (2020) also found that advanced students were faster and had increased focus on relevant chemical structures compared to beginner students. Together, these studies demonstrate the usefulness of case comparisons for eliciting students' reasoning and point to a need for further research into how students consider each part of a case comparison problem. It is necessary to understand how students use concepts when solving case comparisons, if the concepts students consider change during a semester, and how to elicit such reasoning in a classroom setting.

Prior research on case comparison problems for organic reaction mechanisms using an instructional scaffold is reported by Caspari and Graulich (2019), who described interviews in which students compared activation energies for the leaving group departure step of similar E1 reactions. The interviewer first asked students to reason about the problem without the scaffold, followed by asking students to complete the same problem using an instructional scaffold. The researchers designed this scaffold to help students engage in reasoning with multiple variables by separating structural differences from mechanistic changes and by delineating the different influences each structural difference has on each change. By asking students to complete the same case comparison problem with and without the scaffold, the researchers found that the scaffold successfully built upon the reasoning structures students exhibited without the scaffold. Furthermore, they found that students' use of the scaffold was correlated with an increase in the number of influences students considered in their reasoning. To build upon this work, our goal was to identify

if students' consideration of multiple properties changes during a semester for similar case comparison tasks, including students' reasoning as presented on an in-class activity similar in structure to the scaffold used in the work by Caspari and Graulich (2019).

Theoretical framework

Hammer's resources framework

This research is guided by Hammer's (Hammer and Elby, 2000; Hammer *et al.*, 2004) resources framework that describes an approach towards understanding the cognitive structures people use to construct explanations. This framework is influenced by previous literature seeking to define units of cognition, including diSessa's (diSessa, 1993; diSessa and Sherin, 1998) "phenomenological primitives" and "coordination classes" and Thagard's (1989) "propositions." Within Hammer's framework, these fine-grained cognitive elements of knowledge are referred to as "resources" (Hammer and Elby, 2000; Hammer *et al.*, 2004). Resources are, generally, ideas held about a phenomenon that are neither right nor wrong, and which are activated within certain situations to construct explanations. Activated resources can then be deemed productive or unproductive, depending on how the person relates the resources to the problem at hand. The resources framework contrasts with frameworks that suggest conceptions of a phenomenon are stable, fully formed ideas that are either correct or incorrect. Within these frameworks, conceptions are the cognitive units a person uses to build an explanation. Hammer suggests that conception-based frameworks do not accommodate the flexibility often observed in people's reasoning. That is, a conceptions framework does not explain situations in which a person seems to significantly alter a conception or misconception when encountering similar problems. Furthermore, the resources framework suggests that the resources activated when constructing explanations for similar phenomena differ depending on the situation, positing that the resources people activate for a particular phenomenon are not immutable. Since people use activated resources to construct explanations, the framework is useful for understanding how people's explanations for similar phenomena might change across time.

The resources framework provides a way for understanding how students construct explanations while engaging in organic chemistry case comparison reaction mechanism problems. Students have resources, or units of knowledge relating to structural features of a representation or concepts such as resonance, induction, or electronegativity, that they should be able to use to construct explanations. Students activate these resources for case comparison problems to explain the mechanistic question within the problem. The resources framework offers a way to understand what specific resources students activate and how they use ideas from activating multiple resources to produce explanations. Furthermore, the framework can be useful for understanding how students' reasoning changes, in terms of what resources a task activates at multiple time points and how students can use these resources when constructing explanations.

Research questions

This research aims to understand how students engage in reasoning and constructing explanations for case comparison problems about acyl transfer mechanism problems in organic chemistry. The goal is to qualitatively understand if and how students' reasoning for these problems might develop during the semester. This research addresses this goal by focusing on two aspects of students' reasoning for organic mechanism case comparison problems, namely what resources they activate and how they weigh activated resources across three time points. To address this goal, we seek to answer the following questions:

RQ1. What resources do students activate when considering the case comparison problems and how do the resources students activate change across time?

RQ2. How do students weigh resources when constructing explanations for the case comparison problems across time?

Methods

Instrumental case study methodology and research design

This research presents an exploratory, instrumental case study to investigate how students engage with case comparison problems during a second-semester introductory organic chemistry course for majors and non-majors at the University of Michigan. Instrumental case studies are those in which specific cases are used to understand a phenomenon, in contrast to traditional case studies that seek to understand something about the cases themselves (Grandy, 2012). In this research, we are studying the phenomenon of how students' engagement with case comparison problems develops throughout a second-semester organic chemistry laboratory course. Hence, this study aims not to understand the students themselves but to understand how their reasoning with case comparison problems develops. The research design involved collecting data at three times during the semester: a pre-interview at the beginning of the semester, completed worksheets from an in-class activity, and a post-interview in the weeks following the activity. Because students' reasoning is complex, particularly concerning how students activate and weigh multiple resources, the instrumental case study methodology is appropriate to provide a detailed, qualitative characterization of differences in how students respond to case comparison tasks. The case study methodology allows for a detailed analysis of how students reason with these problems across time points. Furthermore, the case study methodology is useful for guiding future research of students' reasoning on case comparison tasks (Grandy, 2012; Yin, 2014).

Setting

This study was conducted as part of a larger study at a research university in the Midwestern United States. The research was situated within the second-semester introductory organic chemistry laboratory course, which is offered separately from the lecture course at the study institution. The laboratory course consists of a weekly one-hour lecture taught by the course instructors and a four-hour laboratory taught by graduate

student instructors. Students worked with case comparisons in different aspects of the course, including in the laboratory itself and on assignments. The first-semester lecture and laboratory courses are both prerequisites for the second-semester laboratory course, and the second-semester lecture course is an advisory prerequisite or co-requisite. Students usually take the introductory organic chemistry sequence in their first or second year, followed by inorganic, analytical, and physical chemistry courses in later years for chemistry majors.

Participants

This study is part of a larger research effort for which we recruited nine students to participate. We selected three participants, given the pseudonyms Brooke, Violet, and Chad, to focus on for the instrumental case study. These students were selected due to the comparative level of detail in their responses across time points. Furthermore, the observed similarities and differences between these participants during the data collection process presented a rich set of data for which it was valuable to employ the instrumental case study methodology. These specific students were also chosen because their reasoning was representative of the reasoning observed during the data collection process across all nine participants. By focusing on three students, we can provide a detailed account of how they responded to the case comparison problems across time points. The pseudonyms assigned to participants do not reflect their race, ethnicity, gender, or other identities.

Data collection

The data collected for this study includes pre-interviews, students' responses to an in-class activity, and post-interviews, across which students responded to two different case comparison problems. All data collection procedures received Institutional Review Board approval for human subjects research, and all students consented to participate in data collection. The semester was fourteen weeks long, and the pre-interviews took place in the second and third weeks, the in-class activity took place in the sixth week, and the post-interviews took place in the ninth and tenth weeks.

Pre- and post-interviews. For the data used in this study, all three pre-interviews and one post-interview were conducted in-person and audio recorded. Documents annotated by both

the interviewer and interviewee were collected. Two post-interviews were conducted and recorded *via* video conferencing software, and documents were shared and annotated by participants using Google Drive. We conducted all interviews as think-aloud interviews, in which the interviewer instructed participants to verbalize their thinking as they considered the case comparison and responded to the guiding question described below (Herrington and Daubenmire, 2014). Students were provided with a copy of a periodic table and pK_a table to use if needed during each interview.

Both interview protocols involved asking students to reason through a case comparison problem involving acyl transfer reactions, shown as presented to students in Fig. 1(A) and (B). We asked students to decide which of the mechanistic steps shown in the case comparison has the lowest activation energy. The acyl transfer reactions chosen for this case comparison were selected because they required considering two variables: the different substituent on the electrophile (methoxy *vs.* chlorine) and the different nucleophiles (hydroxide *vs.* methylamine). Through the interviews, we aimed to capture how students considered the two variables in the contrasting cases as they decided about the relative activation energies for the represented steps.

The pre-interview protocol included a warm-up for the case comparison problem, in which the interviewer asked students to think-aloud while describing the two reactions. The interviewer asked probing questions to encourage students to identify similarities and differences between the reactions and describe the depicted mechanistic steps. After the warm-up, the interviewer asked students the guiding question, "Which reaction has a lower activation energy for the represented step? Make a prediction." Students were instructed to annotate the document as needed and to think-aloud as they responded to the question. Probing questions were asked after students completed their responses to clarify their intended meaning as well as their working definitions of activation energy and other concepts brought up during the task. Students appeared to have an appropriate working definition of the concept of activation energy. After the probing questions, the interviewer asked students to formulate a final statement of their response to the guiding question along with a summary of their reasoning. At the time of the pre-interview, students had not yet learned about acyl transfer reactions in their organic chemistry course but could

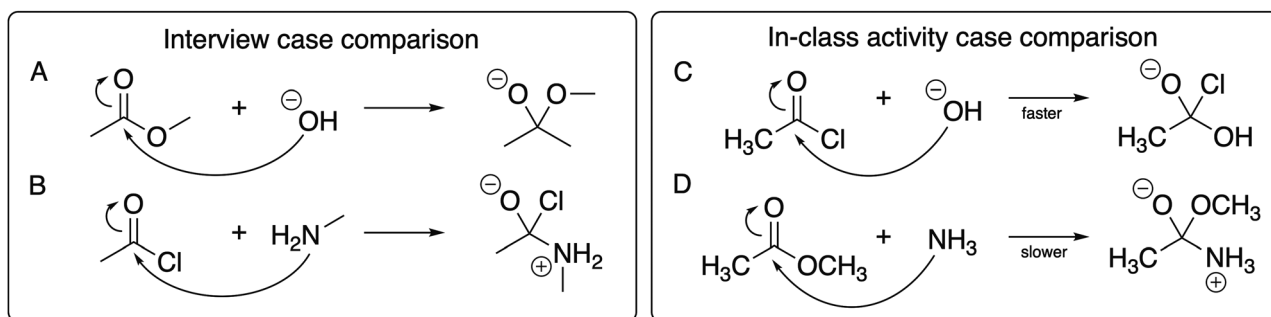


Fig. 1 The case comparison reactions for the pre- and post-interviews (reactions A and B) and for the in-class activity (reactions C and D), shown as presented to the students.

reasonably be expected to respond to the problem based on the material covered earlier in the course sequence.

The post-interview protocol presented students with the same case comparison problem and asked the same guiding question from the pre-interview. Probing questions were asked to clarify students' statements, and then interviewers asked students to formulate a final statement and summary of their reasoning. At the time of the post-interview, students had experience with acyl transfer reactions both in the lecture and laboratory.

In-class activity and implementation. The in-class activity was developed based upon the activity described in prior research reported by Caspari and Graulich (2019). The development of the activity followed the process described in detail by Graulich and Caspari (2020). The activity took the form of a scaffold implemented during the one-hour lecture component of the course to guide students through reasoning about organic chemistry case comparison problems. The activity asked students to explain relative reaction rates for contrasting cases of single mechanistic steps. During the implementation, students worked in small groups to complete the activity as the instructor and multiple graduate student instructors circulated through the lecture hall to address student questions. Afterward, activity worksheets were collected from students, including the three participants' completed activities. The acyl transfer reactions in the activity, shown as presented to students in Fig. 1(C) and (D), were similar to those used in the pre-interviews in that the case comparison required students to consider two variables in their reasoning: the different substituents on the electrophile (chlorine *vs.* methoxy) and the different nucleophiles (hydroxide *vs.* ammonia). We collected this data as a mid-point in the data collection process to provide insight into how students' reasoning developed from pre- to post-interview. The data served as an artefact of students' reasoning for a problem similar to the case comparison problem in the pre- and post-interviews. The complete activity is reproduced in Appendix 1.

Data analysis

Throughout the analysis, the research team focused on presenting what students were doing when responding to the case comparison problems rather than evaluating or assessing students' responses. As such, the three data sources were used to develop detailed, complete profiles of students' reasoning for each participant (Miles *et al.*, 2014). The interview recordings were transcribed verbatim, and we used the transcripts and recordings to write detailed descriptions of the resources students activated as they reasoned about the case comparison problems. Similarly, the in-class activity worksheets were scanned, and we used students' writings and annotations to write detailed descriptions of how students' reasoning was presented on the in-class activity. All profile descriptions incorporated annotations for how students used resources to guide their reasoning and how they weighed different resources when constructing explanations. These descriptions were read by another member of the research team and cross-referenced with the original data sources to ensure they accurately represented the students' reasoning.

The detailed descriptions of students' reasoning in each data source were then inductively coded (Miles *et al.*, 2014) by two research team members. They independently analysed the data for (1) the resources students activated when considering the problems and (2) how students weighed multiple resources, which we defined as the process of identifying resources to be more or less important for constructing an explanation in response to the guiding question. Afterward, the researchers discussed the inductive coding, organized the codes into themes, and developed finalized coding schemes (Miles *et al.*, 2014). The coding schemes are presented in Appendix 2. With the final coding schemes, the two researchers then re-coded the data, discussed the coding, and reached a consensus on the final set of codes applied to each data source. The detailed descriptions of students' reasoning and the coding results were then used to develop more concise profiles to represent the students' reasoning, presented in Appendix 3. We then used cross-case analysis, in which members of the research team discussed the coding results and profiles to make comparisons between students and across time points, to identify the key findings from the data. By using the cross-case analysis methodology, we were able to examine the similarities and differences across profiles to more deeply understand how the students approached the case comparison problems (Miles *et al.*, 2014). Discussions with the research team took place throughout the process of profile development, coding, and cross-case analysis to ensure the reliability and trustworthiness of the results.

Results and discussion

This study aims to understand how students approach case comparison problems focused on acyl transfer reactions during the second-semester organic laboratory course. The presented analysis focuses on the resources students activated when considering case comparison problems and how students weighed resources when constructing explanations. The profiles and coding that describe students' explanations on the case comparison problems in the pre-interviews, on an in-class activity, and in the post-interviews serve as the basis of the cross-case analysis for the results and discussion and are presented in the appendices. The presented analysis seeks to identify the similarities and differences for each student across time points and between students at each time point.

RQ1: What resources do students activate when considering the case comparison problems and how do the resources students activate change across time?

To address this research question, we will first discuss the resources students considered across time points, which are directly tied to the differences that students observed between the electrophiles, nucleophiles, and products in the presented reactions. Then we will discuss how students' activated resources changed across the time points of the study. This analysis is drawn from the inductive coding for resources

students used to guide their reasoning (with the detailed coding scheme presented in Appendix 2) and the individual themes, as they emerged from the data, are presented across Tables 1–3.

Resources activated when considering differences between the electrophiles. The two electrophiles for both the interview and in-class activity case comparison problems were carboxylic acid derivatives, with only one difference between them: the chlorine of the acid chloride *versus* the methoxy group substituent of the ester, as shown in Fig. 2. When considering this difference, all three students activated the resources of resonance, induction, and sterics at different points during the data collection. Only one student, Brooke, activated the resource of reactivity trends. When students activated each of these resources is summarized in Table 1.

Students activated the concept of resonance by recognizing the nonbonding electron pairs in the functional groups. For example, in Brooke's pre-interview they identified that both reactions have atoms with nonbonding electron pairs adjacent to the carbonyl that "have resonance with the oxygen in the carbonyl" which "should have some stabilizing effect." Similarly, in the post-interview, Chad identified that the methoxy group allows for resonance with the carbonyl, which "makes the negative charge more spread out and more present in the carbonyl" and "less electrophilic." That students appealed to resonance aligns with prior research demonstrating students' reliance on the concept (Ferguson and Bodner, 2008; Finkenstaedt-Quinn *et al.*, 2020; Petterson *et al.*, 2020). Notably, when the students considered resonance within the electrophiles, they did not exhibit the challenges students often have with resonance, such as describing resonance structures as distinct entities rather than contributors to a resonance hybrid (Taber, 2002; Kim *et al.*, 2019; Finkenstaedt-Quinn *et al.*, 2020; Petterson *et al.*, 2020; Xue and Stains, 2020).

Students similarly activated the concepts of induction or sterics when comparing the chlorine and methoxy functional groups. On the in-class activity, Violet considered sterics and induction in tandem, stating, "Reaction [C] is less sterically hindered as it doesn't have a bulky CH₃ [*sic*] group and also has an electron-withdrawing group (Cl) that makes the electrophilic carbon site more favourable for attack." The other students similarly used these concepts in their reasoning to suggest that

the electron-withdrawing chlorine would increase the reaction rate (or lower activation energy) while the steric bulk of the methoxy group would decrease the reaction rate (or raise activation energy). The students' consideration of both steric and inductive effects within the electrophiles exemplifies how they may reason using both explicit and implicit structural features without necessarily needing to infer implicit electronic properties from the explicit features, a possibility suggested in prior research (Caspari *et al.*, 2018a). However, students' use of steric considerations aligns with prior studies of students applying this concept when other, electronic properties are more appropriate (Bodé *et al.*, 2019).

Brooke activated the resource of reactivity trends, which involved discussing specific knowledge of which functional groups are generally more or less reactive. For instance, on the in-class activity, Brooke wrote that the acid chloride is "highly reactive" while the ester is "moderately reactive." This type of reasoning similarly appeared in Brooke's post-interview, in which they began by identifying that Reaction B has "the most reactive possible carboxylic acid derivative," a statement which they used to claim that the "energetics of this reaction are very favourable." This type of reasoning reflects the acyl compound reactivity trends students learn in the lecture course at the study institution, in which it is emphasized that acyl halides are the most reactive carboxylic acid derivatives and amides are the least reactive. Brooke's recollection of this trend could be reflective of students' rote memorization of rules in organic chemistry that has been documented in the literature (Kraft *et al.*, 2010; Christian and Talanquer, 2012; Grove and Bretz, 2012; Cruz-Ramírez De Arellano and Towns, 2014; Weinrich and Talanquer, 2016).

Resources activated when considering differences between the nucleophiles. The nucleophiles in the two case comparison problems were hydroxide and an amine. The amine was methylamine for the interview problem, whereas the amine was ammonia for the in-class activity, as illustrated in Fig. 3. When considering the different nucleophiles in the case comparisons, students reasoned with resources including charge, sterics, basicity, electronegativity, and reactivity trends. The resources that students activated at each time are summarized in Table 2.

The most frequently activated resources were the charges of the nucleophile. When considering charges, students focused on the negatively charged hydroxide compared to the neutrally charged amine. For example, when considering the hydroxide in the pre-interview, Violet stated, "You've got a negative charge already present, which I know already would want to not have that. So it's going to especially be attracted to getting rid of that negative charge." Similarly, on the in-class activity, Chad wrote that the hydroxide "is a better nucleophile because [of the] negative charge." In the post-interview, Brooke initially indicated similar reasoning, stating that their "immediate inclination is to think that the negatively charged species will be more reactive, but if I think about some trends here, I know that in general nitrogen species are going to be more nucleophilic than oxygen species." In Brooke's case, they first noted a resource they did not use in their reasoning (the negative formal charge)

Table 1 Resources students activated when considering the electrophile in the case-comparison problems. For each student, the checked (✓) boxes indicate the resources students activated at each time point (pre-interview, in-class activity, and post-interview). The crossed (✗) boxes indicate resources the students did not activate

Resource	Brooke			Violet			Chad		
	Pre	In-class	Post	Pre	In-class	Post	Pre	In-class	Post
Resonance structures	✓	✓	✗	✗	✗	✓	✓	✓	✓
Inductive effects	✗	✓	✗	✗	✓	✓	✗	✓	✓
Steric bulk	✓	✗	✓	✗	✓	✗	✓	✗	✗
Reactivity trends	✗	✓	✓	✗	✗	✗	✗	✗	✗

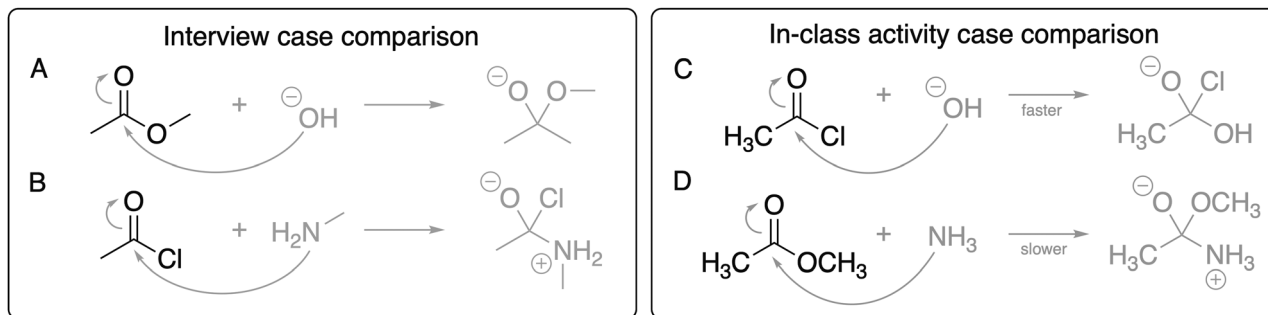


Fig. 2 The case-comparison reactions for the pre- and post-interviews (reactions A and B) and for the in-class activity (reactions C and D), with the electrophiles emphasized.

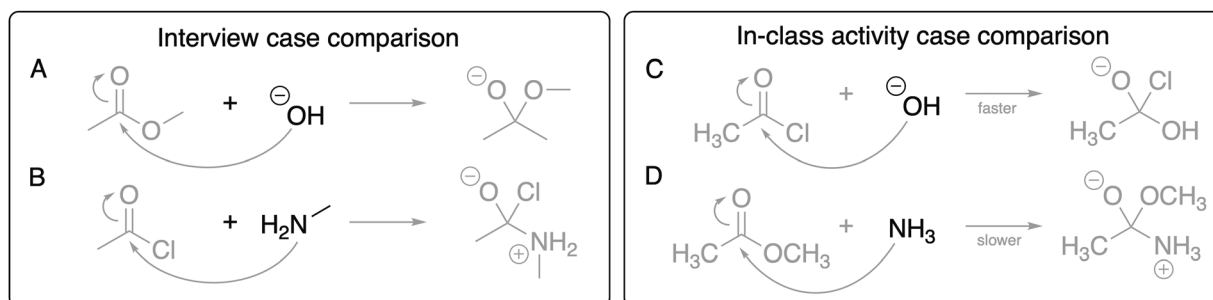


Fig. 3 The case-comparison reactions for the pre- and post-interviews (reactions A and B) and for the in-class activity (reactions C and D), with the nucleophiles emphasized.

Table 2 Resources students activated when considering the nucleophile in the case-comparison problems. For each student, the checked (\checkmark) boxes indicate the resources students activated at each time point (pre-interview, in-class activity, and post-interview). The crossed (\times) boxes indicate resources the students did not activate

Resource	Brooke			Violet			Chad			
	Pre	In-class	Post	Pre	In-class	Post	Pre	In-class	Post	
Formal charge	\times	\times	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\times	\checkmark	\checkmark
Steric bulk	\checkmark	\times	\checkmark	\times	\times	\times	\checkmark	\times	\times	\times
Basicity	\times	\times	\times	\times	\checkmark	\times	\checkmark	\times	\times	\times
Electronegativity	\times	\times	\times	\times	\times	\times	\checkmark	\times	\times	\times
Reactivity trends	\times	\times	\checkmark	\times	\times	\times	\times	\times	\times	\times

which activated a resource they did use in their reasoning (the reactivity trends). Brooke was the only student who mentioned this reactivity trend for the nucleophiles, similarly to how Brooke was the only student to mention the reactivity trends for the electrophiles. In general, the students' tendency to associate negative charges with nucleophilicity is similar to findings in the literature regarding how students conceptualize nucleophiles (Anzovino and Bretz, 2015, 2016; Dood *et al.*, 2020).

Students also reasoned by comparing the relative steric bulk of the nucleophiles, such as in Chad's pre-interview where they stated, "the nucleophile in Reaction [B] is more bulky and I think it's going to have a harder time trying to attack," suggesting that the increased steric bulk would correspond to higher activation energy for Reaction B. Brooke exhibited similar reasoning when considering sterics. Violet and Chad

also activated resources relating to basicity when considering the nucleophile—exemplified by Violet writing "[hydroxide] strong base" and "[ammonia] weaker base" on their response to the in-class activity. Chad reasoned by using the provided pK_a table to identify that hydroxide "has the higher pK_a of its conjugate base," connecting this to the fact that "[oxygen] is more electronegative" to identify that "[hydroxide] is going to be the better nucleophile" compared to methylamine. These students' consideration of sterics and basicity is similar to previous work that has demonstrated students' alignment of these concepts with nucleophilicity (Cruz-Ramírez De Arellano and Towns, 2014; Anzovino and Bretz, 2015).

Resources activated when considering differences between the product. The products for the case comparison reactions are emphasized in Fig. 4. All students activated one particular resource, the differences between formal charges, across the data collection time points. One student, Violet, also activated the concept of resonance when comparing the products in the pre-interview. The resources students activated for the products and for which time point are summarized in Table 3.

All three students considered the charges in the product during the pre-interview. When using charges to reason, students attempted to make a connection between charges and stability. In Chad's pre- and post-interview, they reasoned about charges in the product by stating that "it seems like Reaction [B] is going to have a more stable product just because the overall net charge is zero, and in Reaction [A] the products has an overall net charge of minus one." In contrast,

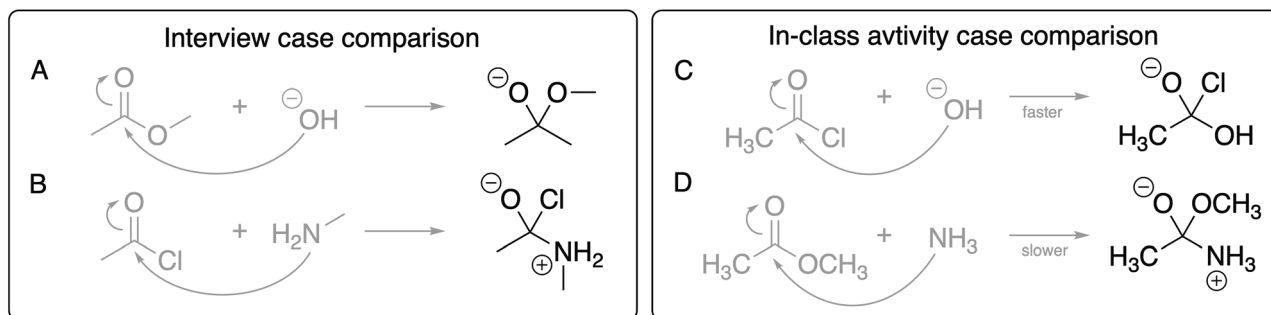


Fig. 4 The case-comparison reactions for the pre- and post-interviews (reactions A and B) and for the in-class activity (reactions C and D), with the products emphasized.

Table 3 Resources students activated when considering the product in the case-comparison problems. For each student, the checked (✓) boxes indicate the resources students activated at each time point (pre-interview, in-class activity, and post-interview). The crossed (✗) boxes indicate resources the students did not activate

Resource	Brooke			Violet			Chad		
	Pre	In-class	Post	Pre	In-class	Post	Pre	In-class	Post
Formal charge	✓	✗	✗	✓	✓	✓	✓	✗	✓
Resonance structures	✗	✗	✗	✓	✗	✗	✗	✗	✗

Brooke and Violet inferred the opposite relationship between charge and stability. For example, Violet stated in the pre-interview that it “is not favourable, generally, to have two charges within a molecule” and wrote on the in-class activity that the product in Reaction [D] is “less stable (has more charges).” Similarly, Brooke stated in the pre-interview that “the charges in the final product [of Reaction A] are more favourable than the charges in the final product of Reaction [B].” All of the students used their reasoning about charges to make claims about which of the two products was more stable. They all reasoned that the more stable product would correspond to the reaction with the lower activation energy. Such connections between formal charges, stability, and activation energy align with prior research demonstrating how students reason about charges focusing on the products when asked about the activation energy of a mechanistic step (Caspari *et al.*, 2018a).

Only one student, Violet, considered resonance in the products, when neither product structure is capable of resonance stabilization. When comparing the reactions, Violet initially indicated that the product in Reaction A, but not the product of Reaction B, was capable of resonance stabilization: “There’s also resonance stabilization in the molecule afterwards. I don’t think there could be for Reaction [B].…” However, Violet goes on to realize the incorrect application of this resource, stating, “Well actually wait... No, I don’t think you could do that... So maybe you wouldn’t have resonance stabilization, which would put a slight problem in my theory.” This was the only case of a student activating a resource and then recognizing that it would be conceptually incorrect to

apply the resource. As identified by Carle and Flynn’s (2020) research on learning objectives for resonance, the ability to recognize molecular structures able to engage in resonance stabilization is the first learning objective for the resonance concept, whereas using delocalization concepts to explain reactivity are among the final learning objectives. That Violet considered resonance but did not identify the correct molecules in the reactions in which the concept applies aligns with students’ application of the concept in prior studies (Petterson *et al.*, 2020). Furthermore, Violet’s reasoning suggests that their ability to use resonance as an explanatory concept is at an early stage in Carle and Flynn’s (2020) proposed set of resonance learning outcomes—though it is important to note that Violet recognized that the concept, as they initially applied it, was not correct.

Changes in activated resources across time points. For each student, the resources activated when examining the differences in the case comparisons changed from pre-interview to in-class activity to post-interview, as seen across Tables 1–3. The different resources students activated on the in-class activity may be tied to the slight differences in the framing of the problem—specifically, that the in-class activity problem identified the faster reaction and asked students to explain, whereas the interview problem asked the student to identify which reaction had the lower activation energy and explain their reasoning. Furthermore, there was an inherent difference in the setting between the in-class activity and the interviews. Hence, the comparisons made between the resources activated in the pre- and post-interviews are the most insightful, as the framing remained the same across these time points. However, due to the similarities between the reactions on the in-class activity and in the interviews, noting the resources students activated on the in-class activity is useful as a midpoint for gaining insight into how students’ reasoning changed across the semester.

In the pre-interview, Brooke considered resonance and sterics for the electrophile; sterics for the nucleophile; and charge for the product. After activating these resources, Brooke selected Reaction A by focusing on charges. In the post-interview, Brooke activated sterics and reactivity trends for the electrophile and charge, sterics, and reactivity trends for the nucleophile. In contrast to the pre-interview, Brooke selected

Reaction B by focusing on the reactivity trends in the electrophiles. A commonality in the resources Brooke activated from the pre- to post-interview was their discussion of sterics, which they described as having a possible influence on reactivity. However, in both interviews, they ultimately decided that sterics was not a factor that would change their response to the guiding question. Brooke was the only student to discuss reactivity trends, and most of the other resources Brooke considered for these time points (e.g., charges, resonance, and inductive effects) were related to how Brooke described these trends. The increased focus across time on the reactivity trends for the electrophile is evident from Brooke's response to the in-class activity, in which they only activated resources for the electrophile. Brooke's tendency in both the in-class activity and post-interview to activate resources focusing only on the electrophile or nucleophile aligns with prior research that identifies students' focus on reactants over products (Rodemer *et al.*, 2020).

In Violet's pre-interview, the only resources they activated were charges for the nucleophile and product, and—incorrectly—resonance for the product. That Violet did not activate any resources when considering the electrophile aligns with prior studies suggesting students do not necessarily focus on every feature when examining case comparison reactions (Bodé *et al.*, 2019; Rodemer *et al.*, 2020). Because Violet recognized that considering resonance in the product was incorrect, they selected Reaction A by only considering one resource: charges. For Violet's post-interview, they activated resonance structures and inductive effects in the electrophile and formal charges in the nucleophile and product. To respond to the guiding question, Violet used their consideration of resonance and induction for the electrophiles to select Reaction B. That Violet activated more resources from pre- to post-interview suggests a development in the ability to activate more resources in response to the problem. This possible development is also evident in their response to the in-class activity, where they activated more resources than they did for the pre-interview. Violet's reasoning about charges from pre- to post-interview was similar, but the fact that they considered more resources to ultimately select Reaction B in the post-interview demonstrates how the increase in activated resources shaped Violet's reasoning.

In Chad's pre-interview, they activated the resources of resonance and sterics for the electrophile; sterics, basicity, and electronegativity for the nucleophile; and charge for the product. They were unsure which resource to focus on to respond to the guiding question but selected Reaction A as their final response. In Chad's post-interview, they activated resonance and induction for the electrophile and charges for the nucleophile and product. They selected Reaction B by focusing on inductive effects in the electrophiles and the charges of the products. The resources Chad activated during the in-class activity were nearly the same as those activated during the post-interview, demonstrating how Chad narrowed the number of resources they activated from pre-interview to in-class activity to post-interview. This narrowing of activated resources was possibly valuable for Chad because they may

have been unable to respond to the guiding question in the pre-interview due to the number of resources they activated. Two of the resources Chad activated in the post-interview were the same as those activated in the pre-interview—resonance in the electrophile and charges. While Chad used these resources similarly for the two time points, their more focused consideration of inductive effects in the electrophile guided their reasoning in the post-interview.

RQ2: How do students weigh resources when constructing explanations for the case comparison problems across time?

We address this research question by focusing on how students weighed between multiple resources when constructing their explanations. As described above and in the coding scheme in Appendix 2, we specify weighing resources as the process of students identifying which of the multiple resources they deem relevant to be more or less important in their reasoning. We examine how students engaged in weighing resources differently between students and across time points, as captured by the second layer of coding completed in the analysis process. The results of this analysis are presented in Table 4. In characterizing each student's response holistically for how they weighed resources, we seek to illustrate how students organized the resources activated by the case comparison problems and the in-class activity when constructing their responses.

Students exhibited a range of abilities to weigh resources in the pre-interview. In the pre-interview, all participants selected Reaction A as having the lower activation energy. However, the way students weighed resources differed. Brooke weighed their considerations of resonance and charge, noting that resonance is “not the most important thing to consider” and focusing on the charges in the product to select Reaction A. Later in the interview, Brooke also exhibited placing less importance on their consideration of sterics, stating that the sterics are “going to have some impact, but I don't think it's going to be game-changing in this case.” Violet activated two resources in their pre-interview: charges in the nucleophile and product, and, incorrectly, resonance in the product. Because Violet recognized that considering resonance stabilization within the product was incorrect, as described previously, Violet selected

Table 4 Coding scheme for characterizing how each student weighed resources when constructing their explanations. For each student, the checked (✓) boxes indicate the type of weighing students engaged in at each time point (pre-interview, in-class activity, and post-interview). The crossed (✗) boxes indicate the type of weighing students did not engage in

Demonstrating weighing of resources	Brooke			Violet			Chad		
	Pre	In-class	Post	Pre	In-class	Post	Pre	In-class	Post
Uses one resource	✗	✗	✗	✓	✗	✗	✗	✗	✗
Does not weigh resources	✗	✗	✗	✗	✓	✗	✓	✓	✗
Weighs resources	✓	✓	✓	✗	✗	✓	✗	✗	✓

Reaction A based solely on their consideration of charges. Violet emphasized their focus on this resource, stating, “I always try to keep track of charges as best I can because that I find really helps me.” Because Violet only considered one resource in making their decision, they did not weigh resources. Both Violet and Brooke’s focus on charges in their explanations, despite their differences in considering and weighing other resources, aligns with prior research demonstrating students’ reliance on formal charges when considering mechanisms (Anzovino and Bretz, 2015; Galloway *et al.*, 2017; Caspari *et al.*, 2018a; Watts *et al.*, 2020).

Chad activated the most resources of the three students during the pre-interview. While Chad activated many relevant resources, the number of resources they considered proved to be challenging for them. They used these considerations to alternately support Reaction A (when comparing nucleophiles) and Reaction B (when comparing electrophiles and products). Chad changed their mind frequently throughout the interview, explicitly recognizing that they “keep going back and forth.” Near the end of the interview, Chad stated that “I think I’m just going to have to stick with the charges” to select Reaction B, before changing their mind a final time and selecting Reaction A due to the nucleophile and the resonance delocalization of the electrophile in Reaction A. Ultimately, Chad exhibited considering resources which support different conclusions and explicitly indicated difficulty in selecting which resources to weigh as most important in constructing their response.

While all students activated at least two resources in the pre-interview, they demonstrated differences in how they weighed resources. Brooke weighed between multiple resources and ultimately focused their reasoning on charges in the product as the most important resource. Violet similarly focused on charges in the product to make a decision, only after recognizing their incorrect consideration of resonance in the products. In contrast to Violet, Chad activated many resources and demonstrated difficulty with weighing resources. While Violet was able to make a decision and Chad was not, it was likely that Violet’s decision-making was possible because Violet only considered a single, relevant resource—a reasoning strategy that aligns with students’ one-reason decision-making, as identified in the literature (Maeyer and Talanquer, 2013). In contrast, Brooke and Chad considered multiple resources, with only Brooke exhibiting the ability to weigh between resources to make a decision. However, Chad’s engagement in a productive struggle—in contrast to Violet’s consideration of a single resource and apparent lack of difficulty in making a final decision—could have been useful for Chad’s learning.

Students all considered multiple resources but did not necessarily demonstrate evidence of weighing resources on the in-class activity. On the in-class activity, students were prompted to explain why one of the shown reactions was faster than the other after considering all relevant properties. The three participants exhibited similarities in terms of the number of resources activated. However, there was a difference in whether students weighed resources. For the in-class activity, students demonstrated weighing resources by activating resources

that would support different claims about which reaction was faster but ultimately building their explanation by focusing on the specific resources that support the claim for one of the reactions to be faster.

Brooke, who engaged in weighing resources during the pre-interview, also weighed resources on the in-class activity. This consistency in weighing was evident from how Brooke activated resources that supported different claims: the inductive effects of the chlorine and the resonance effects of the methoxy group, and the resource of general reactivity trends. Brooke indicated that the inductive effects of the chlorine would increase the rate of Reaction C but make the “double bond harder to break,” whereas the resonance effects of the methoxy group would discourage Reaction D despite making the “double bond easier to break.” By presenting reasoning with these resources that support either reaction being faster, Brooke demonstrated evidence of weighing the inductive effects of the chlorine over the resonance effects of the methoxy group.

While Brooke weighed resources on the in-class activity, Violet and Chad did not demonstrate evidence of weighing resources. They both only wrote about resources that supported the same conclusion and thereby did not necessarily weigh between resources that would have supported different conclusions. They each activated different resources, with some overlap. However, all of the resources that both Violet and Chad considered were used to support their explanation for why Reaction C was faster. As such, both students did not make visible any considerations that would provide a counterargument for Reaction D being faster, in the way that Brooke did. Hence, while Violet and Chad considered multiple resources, they did not produce evidence of weighing resources. However, it is possible that these students did activate resources that they weighed as less important by not including them in their written response. This could be an artefact of the activity itself—both the prompting on the activity and that it took place in-class rather than in an interview setting—as prior research suggests that prompt changes can influence students’ exhibited mechanistic reasoning (Cooper *et al.*, 2016; Crandell *et al.*, 2020).

Students exhibited convergence in ability to weigh resources in the post-interview. In the post-interview, all participants selected Reaction B as having the lower activation energy. Additionally, all students engaged in explicitly weighing resources. In Brooke’s reasoning, they identified how each activated resource would support different reactions having lower activation energy. Similarly to their response during the pre-interview, Brooke explicitly identified that they “don’t think that sterics are going to play a huge role.” Brooke placed the most weight on reactivity trends of the electrophiles and nucleophiles in selecting Reaction B. However, given Brooke’s abilities to weigh resources across the semester, it could be likely that Brooke’s appeal to reactivity trends represents the use of the resource as an explanatory concept—*i.e.*, a resource that is a collection of appropriate resources, such as resonance and inductive effects, that are responsible for determining reactivity trends.

Violet activated more resources on the post-interview as compared to their pre-interview. They explicitly identified that the electrophilicity of the acid chloride in Reaction B outweighs differences in charges on the nucleophiles: “Even though the nitrogen is not negatively charged like the oxygen, [the acid chloride] still makes [the nitrogen] better to possibly be attacking that [carbonyl carbon].” Violet also placed less emphasis on their consideration of charges in the products, selecting Reaction B despite stating that the products have “a negative charge and a positive charge in [the] product, which isn’t great.” Ultimately, Violet selected Reaction B after considering inductive and resonance effects in the electrophiles and charges in the nucleophiles and products, placing most emphasis on the differences between the electrophiles.

Chad activated a similar number of resources as the other participants during the post-interview. After considering each difference between the electrophiles to support Reaction B, Chad considered how the differences between the nucleophiles would support Reaction A. Then, when considering the products, Chad stated that “I still think actually Reaction [B] is going to be faster or have the lower activation energies, because I think the products also play a role.” In concluding the interview, Chad reiterated their choice of Reaction B by weighing the inductive effects in the electrophile and the overall neutral charge of the product over the different nucleophile strengths.

Across the post-interviews, students activated a similar number of resources and demonstrated similar abilities to weigh which resources are the most important. Furthermore, Brooke and Violet emphasized their considerations of the electrophiles over the resources activated when considering the nucleophiles or products, with Brooke focusing on reactivity trends and Violet on inductive and resonance effects. Chad similarly weighed inductive effects in the electrophile alongside charges in the product as the most important resources. Ultimately, while the students activated and focused on slightly different resources in the post-interview, they all successfully engaged in weighing resources to construct their explanations.

Limitations

The primary limitation of this research stems from the case study methodology. While the instrumental case study allows for a thorough investigation and description of how the students engaged with the organic case comparison problems across the three data collection time points, it inherently limits the scope of claims that can be made from the study. While the three case study participants selected for analysis were representative of the range in reasoning observed across the nine participants during the data collection process, the findings reported in this study are not meant to be generalizable to larger populations of students or students at different institutional settings with different backgrounds and experiences. Additionally, the case study participants were recruited on a voluntary basis, possibly contributing to self-selection bias.

Because of the methodological limitations, this research is also limited in the claims that can be made. While this study identified a broad range of resources activated across the case study participants, we cannot claim whether this demonstrated range captures all possible variability in students’ reasoning. In particular, the resources activated on the in-class activity may have been influenced by the inherent differences between the activity and the interviews. Thus, the discussion of students’ reasoning on the in-class activity is limited to identifying how their reasoning presented itself on the activity to provide context for the development in students’ reasoning observed from the pre- to post-interviews.

Conclusions

This case study provides an analysis of three second-semester organic chemistry students’ reasoning for acyl transfer case comparison problems across three time points: a pre-interview, an in-class activity, and a post-interview. The analysis of students’ reasoning focused specifically on the resources students activated when considering these problems and how students weighed the different resources. Our results demonstrate how case comparison problems can elicit multiple resources, both in the interview setting and on the in-class activity. Furthermore, our findings indicate a range of students’ abilities to engage in weighing resources.

Students activated a variety of resources for each stage of data collection, and the resources were not necessarily uniform across time points or between students. When students activated the same resources, they tended to reason similarly, both for each student and across time points. That is, when students considered the concepts of resonance, induction, and sterics, for example, they tended to use these concepts to support their explanations in similar ways. The findings regarding students’ use of resources contributes to the literature on students’ reasoning in organic chemistry by identifying the concepts students use when considering case comparisons of acyl transfer reactions. In particular, all resources students considered were directly related to the differences they identified between the case comparison reactions. That is, each resource tied directly to the explicit differences between the electrophiles, nucleophiles, and products in the reactions.

Students’ abilities to weigh resources ranged from basing their explanations on one activated resource to explicitly considering multiple resources and making their decisions based on what they deemed to be the most important. One student also exhibited the ability to identify an incorrect resource that they activated. When considering multiple resources, students did not necessarily weigh these resources in the pre-interview or during the in-class activity. This tendency was present when students indicated not knowing how to balance the resources they were considering and when students only considered resources that supported the same conclusion. Students also demonstrated explicitly weighing resources by stating the importance of certain resources over others when providing

their explanations. Most notably, students' activated resources and ability to weigh them differed between students and changed over time. While students began at different abilities, their ability to activate and weigh multiple resources converged over the semester.

Implications

Implications for research

Findings from this study indicate how students activate and weigh different resources when producing explanations for case comparison problems. However, future research is merited for furthering our understanding of how in-class activities can support students' reasoning. For instance, some of the differences observed in students' responses to the in-class activity and the interviews suggest the need for further research into how the framing of a prompt may influence the resources students activate and if they weigh resources. Furthermore, this research did not specifically examine the processes by which students deemed particular resources to be more productive than others, which is worth further study. The research design for this project—using pre- and post-interviews and artefacts of students' reasoning from an in-class activity—could also be extended to activities implemented with other instructors or at other institutions to increase the generalizability of the results presented herein. This exploratory case study focused on providing a detailed analysis of the resources students' activated and indicates a range in students' abilities to weigh resources across time points. Future research could seek to identify variations and nuances across students who demonstrate different reasoning abilities. In particular, the characterization of how students weighed resources in this study can be extended and applied to students' reasoning across the organic chemistry curriculum and into graduate programs. Future research could also develop and use quantitative measures of students' reasoning abilities to measure the effects of in-class activities on students' reasoning. Additionally, there is a need for further research exploring how instructors can use similar in-class activities to elicit and scaffold students' reasoning—*i.e.*, both the activation and weighing of multiple resources—with the use of case comparison problems in classrooms on a larger scale. For example, it would be valuable to research modifications of the instructional scaffold used in this study to explore how the prompting might better support students' engagement in weighing multiple resources.

Implications for practice

This research includes an in-class scaffold activity and demonstrates how it can elicit students' reasoning for case comparison problems within a large lecture. In particular, the case comparison problem in this study was found to be particularly useful for focusing students' reasoning on all of the differences between reactants and supporting their identification of numerous explicit and implicit structural properties to guide

their explanations. Furthermore, over the data collection period, our results illustrate that students were able to converge in their ability to weigh multiple resources, suggesting that students starting at different stages of reasoning ability can improve over a relatively short period of time.

Our findings indicate that, while students in general might consider a variety of concepts, specific students can consider different concepts for similar problems while arriving at the same answers. Furthermore, our results demonstrate that how students weigh concepts can differ. Together, these findings suggest incorporating instructional practices, such as the task design recently reported by Lieber and Graulich (2020), that both elicit and assess how students reason rather than focusing on the product of students' reasoning. For example, the students in this case study all provided the same final answer at each time point of data collection but differed in how they arrived at the answer. Hence, to support students' reasoning rather than ability to arrive at an answer, instructional practices must engage in eliciting, supporting, and assessing students' reasoning itself rather than the product of students' reasoning.

Other implications for practice relate to considerations specifically for teaching reaction mechanisms. We demonstrated how students might have incorrect or different understandings of fundamental topics in organic chemistry at the beginning of a second-semester, introductory organic chemistry course taken by both chemistry majors and non-majors. For example, one student within the case study indicated potential challenges with knowing what structures are capable of resonance stabilization; additionally, our participants exhibited different understandings of the relationship between charges in a molecule and relative stability. Hence, it is necessary for instructors in the middle of an organic chemistry course sequence to identify the core concepts from earlier semesters for which students may still need support in learning. Lastly, our results indicate a range in students' abilities to consider and weigh different resources across the semester. Instructors can use this finding to inform how they model reasoning strategies while connecting to students' existing problem-solving skills during instruction.

Conflicts of interest

There are no conflicts to declare.

Appendix 1. The in-class activity

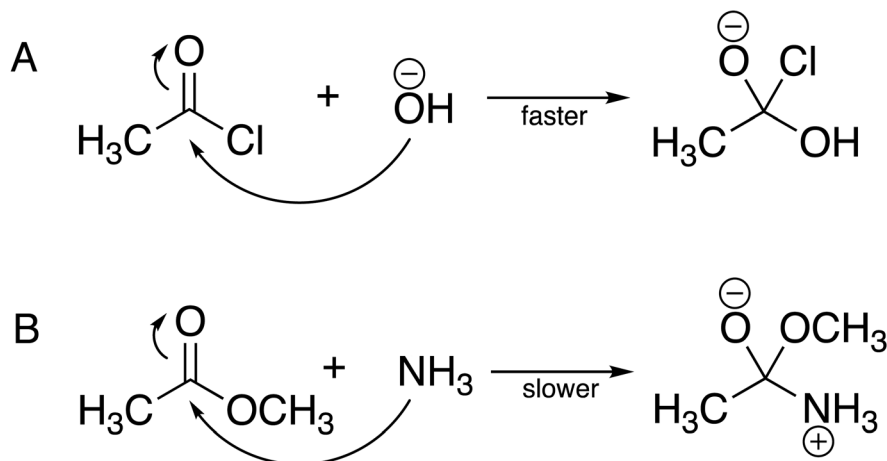
The three pages of the in-class activity worksheet are presented below in Fig. 5–7.

Appendix 2. The coding schemes

The two coding schemes are presented in Table 5.

Name:

Why do reactions A and B occur at different speeds? Make a prediction. Think about explaining your prediction by making a claim and backing that claim up with evidence and reasoning.



Address the points below. Use the table on page two to organize your thoughts and the space on page three to write your final answer. The final answer should be an in-depth explanation at least a paragraph.

1. What structures differ in both reactions A and B? Specify the functional groups in which the reactants differ. Note these differences in the first row of the table (boxes 1).
2. What chemical and physical properties do the functional groups in (boxes 1) have? Note the properties in the second row of the table (boxes 2).
3. What changes occur from reactants to products in both reactions A and B? Focus on the functional groups specified in (boxes 1). Note the changes, such as forming a charge, breaking a bond, or making a bond, in the first column of the table (boxes 3).
4. Why do the changes in (boxes 3) occur? Consider concepts from an energy viewpoint or drawing a reaction coordinate diagram for both reactions A and B.
5. Describe as precisely as possible how the properties in (boxes 1) and (boxes 2) influence the property changes in (boxes 3). Do the influences of the properties accelerate the reaction step or slow it down? Do they have no effect at all compared to the other reaction? Note the influences in the middle boxes (boxes 4/5).
6. Provide a statement that answers the question: **why do reactions A and B occur at different speeds?**

Fig. 5 The first page of the in-class activity.

Appendix 3. Profile descriptions for each case

Case 1. Brooke

Pre-interview. Brooke begins by noting similarities in the bonds being broken and formed in the presented mechanism and noting differences in functional groups between the reactions. In response to the guiding question, Brooke considers the resonance effects in the electrophile then the charges on the products. Brooke weighs these two considerations, noting that resonance is “not the most important thing to consider” and focusing on charges to select Reaction A because it does not form charges in the product. After selecting Reaction A, Brooke mentions considering the sterics of the nucleophiles and electrophiles but states that the sterics are “going to have some impact, but I don’t think it’s going to be game-changing in this case.” Ultimately, Brooke considers resonance in the electrophiles, the charges of the products, and sterics of all reactants but weighs charges most in selecting Reaction A.

In-class activity. Brooke identifies all the differences in the reactions and indicates properties relating to the electrophiles, writing about the electron withdrawing properties of the chlorine *versus* methoxy substituents and that the acid chloride is “highly reactive” while the ester is “moderately reactive.” Brooke identifies that the electron withdrawing chlorine increases the reaction rate but makes the carbonyl pi-bond harder to break. For the ester, Brooke identifies that resonance effects discourage the reaction at the carbonyl carbon but that resonance makes the carbonyl pi-bond easier to break. Brooke ultimately explains that Reaction C is faster, though they presented reasoning that would support selecting either reaction. Hence, they implicitly weighed the electron withdrawing property of chlorine to be more important than the resonance effects of the methoxy group.

Post-interview. Brooke begins by identifying that Reaction B has “the most reactive possible carboxylic acid derivative” to claim that the “energetics of this reaction are very favourable.”

Reminder: The numbers on the boxes correspond with the numbers from the instructions on the first page.

DIFFERENCES		
	(1) atoms and functional groups A	(1) atoms and functional groups B
	vs.	
	(2) physical and chemical properties of atoms and functional groups A	(2) physical and chemical properties of atoms and functional groups B
	vs.	
SIMILARITIES		
(3) change 1	(4/5) influence of properties A on change 1	(4/5) influence of properties B on change 1
(3) change 2	(4/5) influence of properties A on change 2	(4/5) influence of properties B on change 2

Fig. 6 The second page of the in-class activity.

(6) **Why do reactions A and B occur at different speeds?** If stuck, consider using this sentence stem:
 "Reaction _____ occurs at a faster speed because _____ affects _____ by _____."

Fig. 7 The third page of the in-class activity. The remainder of the page left space for students' responses.

Brooke then considers the possible steric influence of methylamine before selecting Reaction B. Next, Brooke discusses the nucleophiles, stating that "the negatively charged species will be more reactive" but "in general nitrogen species are going to be more nucleophilic," and later changing their mind to state that "OH might be a little bit more reactive, but I don't think it's going to be by a huge amount." Brooke also considers the steric influences of the electrophiles but states that they "don't think that sterics are going to play a huge role." To conclude, Brooke reiterates selecting Reaction B after weighing the reactivity trends of the electrophiles more heavily than the reactivity of the nucleophiles or possible steric influences.

Case 2. Violet

Pre-interview. Violet begins by noting the differences in functional groups and focusing on the charges present in the reactions, stating "I always try to keep track of charges as best I can because that I find really helps me." In response to the guiding question, Violet immediately chooses Reaction A and then justifies their choice with their consideration of charges on the nucleophiles: "You've got a negative charge already

present, which I know already would want to not have that. So it's going to especially be attracted to getting rid of that negative charge." They then consider the possibility for resonance stabilization within the product but realize their mistake in this reasoning because the products do not have delocalizable electron pairs, stating "so maybe you wouldn't have resonance stabilization, which would put a slight problem in my theory." After recognizing that it was incorrect to consider resonance, Violet concludes by reiterating their choice of Reaction A "because it's less charges in the first place." By only considering one resource, Violet does not engage in weighing between resources.

In-class activity. Violet identifies all differences between the reactions, indicating the steric and inductive effects influencing the electrophiles and the relative basicity of the nucleophiles. Violet indicates that the nucleophile in Reaction C will "attack faster" and that the product of Reaction C is "likely an intermediate and less reversible." They also indicate that the product in Reaction C will continue to react and form a more stabilized product, while the product in Reaction D is more likely to reverse to the starting materials due to the presence of

Table 5 The coding schemes for the analysis, including the resources students activated when considering the electrophiles, nucleophiles, and products, and the characterization of how each student weighed resources when constructing their explanations. Definitions and exemplars are provided for the three themes centred around resources students activated. Definitions are given for the characterization of how students weighed resources. As applying these codes required holistic evaluation of each data source, the profile descriptions corresponding to each code are indicated

Code	Definition	Profiles/exemplars
Coding scheme for resources activated when considering the electrophiles		
Resonance structures	The student considers the resonance structures or electron donating effects <i>via</i> resonance.	<i>Brooke, in-class activity</i> : “resonance effects make double bond easier to break”
Inductive effects	The student considers the inductive, electron-withdrawing effects of substituents.	<i>Violet, post-interview</i> : “Both of the groups on that first reagent are electron-withdrawing, which makes that carbon super partially positive.”
Steric bulk	The student considers the relative sterics of substituents.	<i>Chad, pre-interview</i> : “The methoxy group is going to have more sterics than the chlorine group.”
Reactivity trends	The student considers remembered reactivity trends for acyl compounds.	<i>Brooke, post-interview</i> : “I have the most reactive possible carboxylic acid derivative.”
Coding scheme for resources activated when considering the nucleophiles		
Formal charge	The student considers the charge of the nucleophiles.	<i>Chad, in-class activity</i> : “[hydroxide] is a better nucleophile because negative charge”
Steric bulk	The student considers the relative sterics of the nucleophiles.	<i>Brooke, pre-interview</i> : “There are some steric considerations here, as well. The hydroxyl group is not as bulky as the group in Reaction [B].”
Basicity	The student considers the basicity of the nucleophiles or the pK_a values of the nucleophiles' conjugate acids.	<i>Violet, in-class activity</i> : “[hydroxide] strong base” and “[ammonia] weaker base”
Electronegativity	The student considers the electronegativity of atoms in the nucleophiles.	<i>Chad, pre-interview</i> : “Also, [oxygen] is more electronegative so maybe that plays a role into it as well”
Reactivity trends	The student considers remembered reactivity trends for nucleophiles.	<i>Brooke, post-interview</i> : “If I think about some trends here, I know that in general nitrogen species are going to be more nucleophilic than oxygen species.”
Coding scheme for resources activated when considering the products		
Charge	The student considers the charges of the products.	<i>Chad, post-interview</i> : “Because the products, you have an overall neutral charge for reaction [B] and for reaction [A] you have a negative charge.”
Resonance structures	The student considers the possibility of resonance stabilization.	<i>Violet, pre-interview</i> : “There's also resonance stabilization in the molecule afterwards.”
Coding scheme for characterizing how each student weighed resources when constructing their explanations		
Uses one resource	The student makes their decision based on one resource and thereby does not engage in weighing resources.	<i>Violet, pre-interview</i>
Does not weigh resources	The student considers two or more resources that support alternative conclusions, and explicitly indicates not knowing which resource(s) have more (or less) influence on their decision-making OR the student considers two or more resources that support the same conclusion, thereby suggesting that no resource has more (or less) influence on their decision-making.	<i>Violet, in-class activity; Chad, pre-interview; Chad, in-class activity</i>
Weighs resources	The student considers two or more resources that support alternative conclusions, and suggests that at least one resource has more (or less) influence on their decision-making.	<i>Brooke, pre-interview; Brooke, in-class activity; Brooke, post-interview; Violet, post-interview; Chad, post-interview</i>

more charges. Violet also states that Reaction C is faster “because of sterics and electron density,” stating that the less sterically hindered and more electron withdrawing chlorine (*versus* the methyl group) increases the reaction rate. They also indicate the negative charge on the hydroxide to support Reaction C being faster. Each line of reasoning that Violet considers is in support of Reaction C. That is, Violet does not write about considerations that would support Reaction D and thus does not engage in weighing the importance of different resources.

Post-interview. Violet answers the guiding question by identifying differences in the reactions and considering the resonance effects possible in the electrophile for Reaction A that make the carbonyl carbon “less partially positive” and “less likely to be attacked.” They also consider the inductive effects of the chlorine that increase the electrophilicity of the carbonyl carbon in Reaction

B. They voice that the electrophilicity of the acid chloride in Reaction B outweighs differences in charges on the nucleophiles: “Even though the nitrogen is not negatively charged like the oxygen, [the acid chloride] still makes [the nitrogen] better to possibly be attacking that [carbonyl carbon].” Violet next considers the charges of the products, stating that the Reaction B products have “a negative charge and a positive charge in [the] product, which isn't great.” To conclude, Violet selects Reaction B after considering inductive and resonance effects in the electrophiles and charges in the nucleophiles and products, placing most emphasis on the differences between the electrophiles.

Case 3. Chad

Pre-interview. Chad identifies the differences between the electrophiles and nucleophiles in the reactions and, in response

to the guiding question, select Reaction A. They support their choice using the pK_a table and electronegativity to suggest that hydroxide is a better nucleophile. Chad then considers the resonance and sterics of the electrophile in Reaction A and the steric differences between the nucleophiles, considerations which they recognize support Reaction B. Then Chad states being conflicted about their response, keeping Reaction A as their answer, before identifying the charges in the products. Chad changes their mind to Reaction B because it “is going to have a more stable product just because the overall net charge is zero.” They continue recognizing the conflicts in their thinking, specifically stating that “the nucleophile in Reaction [B] is more bulky and I think it’s going to have a harder time trying to attack.” Chad recognizes that they “keep going back and forth, but I think I’m just going to have to stick with the charges” and their selection of Reaction B, before again considering the resonance delocalization of the electrophile and better nucleophile in Reaction A. In the end, Chad selects Reaction A after considering a number of resources and struggling to weigh which resources are the most important.

In-class activity. Chad identifies all differences between the reactions and considers a number of properties: the inductive effects of the electronegative chlorine in Reaction C’s electrophile, the resonance effects in Reaction D’s electrophile, and the charges and relative strength of the nucleophiles. Chad writes that the charge on the hydroxide in Reaction C “makes it more nucleophilic and willing to react,” while the ammonia is “a more neutral charge, that will be more stable and less likely to react as fast.” Chad also indicates the influences of the differences between the electrophiles: that the carbon in Reaction C is “more electrophilic” while the carbonyl in Reaction D is resonance stabilized which makes the carbon “less electrophilic.” Each influence Chad considers supports Reaction C, and thus Chad does not engage in weighing different resources.

Post-interview. Chad answers the guiding question by selecting Reaction B and supporting their claim by identifying that the electron withdrawing group on the electrophile makes the carbonyl more electrophilic. They identify the resonance effects in Reaction A’s electrophile, which “makes the negative charge more spread out and more present in the carbonyl” and “less electrophilic.” Chad then considers the different nucleophiles, stating that Reaction A has a stronger nucleophile because it has a negative charge, but despite this they suggest they are still in favour of Reaction B. Chad next considers the overall charges in the products, claiming that the overall neutral charge in the products of Reaction B further support their choice. To conclude, Chad reiterates their choice of Reaction B, weighing the inductive effects in the electrophile and the overall neutral charge of the product over the different nucleophile strengths.

Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1256260. The authors would additionally

like to thank Solaire Finkenstaedt-Quinn, Nicholas Garza, Jessica Scott, Chang-Hwa Chiang, Maiya Yu, and members of the Shultz group for assistance with facilitating the in-class activity and discussions related to the preparation of this manuscript.

References

- Alfieri L., Nokes-Malach T. J. and Schunn C. D., (2013), Learning Through Case Comparisons: A Meta-Analytic Review, *Educ. Psychol.*, **48**(2), 87–113.
- Anderson T. L. and Bodner G. M., (2008), What can we do about “Parker”? A case study of a good student who didn’t “get” organic chemistry, *Chem. Educ. Res. Pract.*, **9**(2), 93–101.
- Anzovino M. E. and Bretz S. L., (2015), Organic chemistry students’ ideas about nucleophiles and electrophiles: The role of charges and mechanisms, *Chem. Educ. Res. Pract.*, **16**(4), 797–810.
- Anzovino M. E. and Bretz S. L., (2016), Organic chemistry students’ fragmented ideas about the structure and function of nucleophiles and electrophiles: A concept map analysis, *Chem. Educ. Res. Pract.*, **17**(4), 1019–1029.
- Bhattacharyya G. and Bodner G. M., (2005), “It gets me to the product”: How students propose organic mechanisms, *J. Chem. Educ.*, **82**(9), 1402–1407.
- Bhattacharyya G. and Harris M. S., (2018), Compromised Structures: Verbal Descriptions of Mechanism Diagrams, *J. Chem. Educ.*, **95**(3), 366–375.
- Bodé N. E., Deng J. M. and Flynn A. B., (2019), Getting Past the Rules and to the WHY: Causal Mechanistic Arguments When Judging the Plausibility of Organic Reaction Mechanisms, *J. Chem. Educ.*, **96**(6), 1068–1082.
- Carle M. S. and Flynn A. B., (2020), Essential learning outcomes for delocalization (resonance) concepts: How are they taught, practiced, and assessed in organic chemistry? *Chem. Educ. Res. Pract.*, **21**(2), 622–637.
- Caspari I. and Graulich N., (2019), Scaffolding the structure of organic chemistry students’ multivariate comparative mechanistic reasoning, *Int. J. Phys. Chem. Educ.*, **11**(2), 31–43.
- Caspari I., Kranz D. and Graulich N., (2018a), Resolving the complexity of organic chemistry students’ reasoning through the lens of a mechanistic framework, *Chem. Educ. Res. Pract.*, **19**(4), 1117–1141.
- Caspari I., Weinrich M. L., Seviran H. and Graulich N., (2018b), This mechanistic step is “productive”: organic chemistry students’ backward-oriented reasoning, *Chem. Educ. Res. Pract.*, **19**(1), 42–59.
- Christian K. and Talanquer V., (2012), Modes of reasoning in self-initiated study groups in chemistry, *Chem. Educ. Res. Pract.*, **13**(3), 286–295.
- Cooper M. M., Kouyoumdjian H. and Underwood S. M., (2016), Investigating Students’ Reasoning about Acid-Base Reactions, *J. Chem. Educ.*, **93**(10), 1703–1712.
- Crandell O. M., Lockhart M. A. and Cooper M. M., (2020), Arrows on the Page Are Not a Good Gauge: Evidence for

- the Importance of Causal Mechanistic Explanations about Nucleophilic Substitution in Organic Chemistry, *J. Chem. Educ.*, **97**(2), 313–327.
- Cruz-Ramírez De Arellano D. and Towns M. H., (2014), Students' understanding of alkyl halide reactions in undergraduate organic chemistry, *Chem. Educ. Res. Pract.*, **15**(4), 501–515.
- DeCocq V. and Bhattacharyya G., (2019), TMI (Too much information)! Effects of given information on organic chemistry students' approaches to solving mechanism tasks, *Chem. Educ. Res. Pract.*, **20**(1), 213–228.
- diSessa A. A., (1993), Toward an Epistemology of Physics, *Cogn. Instr.*, **10**(2–3), 105–225.
- diSessa A. A. and Sherin B. L., (1998), What changes in conceptual change? *Int. J. Sci. Educ.*, **20**(10), 1155–1191.
- Domin D. S., Al-Masum M. and Mensah J., (2008), Students' categorizations of organic compounds, *Chem. Educ. Res. Pract.*, **9**(2), 114–121.
- Dood A. J., Fields K. B. and Raker J. R., (2018), Using Lexical Analysis to Predict Lewis Acid-Base Model Use in Responses to an Acid-Base Proton-Transfer Reaction, *J. Chem. Educ.*, **95**(8), 1267–1275.
- Dood A. J., Dood J. C., Cruz-Ramírez De Arellano D., Fields K. B. and Raker J. R., (2020), Analyzing explanations of substitution reactions using lexical analysis and logistic regression techniques, *Chem. Educ. Res. Pract.*, **21**(1), 267–286.
- Ferguson R. and Bodner G. M., (2008), Making sense of the arrow-pushing formalism among chemistry majors enrolled in organic chemistry, *Chem. Educ. Res. Pract.*, **9**(2), 102–113.
- Finkenstaedt-Quinn S. A., Watts F. M., Petterson M. N., Archer S. R., Snyder-White E. P. and Shultz G. V., (2020), Exploring student thinking about addition reactions, *J. Chem. Educ.*, **97**(7), 1852–1862.
- Galloway K. R., Stoyanovich C. and Flynn A. B., (2017), Students' interpretations of mechanistic language in organic chemistry before learning reactions, *Chem. Educ. Res. Pract.*, **18**(2), 353–374.
- Galloway K. R., Leung M. W. and Flynn A. B., (2019), Patterns of reactions: A card sort task to investigate students' organization of organic chemistry reactions, *Chem. Educ. Res. Pract.*, **20**(1), 30–52.
- Goodwin W., (2012), *Mechanisms and chemical reaction*, Elsevier B.V.
- Grandy G., (2012), Instrumental Case Study, in Mills A. J., Durepos G. and Wiebe E. (ed.), *Encyclopedia of Case Study Research*, SAGE Publications, Inc., pp. 474–476.
- Graulich N., (2015), The tip of the iceberg in organic chemistry classes: How do students deal with the invisible? *Chem. Educ. Res. Pract.*, **16**(1), 9–21.
- Graulich N. and Bhattacharyya G., (2017), Investigating students' similarity judgments in organic chemistry, *Chem. Educ. Res. Pract.*, **18**(4), 774–784.
- Graulich N. and Caspari I., (2020), Designing a scaffold for mechanistic reasoning in organic chemistry, *Chem. Teach. Int.*, DOI: 10.1515/cti-2020-0001.
- Graulich N. and Schween M., (2018), Concept-Oriented Task Design: Making Purposeful Case Comparisons in Organic Chemistry, *J. Chem. Educ.*, **95**(3), 376–383.
- Grove N. P. and Bretz S. L., (2012), A continuum of learning: From rote memorization to meaningful learning in organic chemistry, *Chem. Educ. Res. Pract.*, **13**(3), 201–208.
- Grove N. P., Cooper M. M. and Cox E. L., (2012a), Does mechanistic thinking improve student success in organic chemistry? *J. Chem. Educ.*, **89**(7), 850–853.
- Grove N. P., Cooper M. M. and Rush K. M., (2012b), Decorating with arrows: Toward the development of representational competence in organic chemistry, *J. Chem. Educ.*, **89**(7), 844–849.
- Hammer D. and Elby A., (2000), Epistemological resources, *Fourth Int. Conf. Learn. Sci.*, **8406**(February), 4–5.
- Hammer D., Elby A., Scherr R. E., Redish E. F., Hammer D., Elby A., *et al.*, (2004), Resources, framing, and transfer Resources, framing, and transfer, (Rec 0087519), 1–26.
- Herrington D. G. and Daubenmire P. L., (2014), Using interviews in CER projects: Options, considerations, and limitations, *ACS Symp. Ser.*, **1166**, 31–59.
- Kim T., Wright L. K. and Miller K., (2019), An examination of students' perceptions of the Kekulé resonance representation using a perceptual learning theory lens, *Chem. Educ. Res. Pract.*, **20**(4), 659–666.
- Kraft A., Strickland A. M. and Bhattacharyya G., (2010), Reasonable reasoning: Multi-variate problem-solving in organic chemistry, *Chem. Educ. Res. Pract.*, **11**(4), 281–292.
- Lieber L. and Graulich N., (2020), Thinking in Alternatives—A Task Design for Challenging Students' Problem-Solving Approaches in Organic Chemistry, *J. Chem. Educ.*, **97**(10), 3731–3738.
- Maeyer J. and Talanquer V., (2013), Making predictions about chemical reactivity: Assumptions and heuristics, *J. Res. Sci. Teach.*, **50**(6), 748–767.
- McClary L. and Talanquer V., (2011), College chemistry students' mental models of acids and acid strength, *J. Res. Sci. Teach.*, **48**(4), 396–413.
- Miles M. B., Huberman A. M. and Saldana J., (2014), *Qualitative data analysis: A methods sourcebook*, 3rd edn, Los Angeles, CA: Sage.
- Petterson M. N., Watts F. M., Snyder-White E. P., Archer S. R., Shultz G. V. and Finkenstaedt-Quinn S. A., (2020), Eliciting student thinking about acid–base reactions via app and paper–pencil based problem solving, *Chem. Educ. Res. Pract.*, **21**(1), 878–892.
- Rodemer M., Eckhard J., Graulich N. and Bernholt S., (2020), *Decoding Case Comparisons in Organic Chemistry: Eye-Tracking Students' Visual Behavior*.
- Schmidt-McCormack J. A., Judge J. A., Spahr K., Yang E., Pugh R., Karlin A. *et al.*, (2019), Analysis of the role of a writing-to-learn assignment in student understanding of organic acid–base concepts, *Chem. Educ. Res. Pract.*, **20**(2), 383–398.
- Taber K. S., (2002), Compounding Quanta: Probing the Frontiers of Student Understanding of Molecular Orbitals, *Chem. Educ. Res. Pract.*, **3**(2), 159–173.
- Thagard P., (1989), Explanatory coherence, *Behav. Brain Sci.*, **12**, 435–502.
- Watts F., Schmidt-McCormack J., Wilhelm C., Karlin A., Sattar A., Thompson B. *et al.*, (2020), 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.*, **21**(3), 1148–1172.
- Webber D. M. and Flynn A. B., (2018), How Are Students Solving Familiar and Unfamiliar Organic Chemistry Mechanism Questions in a New Curriculum? *J. Chem. Educ.*, **95**(9), 1451–1467.
- Weinrich M. L. and Talanquer V., (2016), Mapping students' modes of reasoning when thinking about chemical reactions used to make a desired product, *Chem. Educ. Res. Pract.*, **17**(2), 394–406.
- Xue D. and Stains M., (2020), Exploring Students' Understanding of Resonance and Its Relationship to Instruction, *J. Chem. Educ.*, **97**(4), 894–902.
- Yin R., (2014), *Case Study Research: Design and Methods*, London: SAGE Publications.