Exploring undergraduate organic chemistry students' strategies and reasoning when solving organic synthesis problems

by

Nicholas Bodé

A thesis submitted in partial fulfillment of the requirements for the
Master's degree in Chemistry

Department of Chemistry and Biomolecular Science
Faculty of Science
University of Ottawa

Abstract

Organic synthesis problems are a common assessment tool in organic chemistry courses, as they give instructors the opportunity to determine students' ability to integrate and apply their knowledge of reactions and skills learned in the course. However, students often tend to be unsuccessful in solving them, even if they appear to have a strong grasp on other course material. We hypothesized that part of the reasoning behind this issue is because it can be challenging to integrate learning activities into the curriculum that give students the opportunity to apply their knowledge to synthetic problem solving, while still giving students the opportunity to master the underlying concepts (knowledge of organic reactions and reaction mechanisms). In addition, there is a gap in our understanding of the mental models students construct while solving these problems, as there is no evidence that they approach these problems in the same manner that experts do (*i.e.*, retrosynthetic analysis). The research described in this thesis was performed to address these issues in two ways. First, we designed learning activities for students that were meant to help them develop more systematic approaches (whose benefits are supported by evidence) to solving synthesis problems, and determining if those learning activities could produce significant learning gains. The learning activities we designed were made available to students through out-of-class learning workshops, where learning gains were primarily measured through the analysis of students' synthetic problem-solving abilities, assessed immediately before and after the workshops. Second, we sought to obtain a better understanding of students' mental models when solving synthesis problems; specifically, we wanted to see if they had well-defined strategies for approaching these problems, and if they had a canonical understanding of how these strategies were meant to be applied. To do so, we invited students to participate in semi-structured think-aloud interviews, where participants were asked to solve synthesis problems. We investigated both of these topics using a constructivist paradigm for learning, which states that knowledge is constructed in the mind of the learner rather than passively imparted. The process of knowledge construction is heavily influenced by the prior knowledge and experiences of the learner, and meaningful understanding of new knowledge is unlikely to occur if new knowledge cannot be accommodated by existing knowledge structures. Results from these studies indicated that the workshop-style intervention did not have any effect on students' ability to successfully solve synthesis problems, but we did observe proficiency in the ability to use expert-like strategies, suggesting that more practice over time could lead to the ability to solve synthesis problems more effectively. Our analysis of the interview data showed that some students can proficiently use strategies in situations that are familiar to them, but do not appear to be able to apply those strategies to predict outcomes in unfamiliar situations; further, we observed a strong reliance on the use of reasoning that was based on memorized rules. Future work could further explore the mental models that students construct for solving synthesis problems; we recommend the incorporation of specific instruction on the use of synthesis problem-solving strategies, and research could explore the relationship between students' abilities, and how synthesis is taught, practiced, and assessed in the organic chemistry curriculum.

Keywords: Chemistry Education Research, Organic Chemistry, Synthesis, Problem Solving

Acknowledgements

I would first like to thank my supervisor, Dr. Alison Flynn, for giving me the opportunity to work in her research group over the past years. Alison, I have benefitted significantly from your knowledge and encouragement over the course of these projects. Apart from endless helpful discussions about my research, you've always taken a genuine interest in my life outside of the office, and I can't thank you enough for the dedicated role you've had as my supervisor.

I would also like to thank my committee members, Dr. David Trumpower and Dr. Chris Boddy. Chris, I would also like to thank you in particular for your role as my Organic Chemistry 2 professor. This course was where I learned that organic chemistry would always be a part of my career path; therefore, I believe you laid the foundation for what I've been able to accomplish in the area thus far, and whatever else I am able to accomplish in the future.

I had the opportunity to work with three fantastic post-doctoral fellows over the past few years: Dr. Kelli Galloway, Dr. Amanda Bongers, and Dr. Mark Raycroft. Kelli, your patience with my research methods questions and your appreciation of my constant need to reference *The Office* made my adjustment to graduate studies much easier than it would have been otherwise; your feedback on my work was always incredibly helpful, and I'm grateful for it. Amanda, I appreciate the insight you were always able to offer about my research ideas, your expertise in synthetic organic chemistry, which made me a better student, and finally, for your encouragement whenever I doubted myself. Lastly, Mark: thank you for all of the above, and for reminding me "*you are ~~Lisa Simpson~~ Nicholas Bodé*"

I also want to thank all of the other members of the Flynn group that I had the opportunity to work with, especially Joëlle, Delphine, Natalie, Rebecca, Keith, and Myriam: thank you each of you for your helpful input on my research, and for being a friend during our mutual time with the group.

Last, I wish to thank the students who volunteered their time to participate in the studies discussed in this thesis, particularly the eight who participated in both the synthesis workshop and interview portions of my research: I cannot thank you by name, but this research would not have been possible without you.

Dedicated to

J. N.: I'm not going to read yours, either. But hey, we did it.

K. P.: Thanks for your belief that academia was for me. Hopefully I can continue to prove you right.

Mom and Dad: Thanks for everything; particularly your patience.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

^{1}H NMR	Proton Nuclear Magnetic Resonance
ANOVA	Analysis of Variance
Ar	Aryl
Ar–R	A single bond between an aryl ring and an alkyl group
ASCIv2	Revised Attitudes Toward the Subject of Chemistry Inventory
C–C	A single bond between two carbon atoms
C–O	A single bond between a carbon atom and an oxygen atom
C=C	A double bond between two carbon atoms
CER	Chemistry Education Research
χ^2	Chi-squared
CFA	Confirmatory Factor Analysis
CFI	Confirmatory Fit Index
COPUS	Classroom Observation Protocol for Undergraduate STEM
DBER	Discipline-Based Education Research
E1	Unimolecular elimination
E1cb	Unimolecular elimination conjugate base
E2	Bimolecular elimination
EPF	Electron Pushing Formalism
Jones reagent	Sodium dichromate and sulfuric acid in a mixture of acetone/water
MCAI	Metacognitive Activities Inventory
mCPBA	*m*-Chloroperoxybenzoic Acid
OC1	Organic Chemistry 1
OC2	Organic Chemistry 2
PCC	Pyridinium Chlorochromate
POGIL	Process-Oriented Guided Inquiry Learning
R	A general alkyl group
RMSEA	Root Mean Square Error of Approximation
RQ	Research Question
s.d.	Standard Deviation
S_N1	Unimolecular nucleophilic substitution
S_N2	Bimolecular nucleophilic substitution
STEM	Science, Technology, Engineering, and Mathematics
Ts	*p*-toluenesulfonyl (Tosyl)

Chapter 1 Introduction

The domain of interest of this thesis is in Chemistry Education Research (CER). CER is a specific type of discipline-based education research (DBER), which encompasses a group of research fields that are primarily interested in defining expertise in a given discipline (relevant concepts, practices, and ways of thinking), and understanding how that expertise is developed, and identifying/developing learning objectives that help students develop that expertise.[1] Therefore, the goal of chemistry education researchers is to determine what constitutes expertise in the domain of chemistry, and to obtain evidence for how that expertise could be best developed.[2]

This thesis is primarily interested in how students develop expertise in organic chemistry, where one of the most advanced learning outcomes is the ability to apply one's understanding of mechanistic patterns and knowledge of reactions to solving multi-step synthesis problems. While these problems are appropriately bounded, so that they may be solved using reactions that are familiar to the student, the student must still recall all of the reactions and skills from their previous courses. This project is motivated by our prior research on how students approach synthesis problems, which has shown that students who use certain well-defined strategies in association with one another tend to be more successful in solving these problems than students who do not have a well-defined strategy.[3] However, a limitation of this work was that it did not allow us to understand how students are thinking about these problems. Therefore, the goal of this research project was to obtain a further understanding of what strategies and reasoning processes undergraduate students use to solve organic synthesis problems.

To this end, we explore the extent to which expert-like reasoning is present in students' approaches to solving organic synthesis problems. Furthermore, we seek to determine if learning activities centered around the use of specific problem solving strategies we identified in previous research,[3] which we hypothesize should foster higher-level modes of reasoning, are more effective than synthesis learning activities that are centered around reaction types, which we hypothesize should foster simpler modes of reasoning. Finally, we seek to understand how students think about these problems as they work through them, and to what extent their thought process reflects the intended learning outcomes related to organic synthesis in a patterns-of-mechanisms curriculum.

This thesis describes three studies, each of which has its own dedicated chapter. While each study is related, different theoretical frameworks and methods were used for each. Therefore, the relevant background to the research we conducted is provided at the start of each chapter.

Chapter 2 of this thesis will outline the study that investigated the use of workshops to help students learn organic nomenclature using Nomenclature101.com, an online learning tool. This study has been published in *Chemistry Education Research and Practice*, and appears in this thesis unaltered from how it is presented in the journal. This study investigated the following research questions:

RQ1 What learning gains have participants who have used nomenclature101.com (independently and when guided) made compared to participants who are taught nomenclature through a traditional classroom tutorial?

RQ2 What are participants' perceptions of the usefulness, ease of use, and overall learning experience in each learning setting?

This study was important as a proof-of-principle for the notion that one intervention for a single topic outside of class could lead to significant learning gains for participants, and lead to the decision to use the same format for the study outlined in Chapter 3 of this thesis.

Chapter 3 of this thesis will address our second study, which explores RQs 1 and 2. These research questions are based on a major implication of our prior findings: courses that focus on developing skillfulness in synthetic problem solving need to provide students with the opportunity to practice *explicitly* using the strategies we identified in our research to solve authentic problems.[4] This idea is supported by other work that has suggested that the use of abstract visual representations is important in supporting the problem-solving process.[5-8] It follows that teaching and learning activities should include formative assessment[9] of the ability to *properly apply these strategies* as much as the ability to actually propose a synthesis of a given target molecule.

Therefore, the goal of the research described in Chapter 3 of this thesis was to determine if learning activities centered around the use of specific problem solving strategies we identified in previous research,[3] which we hypothesize should foster higher-level modes of reasoning, are more effective than synthesis learning activities that are centered around reaction types, which we hypothesize should foster simpler modes of reasoning. We investigated this hypothesis through the development of three synthesis learning workshops.

Meanwhile, Chapter 4 will address our final study, which explores RQ 3. This research question is based on a major limitation of our prior work investigating students' problem-solving strategies in responses to synthesis problems on exams: we were able to characterize the strategies they chose to write down, but we were unable to elucidate their procedural characteristics (*i.e.*, step-by-step process) and the reasoning behind students' problem solving strategies; we were only able to

analyze the final products of their analysis. We reasoned that it was likely that these details were important to make truly meaningful implications for how learning outcomes related to synthetic problem solving should be enacted, taught, practiced and assessed in the context of an organic chemistry curriculum that is organized by patterns of mechanisms rather than functional group. The goal of this revised curriculum is to "improve students' abilities using the electron pushing formalism, including predicting and explaining unknown reactions." [10]

These two studies, outlined in Chapters 3 and 4, investigated three main research questions, listed below.

RQ1 What differences in the use of key strategies/degree of success might be observed in students' responses to synthetic problems following workshops focusing on different strategic approaches to solving synthetic problems?

RQ2 What differences in metacognitive skillfulness and attitudes toward chemistry might be observed between participants who provide successful and unsuccessful solutions to synthetic problems?

RQ3 How do students use the key strategies (taught in class and/or in the workshops) to evaluate and/or solve synthetic problems?

Finally, Chapter 5 will provide a summary and synthesis of the conclusions from each of these three studies.

References

(1) National Research Council. *Discipline-Based Education Research*; National Academies Press: Washington, D.C., 2012.
(2) Cooper, M. M.; Stowe, R. L. Chemistry Education Research-From Personal Empiricism to Evidence, Theory, and Informed Practice. *Chem. Rev.* **2018**, acs.chemrev.8b00020.
(3) Bodé, N. E.; Flynn, A. B. Strategies of Successful Synthesis Solutions: Mapping, Mechanisms, and More. *J. Chem. Educ.* **2016**, *93* (4), 593–604.
(4) Bhattacharyya, G.; Bodner, G. M. Culturing Reality: How Organic Chemistry Graduate Students Develop Into Practitioners. *J. Res. Sci. Teach.* **2014**, *51* (6), 694–713.
(5) Bowen, C. W.; Bodner, G. M. Problem-Solving Processes Used by Students in Organic Synthesis. *Int. J. Sci. Educ.* **1991**, *13* (2), 143–158.
(6) Moreno, R.; Ozogul, G.; Reisslein, M. Teaching with Concrete and Abstract Visual Representations: Effects on Students' Problem Solving, Problem Representations, and Learning Perceptions. *J. Educ. Psychol.* **2011**, *103* (1), 32–47.
(7) Domin, D.; Bodner, G. Using Students' Representations Constructed During Problem Solving to Infer Conceptual Understanding. *J. Chem. Educ.* **2012**, *89* (7), 837–843.

(8) Sevian, H.; Bernholt, S.; Szteinberg, G. A.; Auguste, S.; Pérez, L. C. Use of Representation Mapping to Capture Abstraction in Problem Solving in Different Courses in Chemistry. *Chem. Educ. Res. Pract.* **2015**, *16* (3), 429–446.
(9) Nicol, D. J.; Macfarlane Dick, D. Formative Assessment and Self-Regulated Learning: a Model and Seven Principles of Good Feedback Practice. *Stud. High. Educ.* **2006**, *31* (2), 199–218.
(10) Flynn, A. B.; Ogilvie, W. W. Mechanisms Before Reactions: a Mechanistic Approach to the Organic Chemistry Curriculum Based on Patterns of Electron Flow. *J. Chem. Educ.* **2015**, *92* (5), 803–810.

Chapter 2 Evaluating students' learning gains and experiences from using Nomenclature101.com

N. E. Bodé, J. Caron and A. B. Flynn, *Chem. Educ. Res. Pract*., **2016**, *17*, 1156.
DOI: 10.1039/C6RP00132G - Reproduced by permission of The Royal Society of Chemistry.

Chemistry Education Research and Practice

PAPER

View Article Online
View Journal | View Issue

Cite this: *Chem. Educ. Res. Pract.*, 2016, **17**, 1156

Evaluating students' learning gains and experiences from using nomenclature101.com

Nicholas E. Bodé,[a] Jeanette Caron[b] and Alison B. Flynn*[a]

Received 7th June 2016,
Accepted 2nd September 2016

DOI: 10.1039/c6rp00132g

www.rsc.org/cerp

Skill in organic chemistry nomenclature is fundamental for communicating more complex concepts. Interpreting and using functional group names is particularly important. With nomenclature101.com, students can create tailored interactive quizzes according to their learning needs; the tool is free and available in English and French. The present study investigated students' nomenclature learning in three different learning environments to determine (1) what learning gains participants make when using nomenclature101.com independently, when guided, or taught nomenclature through a traditional classroom tutorial and (2) students' perceptions of the usefulness, ease of use, and overall learning experience in each of the settings. We invited students from first and second semester organic chemistry courses at a large, research-intensive Canadian university to participate in a nomenclature workshop. When participants arrived, they were randomly sorted into one of three treatment groups: classroom tutorial, independent use of nomenclature101, or guided use of nomenclature101. Before the session, participants completed a pre-test; at the end, they completed a post-test and a questionnaire related to affective aspects of their experience. We analyzed participants' scores and questionnaire responses, and qualitatively analyzed their test answers, including errors. Learning gains were significant with large effect sizes for all three groups although there were no significant differences in learning gains between groups. The largest gains were observed in the ability to correctly identify and draw functional groups. Exploratory factor analysis showed that the post-workshop questionnaire about nomenclature101.com reliably measured participants' perceptions about two latent factors: usefulness and ease of use. Based on questionnaire results, most participants liked the learning tool and found it useful and easy to use. Participants in all groups reported enjoying their learning experience. Few participants postdicted their quiz grades accurately, suggesting that metacognitive skillfulness was lacking among workshop participants. The large learning gains observed after using nomenclature101.com or learning in a classroom setting for just one hour shows the potential that instruction has to help students learn functional group identification skills, which ideally will mitigate barriers to communication and understanding. These results offer flexibility to educators as they make instructional choices such as teaching nomenclature in a course period or tutorial setting or asking students to learn nomenclature independently with nomenclature101.com. Students have the flexibility to work outside of class in the manner of their choosing.

Introduction

Nomenclature skill provides a foundation for learning advanced chemistry. Students must first be able to identify functional groups and name and draw molecules before they can solve advanced problems. While students do not necessarily need to know the correct placement of every comma and dash in a molecule's name, barriers to communication can arise in discussions about reactivity when participants cannot at least identify the functional groups in question. Students in a first year organic chemistry course struggled with these skills, even in the final exam; the class average on a question asking students to identify functional groups (an aldehyde and an alcohol) was only 58%, instead of the expected 98% (Flynn *et al.*, 2013). The results were even lower when students were asked to name a specific molecule. Despite a large number of resources available describing nomenclature, few interactive resources exist that allow students to actively practice and get feedback; even fewer resources exist in French, one of the languages of instruction at the authors' institution. Many resources are instructor controlled (*e.g.*, homework assignments) and charge a subscription fee.

Nomenclature101.com is an online learning tool that was designed and created to address these issues (Fig. 1)

[a] *Department of Chemistry and Biomolecular Sciences, University of Ottawa, 10 Marie Curie Private, Ottawa, ON, K1N 6N5, Canada. E-mail: alison.flynn@uottawa.ca*
[b] *Centre for e-Learning, Teaching and Learning Support Service, University of Ottawa, 136 Jean-Jacques Lussier, Ottawa, ON, K1N 6N5, Canada*

Fig. 1 Nomenclature101.com's home page, which supports student learning in English or French (Flynn *et al.*, 2014).

(Flynn *et al.*, 2014). The tool is designed to help students achieve three learning outcomes associated with nomenclature:

(1) Identify a functional group, given the structure

(2) Name a molecule according to IUPAC rules, given the structure

(3) Draw a molecule, given its IUPAC name

Using nomenclature101.com, students can create tailored, interactive quizzes according to their learning choices; the tool is free and available in English and French. Students receive feedback (right/wrong) when they submit their answer, can consult hints, and see facts about molecules.

The tool has been accessed more than 10 000 times from almost every country in the world. However, students' learning gains and experiences using the learning tool have not been investigated. Studies of interactive learning tools suggest that level of guidance can have a strong effect on students' use of the tool or activity (Bonawitz *et al.*, 2011; Akaygun and Jones, 2014; Chamberlain *et al.*, 2014) and that the level of guidance can benefit students at different times (González-Cruz *et al.*, 2003). The present study measured the learning gains in organic chemistry nomenclature of students using nomenclature101.com compared to those learning in a technology-free classroom setting and assessed students' satisfaction with the resources offered by the learning tool.

Guskey evaluation framework

nomenclature101.com was designed with a constructivist approach modeled on the strategies of cognitive apprenticeship (Oriol *et al.*, 2010) and aligned with many aspects of meaningful learning (Bretz, 2001; Novak, 2011). This study was designed using part of a Guskey evaluation (Guskey, 2002, 2010), a five-level model that examines participants' (or students'): reactions (level 1), learning (level 2), organizational support and change (level 3), use of new knowledge and skills (level 4), and learning outcomes (level 5).

To evaluate the benefits to student learning of nomenclature101.com compared to a traditional nomenclature tutorial, we asked the following research questions:

(1) What learning gains have participants who have used nomenclature101.com (independently and when guided) made compared to participants who are taught nomenclature through a traditional classroom tutorial? (Guskey level 5)

(2) What are participants' perceptions of the usefulness, ease of use, and overall learning experience in each learning setting? (Guskey level 1)

Methods

Setting and course

Participants in the study were taking Organic Chemistry I and II courses at a large, research-intensive Canadian university. These courses have been described in previous work (Bodé and Flynn, 2016). Organic Chemistry I is offered in the winter semester of students' first year of studies, and Organic Chemistry II is offered in the summer and fall semesters. Both of these courses may be taken in either English or French, and consist of two weekly lectures (1.5 hours each, mandatory, lecture or flipped format), and an optional tutorial session (1.5 hours, also called a recitation or discussion group). The Organic I course has a required, associated laboratory section (3 hours biweekly) and the Organic II course has a laboratory course that runs concurrently and is only required for some programs (3 hours weekly). Assessment for the course generally consists of two midterm exams, a final exam, online homework assignments, and class participation using a classroom response system ("Top Hat," 2016). Assessment for the laboratory portion of Organic Chemistry I consists of pre-lab quizzes, completed online before each experiment is performed, six lab reports (one for each experiment), and in-lab performance. These courses are composed of ~75% Faculty of Science students, ~17% Faculty of Health Sciences students, and ~8% students from other faculties. A new curriculum has been used in the courses since 2012 (Flynn and Ogilvie, 2015). The topics taught in Organic Chemistry I and II include: structure, properties, stereochemistry, and conformational analysis of organic compounds, IR and ^{1}H NMR spectroscopy, electron-pushing formalism/symbolism, acid–base chemistry, π bond electrophiles (*e.g.*, 1,2-carbonyl addition reactions, acetals and imine formation, addition–elimination reactions, *etc.*), π bond nucleophiles, aromaticity and electrophilic aromatic substitution, S_N1, S_N2, E1, E2, and α-carbons as nucleophiles (*e.g.*, aldol condensation, alkylation).

Participants

The participant pool was composed of 65% Faculty of Science students (N = 45), ~20% Faculty of Health Science students (N = 13), and ~12% Faculty of Engineering students (N = 8); the remainder did not report their program of study (N = 3). 52% of participants (N = 36) were first year students, 33% (N = 23) were second year students, 6% (N = 4) were third year students, and 4% (N = 3) were fourth year students; the remaining 4% (N = 3) did not report their year of study. Students from organic chemistry courses were verbally invited to participate in a nomenclature workshop, which was held outside of regular course hours. The authors' institutional review board (IRB) reviewed and approved all stages of the research project. Informed consent was obtained for all stages involving human subjects.

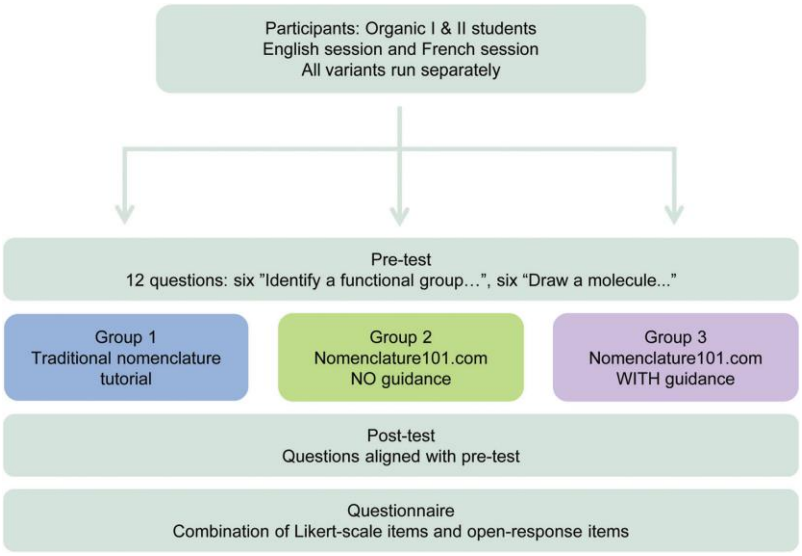


Fig. 2 Methods for the study. The first workshop (February 2015) was offered in both English and French; the second workshop (May 2015) was only offered in English.

The workshop was piloted in the fall of 2014, then conducted at two different times: February 2015 in English and French and May 2015 in English. In February 2015, participants were recruited from Organic Chemistry I courses; in May 2015, participants were recruited from Organic Chemistry II. 59% of participants ($N = 41$) completed the workshop in February 2015, while the remaining 41% of participants ($N = 28$) completed the workshop in May 2015. 85% of participants completed the workshop in English ($N = 59$), and 15% completed the workshop in French ($N = 10$).

Session formats

When participants arrived for the workshop, they were randomly assigned to one of three, one-hour tutorial sessions on organic chemistry nomenclature (Fig. 2). Participants who did not bring a laptop or mobile device (as requested) used the computer in the room or worked with another participant.

All the sessions were guided by videos to ensure consistency between sections (French/English and February/May). All videos welcomed the participants, described the purpose of the study (but did not tell participants which group they were in), and asked participants to complete the pre-test while the video paused. After the pre-test, the videos described the intended learning outcomes (ILOs) for the session (same for each group) with a screenshot of nomenclature101.com (without the site title or URL present) (Fig. 3); the ILOs were also handed out to the participants. Then Group 1 received a traditional classroom tutorial on organic nomenclature, Group 2 used nomenclature101.com independently, and Group 3 received a guided tutorial on nomenclature101.com (Fig. 2) (note: we capitalized "Group" when naming one of these three groups in our results and discussion for clarity). At the end of the session, the video was restarted, participants were instructed to complete the post-test and questionnaire, and they were thanked for attending. These portions of the videos were identical for all three groups. The workshops took approximately two hours to complete, with 45–60 minutes for the tutorial session itself.

The classroom sessions in Group 1 were facilitated by PhD candidates in organic chemistry with multiple years of experience as teaching assistants for organic chemistry courses. After the video introduction, these facilitators presented a one-hour lecture on nomenclature punctuated with exercises for the participants to complete; they answered participants' questions as they arose. During the exercises, participants could work individually or in groups.

After the video introduction, Group 2 (unguided use of nomenclature101.com) participants were prompted to use nomenclature101.com for one hour. They could work individually or in groups. Facilitators for Group 2 could answer technical questions about the site but not chemistry questions.

After the video introduction, Group 3 (guided use of nomenclature101.com) participants were prompted to explore nomenclature101.com for 10 minutes in any way they wanted. Next, Group 3 received additional video guidance that Groups 1 and 2 did not receive, which explained the various features of the learning tool, including the resources available (*e.g.*, find hints, see nomenclature rules, see facts about the molecules, use the quiz dashboard, see a summary of results). Participants could work individually or in groups. The next 5 minutes of video gave recommendations on how to work through the site: first, to

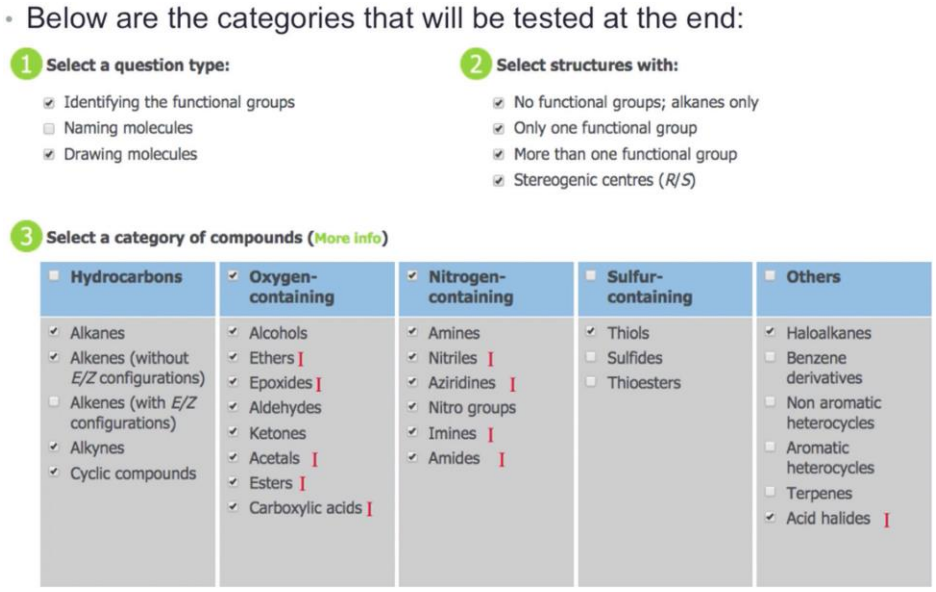


Fig. 3 Intended learning outcomes that were given to all workshop participants *via* video and handout.

practice identifying functional groups then to drawing molecules (going from the simplest to the most complex and those with stereocentres). Participants were also advised to work on aspects of the site that were aligned with the session's ILOs. Participants then had 45 minutes to use the site, for a total of one hour learning time. They could work individually or in groups. The facilitators for Group 3 were chemistry professors or senior teaching assistants who could answer both technical and chemistry questions.

At the beginning of the session, all groups completed the same pre-test on organic chemistry nomenclature (Appendix 1). The pre-test had twelve questions: six questions asked participants to identify functional groups (given a complex structure) and six questions asked participants to draw molecules (given the name). Immediately following the workshops, participants completed a post-test on organic chemistry nomenclature (Appendix 2). This test was in the same format as the first, with different but aligned questions; all groups completed the same post-test. The pre-test and post-test contained the same questions in the English/French and February/May sessions. These questions were aligned with the expected nomenclature learning outcomes from the Organic I course.

Lastly, participants completed a questionnaire (Appendices 3 and 4). The questionnaire was adapted from an existing one (Bolliger and Supanakorn, 2011) and was designed to access students' perception of the tutorial based on four dimensions—usefulness, access, interface and learner performance. Each questionnaire contained sixteen questions answered using a Likert scale from 1 (strongly disagree)—5 (strongly agree). Twelve items were positively worded and four were negatively worded. The items were worded in the same way on each questionnaire, but modified to address the setting; for example, tutorial setting (Group 1) Item 1 statement, "The tutorial was useful in learning nomenclature", was reworded: "The quizzes available in www.nomenclature101.com were useful in learning nomenclature" on the nomenclature101.com workshop questionnaire (Groups 2 and 3). The questionnaire also asked participants to report demographic information and their expected letter grade in the workshop.

Data analysis

We analyzed participants' scores, types and frequencies of errors, and questionnaire responses, making comparisons between the pre- and post-tests, between groups, and within groups. We performed all statistical analyses using IBM SPSS Statistics (version 23). We combined the results from groups that completed the workshop in February and in May because there were no significant differences between their test scores on either the pre-test, $U = 502.0$, $z = -0.391$, $p = 0.696$, or the post-test, $U = 508.5$, $z = -0.306$, $p = 0.760$. The students who participated in the workshop in May had just started Organic Chemistry II, and had already completed Organic Chemistry I, while those who participated in February had not completed a full semester of organic. As such, we might expect that test scores would be higher in the May workshop, as these participants would likely have slightly more prior chemistry knowledge. A possible reason that there was no significant difference in scores is that students who take Organic Chemistry II in the summer may be weaker students who were repeating the course, because students following the Faculty's suggested course sequence normally take this course in the Fall semester. Another possibility is that there was little explicit nomenclature instruction in Organic Chemistry I (nomenclature learning outcomes were communicated to students in the syllabus and they were expected to work independently), and participants in May had not yet been tested on nomenclature in their Organic Chemistry II course.

We also combined the results from groups that completed the workshop in French with those that completed the workshop in English, because there were no significant differences between their test scores on either the pre-test, $U = 257.5$, $z = -0.404$, $p = 0.686$, or the post-test, $U = 256.5$, $z = -0.422$, $p = 0.673$.

To determine the effect of the nomenclature tutorial and of using nomenclature101.com (independently and when guided) on participants' test scores, we performed two-way repeated measures ANOVA. The dependent variable in each ANOVA was participants' scores on the tests that were administered. The within-subjects variable was time (pre-test and post-test) and the between-subjects factor was the workshop treatment group. The main effect of time and the interaction effects of time and the workshop treatment group were evaluated using the criterion of Wilks' lambda (Λ). We performed a parametric ANOVA instead of a non-parametric test, despite having non-normal data. The parametric test was appropriate because of the larger overall sample size ($N = 66$) and an observed power of 1, which indicated a low probability of committing type II errors.

For follow-up analysis of the time effect and due to the smaller sample sizes for the individual workshop groups, we performed Kolmogorov–Smirnov tests of normality on the data to determine the suitability of using *t*-tests. The distribution of scores was significantly non-normal ($p < 0.05$) for all groups on the pre-test, and for Groups 1 and 2 on the post-test. Therefore, we conducted Mann–Whitney *U* tests, the non-parametric equivalent of the *t*-test, to perform *post hoc* analysis of the significant main effect of time in a more statistically powerful fashion. We report effect size measures for Mann–Whitney *U* tests using Pearson's *r*.

To analyze participants' errors on the "circle the functional group" questions, we identified the functional group that each participant circled. To analyze the "draw the molecule" questions, one author coded participants' responses to each question type for certain types of errors using an open-coding system (Patton, 2014); the codes are summarized in Table 1. A second rater then independently coded a subset of the data using the same coding template to determine the degree of reliability in our interpretation of participants' errors in the "draw the molecule" questions. We measured the reliability of the coding system using the criterion of Krippendorff's α because this statistic adjusts for chance agreement, while a simple percent agreement calculation does not (Harshman and Yezierski, 2015). The results of this analysis show that our coding system was reliably applied throughout the data (6 codes, $N_{\text{items}} = 403$, $\alpha = 0.791$).

Table 1 Coding system used to analyze "draw the molecule" questions. Reliability analysis was performed using the criterion of Krippendorff's α

Code	Description
Missing/extra carbon	Number of carbon atoms in the molecule does not match the name (either in the backbone or branches)
Incorrect constitutional isomer	Functional group(s) are not bound to the carbon atom indicated in the name of the molecule
Incorrect or missing stereochemistry[a]	Configuration at stereocentre is not indicated using correct symbolic conventions or incorrect configuration is indicated
Incorrect or missing functional group	Functional group drawn does not match the name, or a functional group indicated in the name is missing from the molecule drawn
Non-meaningful use of symbolism	Parts of a molecule are drawn (*e.g.*, functional groups, Lewis structures) that do not follow the conventions for accurately representing organic molecules

[a] Only relevant to Question 11.

Krippendorff suggests that a value above 0.800 is ideal, but any value above 0.667 is still acceptable to draw any conclusions from a content analysis (Krippendorff, 2004). Once our reliability analysis was complete, the data were recoded until both raters reached complete agreement upon all codes.

To assess the significance of differences in error frequencies between groups and between pre-workshop and post-workshop tests, we performed chi-square (χ^2) tests of independence. We report effect size measures for chi-square tests using the criteria of Cramér's V (ϕ_c).

Using participants' responses to the post-workshop nomenclature101.com questionnaire, we performed principal axis factoring with oblique rotation (Promax with Kaiser normalization). For the questionnaire responses, we reverse-coded negatively worded items (item no. 5, 9, 12, 14—Table 3) to facilitate analysis. We performed reliability analysis to verify the internal consistency of both the nomenclature101.com and nomenclature tutorial questionnaires using Cronbach's α. We also used the questionnaires to assess participants' postdiction accuracy by asking them what they thought their letter grade on the post-test would be. Participants' responses to this item were compared to their actual scores to gain insight about participants' metacognitive skillfulness.

Limitations

Social constructivist theories of learning propose that limitations on an individual's level of potential development are determined through collaborative rather than individual problem solving (Vygotsky, 1978); as such, a factor that may have influenced learning gains was whether or not participants chose to spend their time using nomenclature101.com alone or in groups. We hypothesize that participants who chose to work in groups would have benefitted more from the workshop than those who chose to work alone, but social effects on learning were not measured as part of the study.

We also did not measure the extent to which learning gains from the workshop were retained on a long-term basis, but expect that regular, long-term use of nomenclature101.com would be required to ensure that any learning gains would be retained (Kornell, 2009). Future studies that measure learning gains following an intervention could consider the retention aspect of learning, but controlling for variations in what participants do between a post-test and a delayed post-test (for instance, how much/how often they choose to review the material) may be a challenge.

Results and discussion

All groups showed large learning gains

Our first goal was to determine the learning gains that participants experienced using this tool compared to a traditional tutorial setting. The majority of participants in all groups performed better on the post-workshop test than on the pre-workshop test (Fig. 4 and 5).

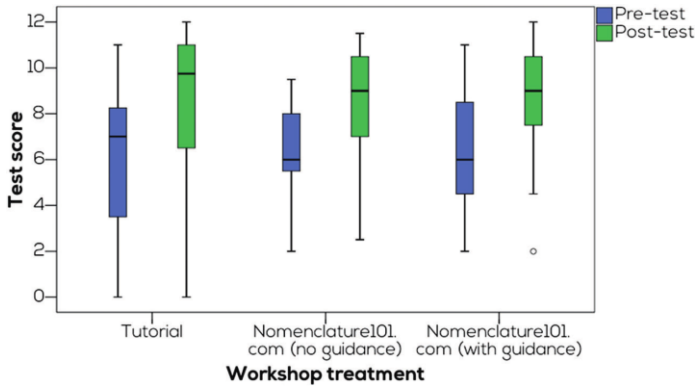


Fig. 4 Box plots showing the distribution of pre-test and post-test scores for each treatment group. $N_{tutorial}$ = 20, $N_{nomenclature101.com\,(no\,guidance)}$ = 21, $N_{nomenclature101.com\,(with\,guidance)}$ = 25.

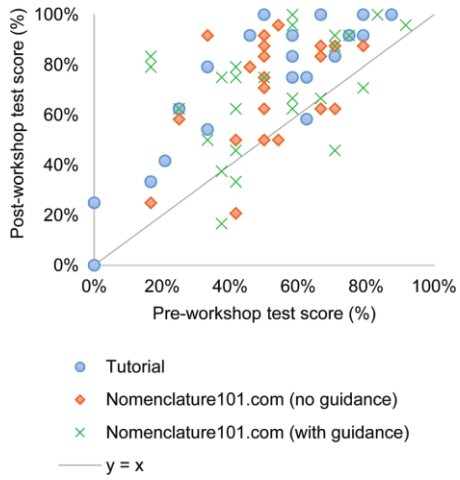


Fig. 5 Most participants' post-test scores were higher than their pre-test scores. $N_{tutorial}$ = 20, $N_{nomenclature101.com\,(no\,guidance)}$ = 21, $N_{nomenclature101.com\,(with\,guidance)}$ = 25.

Learning gains were significant for all groups and there was no significant difference in learning gains between groups (Table 2). The results of a two-way repeated measures ANOVA show the main effect of time was significant, Λ = 0.498, $F(1,63)$ = 63.480, $p < 0.001$, and had a large effect size, partial η^2 = 0.50. The interaction effect was non-significant, Λ = 0.990, $F(2,63)$ = 0.327, p = 0.722, with a small effect size, partial η^2 = 0.01. The test of between-subjects effects shows that the effect of the workshop treatment was non-significant, $F(2)$ = 0.01, p = 0.990, partial $\eta^2 < 0.001$. *Post hoc* analysis showed the learning gains in each individual workshop were significant and had medium effect sizes (Table 2).

Functional group identification: errors and learning gains. All groups had significant gains in the ability to identify functional groups and there were no significant differences between groups. Two-way repeated measures ANOVA showed the time main effect was significant, Λ = 0.637, $F(1,63)$ = 35.953, $p < 0.001$, with a large effect size, partial η^2 = 0.36. The interaction effect was non-significant, Λ = 0.949, $F(2,63)$ = 1.683, p = 0.19, with a small effect size, partial η^2 = 0.05. The test of between-subjects effects shows that the effect of the workshop treatment was non-significant, $F(2)$ = 0.025, p = 0.976, partial η^2 = 0.001.

The most pronounced gains related to functional group identification were in the ability to identify an acetal; participants commonly identified an ester on the pre-test when asked to identify an acetal, but many corrected this error on the post-test (Fig. 6). The increase in the ability to identify this functional group was significant with a large effect size for Group 1, $\chi^2(1)$ = 18.027, $p < 0.001$, ϕ_c = 0.67, significant with a medium effect size for Group 2, $\chi^2(1)$ = 9.882, p = 0.002, ϕ_c = 0.49, and significant with a large effect size for Group 3, $\chi^2(1)$ = 14.286, $p < 0.001$, ϕ_c = 0.54.

Close to significant gains were observed for Group 1 in the ability to distinguish between a nitro group and a nitrile, $\chi^2(1)$ = 3.636, p = 0.057, with a medium effect size, ϕ_c = 0.30. Non-significant gains were observed for Group 2 participants, $\chi^2(1)$ = 1.556, p = 0.212, with a small effect size, ϕ_c = 0.19, and for Group 3 participants, $\chi^2(1)$ = 1.389, p = 0.239, with a small effect size, ϕ_c = 0.17 (Fig. 7). Although learning gains were not significant for any Group, more than 65% of participants in each group correctly identified a nitrile on the post-test.

The difference between an ester and an acetal, which have a similar arrangement of atoms, and between a nitro group and a nitrile, which have similar sounding names, might be considered trivial to the novice, but represent important implications in structure and reactivity. The fact that these learning gains can be observed after just one hour using nomenclature101.com shows the potential this learning tool has to mitigate the small errors that could lead to much larger errors in understanding reactivity later on in the course, without an impact on classroom time. Alternatively, one hour class time dedicated to nomenclature could rapidly lead to learning gains.

Drawing molecules given their name: errors and learning gains. All groups made significant gains in the ability to draw molecules although there was no significant difference between groups in gains in the ability to draw molecules. Two-way repeated measures ANOVA showed the main effect of time was significant for this section of the test, Λ = 0.717, $F(1, 63)$ = 24.815, $p < 0.001$, with a large effect size, partial η^2 = 0.28. The interaction effect was non-significant, Λ = 0.970, $F(2,63)$ = 0.986, p = 0.379, with a small effect size, partial η^2 = 0.03. The test of between-subjects effects shows that the effect of the workshop treatment was non-significant, $F(2)$ = 0.013, p = 0.987, partial $\eta^2 < 0.001$.

Fig. 8 shows the proportions of participants who made each type of error on at least one of the six questions, and shows that each of those frequencies decreased for nomenclature101.com users between the pre-test and post-test. The most pronounced decrease among nomenclature101.com users (Groups 2 and 3) was in incorrect or missing functional groups. This improvement was significant for Group 2, $\chi^2(1)$ = 6.857, p = 0.009, with a medium effect size, ϕ_c = 0.40, and significant for Group 3, $\chi^2(1)$ = 5.903, p = 0.015, with a medium effect size, ϕ_c = 0.36. Meanwhile, the decrease was non-significant in tutorial

Table 2 Mann–Whitney *U* test results, comparing nomenclature pre-test and post-test total scores

Group	$N_{pre\text{-}test}$ [a]	U	z	p	r
Tutorial (Group 1)	20	100.0	−2.712	0.007	−0.429[b]
Unguided use of nomenclature101.com (Group 2)	21	101.5	−3.017	0.003	−0.466[b]
Guided use of nomenclature101.com (Group 3)	25	167.0	−2.829	0.006	−0.400[b]

[a] $N_{pre\text{-}test} = N_{post\text{-}test}$. [b] Medium effect size.

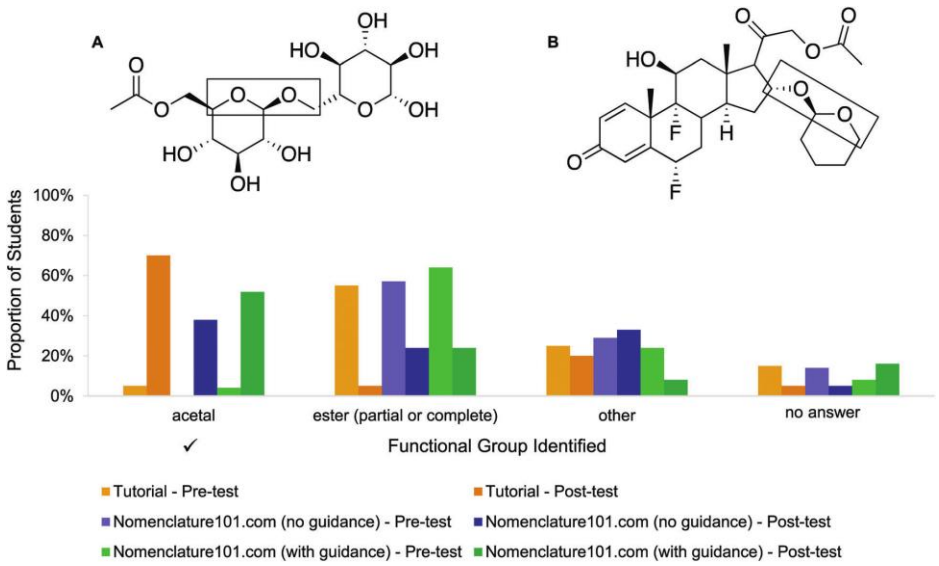


Fig. 6 Types of functional groups circled by workshop participants on (A) pre-test question 3 and (B) post-test question 3. The correct answer to each question is indicated with a box on the molecule and a checkmark on the chart. N_{tutorial} = 20, $N_{\text{nomenclature101.com (no guidance)}}$ = 21, $N_{\text{nomenclature101.com (with guidance)}}$ = 25. Participants made significant gains in functional group identification between the pre- and post-tests in every workshop group.

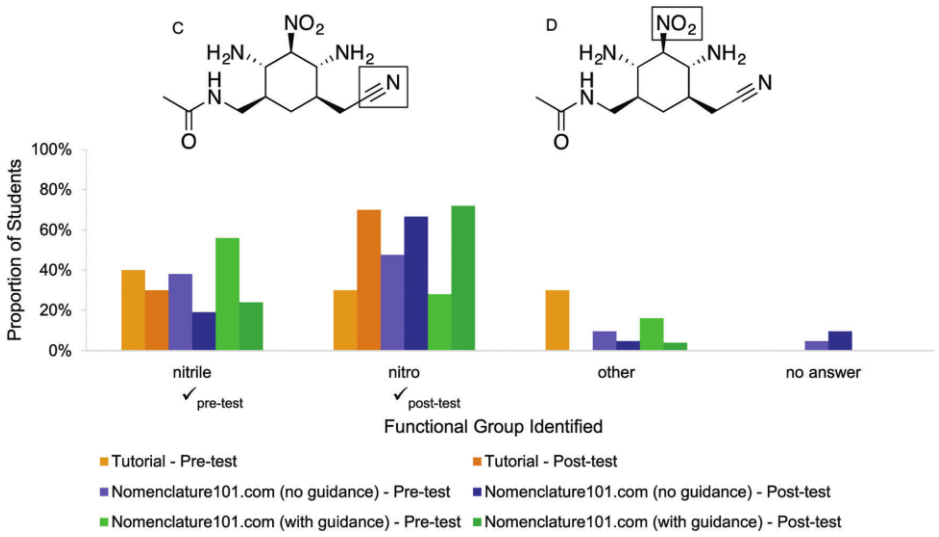


Fig. 7 Types of functional groups circled by workshop participants on (C) pre-test question 4 and (D) post-test question 4. The correct answer to each question is indicated with a box on the molecule and a checkmark on the chart. N_{tutorial} = 20, $N_{\text{nomenclature101.com (no guidance)}}$ = 21, $N_{\text{nomenclature101.com (with guidance)}}$ = 25.

participants who made incorrect or missing functional group errors, $\chi^2(1) = 1.600$, $p = 0.206$, with a small effect size, $\phi_c = 0.20$. The organization of the test bank, which is sorted by functional group, is a possible reason participants using nomenclature101.com may have experienced more significant gains in this area.

One area where tutorial participants appeared to improve much more than nomenclature101.com users was in the ability to draw the correct configuration at stereocentres (Fig. 8). The decrease in this error was significant for tutorial participants (Group 1), $\chi^2(1) = 4.912$, $p = 0.027$, with a medium effect size,

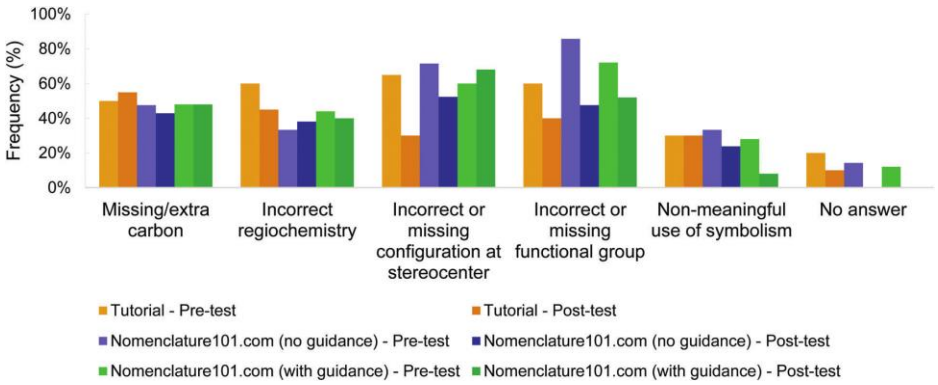


Fig. 8 Types of errors made by participants on the "draw the molecule" questions on pre-workshop and post-workshop tests. Frequencies indicate the proportion of participants who made the indicated error on at least one of the six questions. $N_{tutorial}$ = 20, $N_{nomenclature101.com\ (Unguided)}$ = 21, $N_{nomenclature101.com\ (Unguided)}$ = 25.

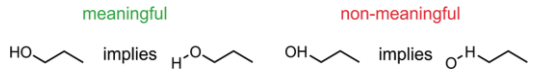


Fig. 9 A common error was the non-meaningful use of symbolism when drawing functional groups (right).

ϕ_c = 0.35. However, there was no significant change in this error for Group 2, $\chi^2(1) = 1.615$, $p = 0.341$, with a small effect size, ϕ_c = 0.20, or Group 3 participants, $\chi^2(1) = 0.347$, $p = 0.769$, with a small effect size, ϕ_c = 0.08. A possible reason for this difference is that nomenclature101.com users were free to practice whichever question types they wanted and may not have chosen to focus on stereochemistry. However, this may illustrate a lack in metacognitive skillfulness among the participants if they are unable to identify these gaps in their understanding of organic nomenclature and practice accordingly.

We found the non-meaningful use of symbolism (Fig. 9) in nearly one-third of participants' answers, but less commonly among nomenclature101.com users than among tutorial participants. While there was no change in frequency of this error among tutorial participants, there was a slight (but non-significant) decrease in the frequency of this error among Group 2 participants, and the decrease in the frequency of this error among participants in Group 3 was close to significant, $\chi^2(1) = 3.388$, $p = 0.066$, with a small effect size, ϕ_c = 0.26. This may have been a less common error among nomenclature101.com users because the drawing tool does not allow users to draw functional groups in the non-meaningful fashion; furthermore, the learning tool shows the answers with the molecules correctly drawn.

Questionnaire analysis: factor analysis and reliability

To answer our second research question, we first investigated whether the nomenclature101.com questionnaire reliably measured participants' perceptions of the usefulness and ease of use of the website (Table 3). The questionnaires are given in Appendices 3 and 4. We performed exploratory factor analysis using participants' responses to see if the questionnaire could reliably measure these latent factors (usefulness and ease of use). The Kaiser–Meyer–Olkin measure verified that our sample size was adequate, KMO = 0.852, which is well above the acceptable limit of 0.500 (Kaiser, 1974).

We performed an initial factor analysis following Kaiser's criterion, retaining factors with an eigenvalue greater than 1. This analysis yielded 3 factors that, in combination, explained 74.65% of the variance. We then estimated 95% confidence intervals for each eigenvalue, following the revised Kaiser's criterion to retain factors with eigenvalues where the entire 95% confidence interval was greater than 1. This procedure helps avoid retaining too many factors due to errors related to sample size (Larsen and Warne, 2010).

We chose to retain two factors that, in combination, explained 66.98% of the variance, but a small portion of the estimated 95% CI for our second factor is below 1 (95% CI = [0.889,2.141]). We decided to keep this second factor because of the qualitatively thematic differences between the items that loaded onto the different factors, which are supported by reasonably high factor loadings. Underlying themes between the items loading on factor 1 suggest that this factor represents participants' perceptions of the usefulness of nomenclature101.com, while factor 2 represents participants' perceptions of the ease of use of nomenclature101.com. Although questionnaire item 10 loads more strongly onto factor 1, we believe that the statement "I like that I am able to use www.nomenclature101.com outside of class" is more relevant to nomenclature101.com's ease of use than its usefulness. As each of these factors contains at least four items with loadings of 0.60 or greater, these latent factors revealed by the pattern matrix can be reliably interpreted (Guadagnoli and Velicer, 1988).

Reliability analysis of each latent factor indicated excellent homogeneity within the subscales pertaining to perceived usefulness and perceived ease of use of nomenclature101.com (Cronbach's α = 0.935 and 0.899 for factors 1 and 2 respectively). Homogeneity is a measure of the extent to which there is a single latent variable that is common to the items in the domain being measured (Revelle and Zinbarg, 2008). A limitation of this reliability analysis is that the α coefficient is intended to be used as a reliability measure of normally distributed data, while the distribution of responses to these

Table 3 nomenclature101.com questionnaire (Groups 2 and 3): principal axis factoring and reliability analysis results

Subscale mean	Item no.	Item statements, by subscales	Rotated factor loadings[a] 1	2	Mean	Std. dev.	α if item deleted
		Perceived usefulness (N = 42)					
3.95	1	The quizzes available in www.nomenclature101.com were useful in learning nomenclature.	0.804		4.10	1.10	0.920
	2	I enjoyed using the nomenclature quizzes as a way to review the material that I learned in class.	0.922		3.83	1.17	0.918
	3	The nomenclature quizzes aided me in understanding the material.	0.793		3.93	1.13	0.924
	4	I will do other nomenclature quizzes in the future.	0.768		4.17	0.93	0.926
	5[b]	It is faster to study the material myself than to learn from the nomenclature quizzes.	0.317		3.33[b]	1.00	0.948
	6	The www.nomenclature101.com website helped me improve my skill.	1.026		3.98	0.92	0.920
	7	I believe using www.nomenclature101.com will help improve my course grade.	0.905		3.98	1.00	0.922
	8	I appreciate that I will be able to access www.nomenclature101.com anywhere (with WiFi).	0.796		4.31	1.02	0.928
		Perceived Ease of Use (N = 45)					
3.94	9[b]	Technical difficulties using www.nomenclature101.com obstructed my experience.		0.562	3.67[b]	1.40	0.895
	10	I like that I am able to use www.nomenclature101.com outside of class.	0.493	0.443	4.33	0.85	0.881
	11	I felt comfortable using www.nomenclature101.com.		0.669	4.18	1.07	0.875
	12[b]	I had a hard time navigating www.nomenclature101.com.		0.717	3.93[b]	1.14	0.882
	13	The www.nomenclature101.com is well designed.		0.798	3.93	1.14	0.876
	14[b]	The pace of the activity on www.nomenclature101.com was too fast.		0.524	3.93[b]	0.78	0.908
	15	The length of time allocated to 'study' using www.nomenclature101.com was appropriate.		0.796	3.53	1.12	0.888
	16	The consistent interface design made it easy for me to follow the material.		0.806	4.02	0.89	0.882
Observed eigenvalues			9.202	1.515			
Estimated 95% confidence interval of the eigenvalues			[5.994,13.00]	[0.889,2.141]			
% of variance			57.51	9.468			
α			0.935	0.899			

[a] Values less than 0.3 are omitted. [b] Items are reverse-coded (*i.e.*, strongly disagree = 5, strongly agree = 1).

Likert-scale questionnaires is non-normal (as is often the case) (Bretz and McClary, 2015).

We also performed principal axis factoring using participants' responses to the tutorial questionnaire (Group 1), but the obtained KMO measure of 0.452 indicated inadequate sampling, as it is below the suggested threshold for exploratory factor analysis of 0.500 (Kaiser, 1974). As such, the results of factor analysis were disregarded, and the questionnaire as a whole was analyzed for reliability. A Cronbach's α value of 0.806 was obtained, indicating good internal consistency of the questionnaire as a whole. A possible limitation of this statistic is the small sample size (N = 17) from which it was obtained.

Participants thought nomenclature101.com was useful and easy to use

The Likert scale scores were fairly high across all statements indicating that participants generally agreed with positively worded items and disagreed with negatively worded items on both questionnaires and that participants in different groups responded to each item in a similar fashion. Boxplots comparing the distribution of scores between groups on individual questionnaire items can be found in Appendix 5. The subscale means that were obtained indicate that participants agree nomenclature101.com is both useful (subscale mean = 3.95) and easy to use (subscale mean = 3.94) (Table 3). Specific features that some participants liked about nomenclature101.com included the test bank's organization and large size, along with the option to customize the difficulty of quizzes. Participants also had some suggestions for improving the learning tool, such as adding the option to customize the number of quiz questions. The post-tutorial questionnaires were not compared in depth between groups. This is because differences in the learning environments may have resulted in possible differences in how the item statements would be interpreted by participants, and as such, these comparisons would not be meaningful.

Comparably high means on analogous items demonstrate that both learning environments were useful, and therefore, environment-related factors such as the pace of the workshops, the quality of the workshop facilitators, and participant comfort levels were likely not different enough between groups to confound the results of this study.

The Group 1 participant who obtained a score of 0 on both the pre- and post-tests (Fig. 5) provided many comments in their questionnaire, including: "the facilitator was good at answering

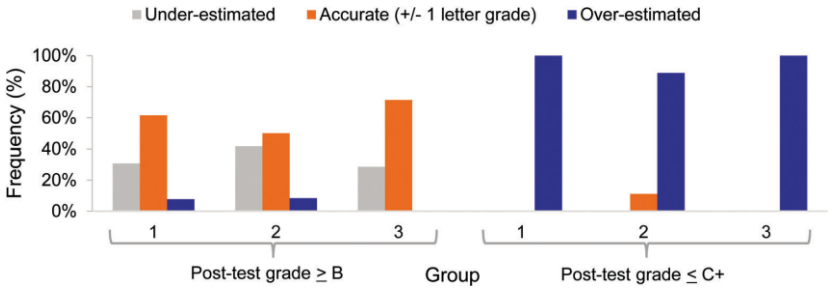


Fig. 10 Postdiction accuracy of tutorial and nomenclature101.com workshop participants. Participants were asked to postdict (*i.e.*, after completing the test) their letter grade on the post-test; this postdiction was compared to their actual score. $N_{Group\,1}$ = 20, $N_{Group\,2}$ = 21, $N_{Group\,3}$ = 25.

questions" but he went "too fast, he thought we knew everything already (I've never seen any of this before)" and "I really like how Dr. [name] does videos outside of class. That way you can pause and take good notes instead of trying to rush it." Although the tutorial was designed to build off of knowledge that we expected most participants to have, this participant did not have the anticipated prior knowledge. A site such as nomenclature101.com can benefit that type of student because the student can start learning at any point, as long as the student can identify where they should most appropriately begin.

Participants' self-evaluations were inaccurate

One of the primary factors that can limit a student's potential gains in a self-regulated learning environment is their metacognitive skillfulness, which refers to the ability of the student to assess their knowledge and regulate their cognition (Azevedo *et al.*, 2012). In a learning environment where the student must independently self-regulate their learning, they must constantly monitor their learning and adapt their strategies as necessary in order to be successful. Aspects of metacognitive skillfulness may include determining what strategies to use to solve problems and setting well-defined learning goals (Azevedo *et al.*, 2012).

The process of setting learning goals should also be informed by one's current levels of knowledge and understanding; therefore, another important aspect of metacognitive skillfulness is the ability to accurately self-assess knowledge and understanding. We asked participants to postdict (*i.e.*, "predict" after the fact) their score on an assessment task (Hawker *et al.*, 2016). Some student-centered learning environments (SCLEs) in science education exist that were designed using theories of metacognition and self-regulated learning (Azevedo *et al.*, 2010; Kinnebrew and Biswas, 2011) but nomenclature101.com does not explicitly guide or scaffold students' metacognitive thinking and the Group 1 tutorial setting moved at a unified pace based on estimated prior knowledge and processing ability.

We compared participants' self-reports of their expected grade on the post-test (postdictions) with their actual post-test scores. Because nomenclature101.com users receive constant feedback using the website, we hypothesized that Group 2 and 3 participants would make more accurate postdictions than the Group 1 (tutorial) participants. We further expected Group 3 participants to make the most accurate postdictions because they were encouraged to work according to a gradient of difficulty and according to their individual existing knowledge and abilities.

However, between groups, there was no significant difference in postdiction accuracy (Fisher's exact test: p = 0.980); weaker participants consistently over-estimated their scores on the post-test (Fig. 10). These trends in postdiction accuracy are consistent with findings from other studies (Karatjas, 2013; Hawker *et al.*, 2016) and with research showing that less skilled individuals often believe they are more competent than they are in reality, a cognitive bias commonly called the Dunning–Kruger effect (Kruger and Dunning, 1999).

We also asked Group 2 and 3 participants how often they would use nomenclature101.com and how often they recommended the tool be used. Nine participants who predicted they would get below an A– (<80%) on the post-test recommended that others use the tool more often than they said they would use it themselves. These lower achieving individuals seemed to be aware that they had room for improvement and recommended that other students use nomenclature101.com at least once per week; however, most of these individuals said they would not use the website more than once every two weeks. Their choices to use the website less often could be due to other conflicting course or academic priorities or other personal reasons.

Metacognitive skillfulness is essential to optimizing learning gains from a self-directed learning environment such as nomencla ture101.com. Based on the significant learning gains for participants that used this learning tool, we expected that they would also be reasonably good at predicting their scores, due in part to both high metacognitive skillfulness and the features of the site that facilitate the development of these skills, such as the constant feedback. One possibility is that longer-term use of the tool or further metacognitive guidance or scaffolding are needed to develop these skills.

Conclusions and implications for instruction

This study compared students' learning gains on organic chemistry nomenclature problems after one of three treatments and used a questionnaire to investigate students' perceptions of the usefulness, ease of use, and overall learning experience in their Group's learning setting. Participants were randomly

sorted into one of three groups; they completed a pre-test, participated in a one-hour session where they either received a nomenclature tutorial (Group 1), used nomenclature101.com without guidance (Group 2), or used nomenclature101.com with guidance (Group 3), then completed a post-test and questionnaire.

The learning gains were large across all three groups although there were no significant differences between groups. The most evident learning gains by the two groups that used nomenclature101.com were in the ability to correctly name and identify functional groups; this was a common error on pre-test questions that was significantly less frequent on post-test questions. nomenclature101.com has the potential to mitigate the small errors that could lead to much larger errors in understanding reactivity later on in the course.

Participants in all groups reported high levels of satisfaction with their experiences and participants in Groups 2 and 3 generally agreed with the items pertaining to the characteristics of the website. However, participants were generally inaccurate in their postdictions of their performance in the workshop, especially students with lower grades. These results are consistent with metacognition research (Kruger and Dunning, 1999; Karatjas, 2013; Hawker *et al.*, 2016) and suggest that participants are either lacking in metacognitive skillfulness (which we did not measure beyond postdiction accuracy) or that nomenclature101.com does not effectively foster the metacognitive skillfulness that is required for student-centered learning environments to be as effective as they can be. We recommend that the importance of metacognitive skillfulness be emphasized when this tool is used as part of a course and that students be provided with clear guidelines for setting effective learning goals. For example, we share the intended learning outcomes with students using a screenshot of nomenclature101.com's homepage (*e.g.*, Fig. 3) and recommend that they identify objectives they have not yet mastered (*e.g.*, by doing a quiz), then work from the simplest to the most complex tasks.

Our results show that nomenclature101.com can be a useful resource for students to practice their skills pertaining to organic nomenclature and seems to be an effective substitute for class time, if instructors wish to use class time in other ways. The most pronounced area of weakness for many Group 2 and 3 participants was the ability to draw the correct configuration of functional groups at stereocentres. As such, we recommend that class or tutorial time still be dedicated to teaching this skill or that learning outcomes related to stereochemical nomenclature be made more explicit, so that students can reflect on their skillfulness in identifying/producing the correct configuration at stereocentres more effectively.

Future work

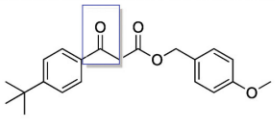
Future research could investigate how the incorporation of this resource into the organic chemistry curriculum is advocated and supported by faculty. Future work could also measure learning gains from the regular use of nomenclature101.com in a more longitudinal fashion. Metacognitive skillfulness of the research participants could be studied in more depth, for example by using existing instruments (Cooper and Sandi Urena, 2009) to measure student-reported gains in metacognitive skillfulness associated with a similar intervention as reported in this study. Metacognitive scaffolding could also be built into future online learning tools. Finally, the impact of greater (or lesser) mastery of organic chemistry nomenclature on students understanding of chemistry concepts could be studied in more depth.

Appendix 1. Pre-test

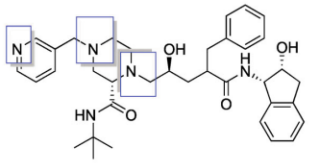
Organic Nomenclature
Pre-test ANSWERS

Write seat code here

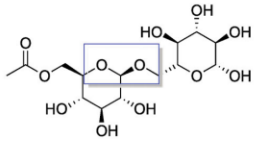
1. Circle a ketone in the molecule below:



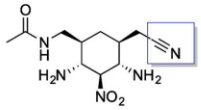
2. Circle an amine in the molecule below:



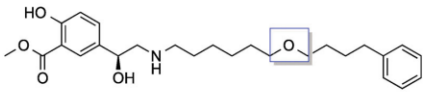
3. Circle an acetal in the molecule below:



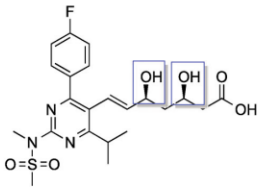
4. Circle a nitrile in the molecule below:



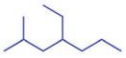
5. Circle an ether:



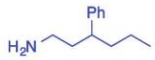
6. Circle an alcohol:



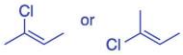
7. Draw the following molecule: 4-ethyl-2-methylheptane



8. Draw the following molecule: 3-phenylhexan-1-amine

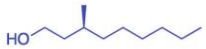


9. Draw the following molecule: 2-chlorobut-2-ene

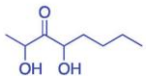


10. Draw the following molecule: 4-methylcyclohexene

11. Draw the following molecule: (*S*)-3-methylnonan-1-ol



12. Draw the following molecule: 2,4-dihydroxyoctan-3-one

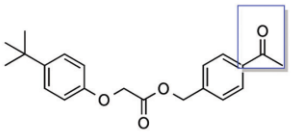


Appendix 2. Post-test

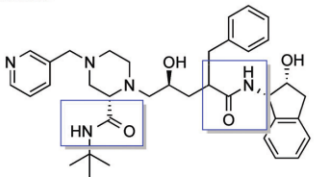
Organic Nomenclature Post-test

Write seat code here

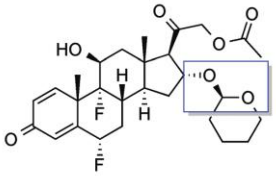
1. Circle a ketone:



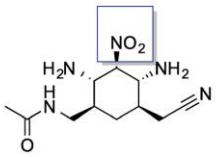
2. Circle an amide:



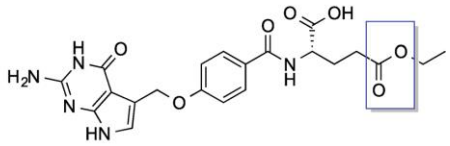
3. Circle an acetal:



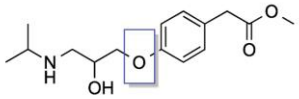
4. Circle a nitro group:



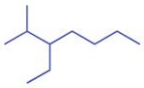
5. Circle an ester:



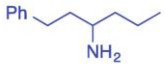
6. Circle an ether:



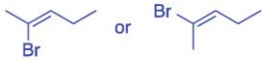
7. Draw 3-ethyl-2-methylheptane



8. Draw 1-phenylhexan-3-amine

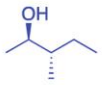


9. Draw 2-bromopent-2-ene

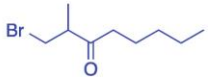


10. Draw 3-ethylcyclobutene

11. Draw (2*R*, 3*S*)-3-methylpentan-2-ol



12. Draw 1-bromo-2-methyloctan-3-one



Appendix 3. Questionnaire for Group 1

Questions for Post-Workshop Survey – Group 1

Part 1

1. What is your research code? _____
2. Which faculty, program are you currently enrolled in? _____
3. Year of program? (1, 2, 3,4) _____
4. Please indicate your student status (FT, PT, Independent, other): _____
5. Please indicate your expected letter grade for this workshop? Circle one of the choices below:

[A+, 90-100] [A, 85-89] [A-, 80-84] [B+, 75-79] [C+, 65-69] [C, 60-64] [D+, 55-59] [D, 50-54]

Part 2

Please rate the following statements	Strongly disagree 1	Disagree 2	Neither agree or disagree 3	Agree 4	Strongly agree 5
1. The tutorial was useful in learning nomenclature.	1	2	3	4	5
2. I enjoyed reviewing my tutorial lecture notes as a way to review the material that I learned in the tutorial.	1	2	3	4	5
3. The tutorial session aided me in understanding the material.	1	2	3	4	5
4. Given an online nomenclature learning tool, I would do other nomenclature quizzes in the future.	1	2	3	4	5
5. It is faster to study the material by myself than to learn from an instructor in a tutorial session.	1	2	3	4	5
6. The tutorial helped me improve my skill.	1	2	3	4	5
7. I believe attending this tutorial session will help improve my course grade.	1	2	3	4	5
8. I appreciate that I will be able to view the tutorial again anywhere with Internet access.	1	2	3	4	5
9. Some technical difficulties during the tutorial obstructed my learning experience.	1	2	3	4	5
10. I like that I will be able to review the video of this lecture outside of class.	1	2	3	4	5
11. I felt comfortable participating in the tutorial.	1	2	3	4	5
12. I had a hard time following the tutorial.	1	2	3	4	5
13. The tutorial was well designed.	1	2	3	4	5
14. The pace of the tutorial was too fast.	1	2	3	4	5
15. The length of the time allocated for the tutorial was appropriate.	1	2	3	4	5
16. The consistent organization of the content made it easy for me to follow the material presented in the tutorial.	1	2	3	4	5

Part 3

17. What did you like about this tutorial?

18. What did you not like about this tutorial?

19. Please provide one suggestion for improvement:

Questions for Post-Workshop Survey – Group 1

20. A video of this tutorial will be made available to you. Will you use the video to review the instructor-led tutorial? Do you think it will have an impact on the success in your course?
21. How is the www.nomenclature101.com tool being promoted in your class, if at all?
22. On average how much time do you anticipate using www.nomenclature1101.com throughout the semester?
    a. Daily
    b. Weekly
    c. Once every two weeks
    d. More than twice a month
23. How much time would you recommend students spend using nomenclature101.com?
24. Any other comments?

Appendix 4. Questionnaire for Groups 2 and 3

Post Workshop (Survey (Group 2, Group 3)

Part 1 Details

1. What is your research code? _____
2. Which faculty, program are you currently enrolled in? _____
3. Year of program? (1, 2, 3,4) _____
4. Please indicate your student status (FT, PT, Independent, other): _____
5. Please indicate your expected letter grade for this workshop? Circle one: [A+, 90–100] [A, 85–89] [A-, 80–84] [B+, 75–79] [C+, 65–69] [C, 60–64] [D+, 55–59] [D, 50–54]

Part 2 Tutorial

Please rate the following statements:	Strongly disagree 1	Disagree 2	Neither agree or disagree 3	Agree 4	Strongly agree 5
1. The quizzes available in www.nomenclature101.com were useful in learning nomenclature.	1	2	3	4	5
2. I enjoyed using the nomenclature quizzes to review the material that I learned in class.	1	2	3	4	5
3. The nomenclature quizzes aided me in understanding the material.	1	2	3	4	5
4. I will do other nomenclature quizzes in the future.	1	2	3	4	5
5. It is faster to study the material by myself than to learn from the nomenclature quizzes.	1	2	3	4	5
6. The www.nomenclature101.com website helped me improve my skill.	1	2	3	4	5
7. I believe using www.nomenclature101.com will help improve my course grade.	1	2	3	4	5
8. I appreciate that I am able to access www.nomenclature101.com anywhere (with WIFI).	1	2	3	4	5
9. Technical difficulties using www.nomenclature101.com obstructed my experience.	1	2	3	4	5
10. I like that I am able to use www.nomenclature101.com outside of class.	1	2	3	4	5
11. I felt comfortable using www.nomenclature101.com .	1	2	3	4	5
12. I had a hard time navigating www.nomenclature101.com.	1	2	3	4	5
13. The www.nomenclature101.com is well designed.	1	2	3	4	5
14. The pace of the activity on www.nomencature101.com was too fast.	1	2	3	4	5
15. The length of the time allocated to 'study' using www.nomenclature101.com was appropriate.	1	2	3	4	5
16. The consistent interface design made it easy for me to follow the material.	1	2	3	4	5

Post Workshop (Survey (Group 2, Group 3)

Part 3 Nomenclature101.com

17. What did you like about www.nomenclature101.com ?

18. What did you not like about www.nomenclature101.com ?

19. Navigation

When navigating the quiz, there are many buttons and icons with different functions available to you. Please rate the following buttons/functionality:

The following buttons/tools found in the www.nomenclature101.com tool were **useful/helpful**:	Strongly disagree 1	Disagree 2	Neither agree or disagree 3	Agree 4	Strongly agree 5	N/A Didn't use this button
a) Submit your answer	1	2	3	4	5	
b) See the answer	1	2	3	4	5	
c) Submit the quiz, see your results	1	2	3	4	5	
d) Export quiz to PDF	1	2	3	4	5	
e) Important Facts	1	2	3	4	5	
f) Hints	1	2	3	4	5	
g) Decision Tree	1	2	3	4	5	
h) Important Notes	1	2	3	4	5	
i) Toggle to hide/show the correct response.	1	2	3	4	5	
j) Quiz Dashboard	1	2	3	4	5	

20. Please provide one suggestion for improvement.

21. Will your use of www.nomenclature101.com have an impact on your success in your course?

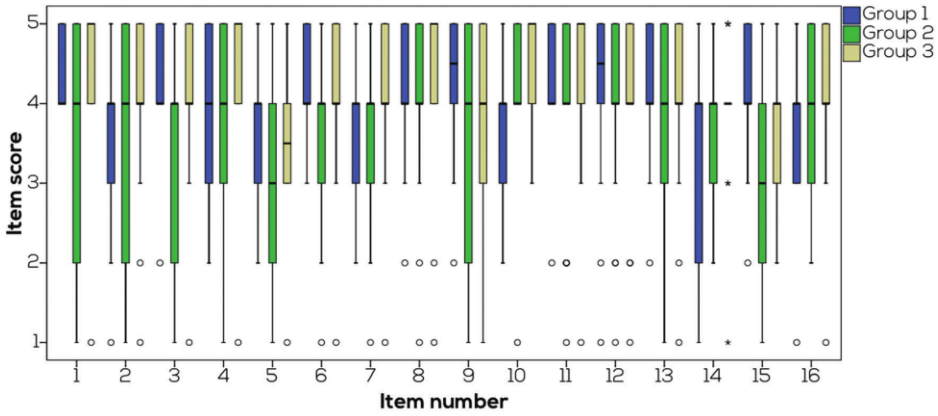
22. How is the www.nomenclature101.com tool being promoted in your class, if at all?

23. On average how much time do you anticipate using www.nomenclature1101.com throughout the semester?
    A. Daily B. Weekly C. Once every two weeks D. More than twice a month

24. How much time would you recommend students spend using nomenclature101.com?

25. Any other comments?

Appendix 5. Distribution of responses to questionnaire items



Box plots to compare distribution of scores on analogous questionnaire items. Item statements are listed Appendices 3 and 4.

Acknowledgements

We thank the University of Ottawa for funding this research, Hayley Sprigings for her contributions, Bradley Cousins and Samuel Pazicni for helpful discussions, facilitators for giving their time and expertise, and the participants.

References

Akaygun S. and Jones L. L., (2014), How Does Level of Guidance Affect Understanding When Students Use a Dynamic Simulation of Liquid–Vapor Equilibrium? in Devetak I. and Glažar S. A. (ed.) *Learning with Understanding in the Chemistry Classroom*, Dordrecht: Springer Netherlands, pp. 243–263. DOI: 10.1007/978-94-007-4366-3_13.

Azevedo R., Johnson A., Chauncey A. and Burkett C., (2010), Self-regulated Learning with MetaTutor: Advancing the Science of Learning with MetaCognitive Tools, in Khine M. S. and Saleh I. M. (ed.) *New Science of Learning*, New York, NY: Springer New York, pp. 225–247, DOI: 10.1007/978-1-4419-5716-0_11.

Azevedo R., Behnagh R. F., Duffy M., Harley J. M. and Trevors G., (2012), Metacognition and Self-Regulated Learning in Student-Centered Learning Environments, in D. Jonassen and S. Land (ed.) *Theoretical Foundations of Learning Environments*, 2nd edn, New York, NY: Prentice Hall.

Bodé N. E. and Flynn A. B., (2016), Strategies of Successful Synthesis Solutions: Mapping, Mechanisms and More, *J. Chem. Educ.*, **93**(4), 593–604. DOI: 10.1021/acs.jchemed.5b00900.

Bolliger D. U. and Supanakorn S., (2011), Learning styles and student perceptions of the use of interactive online tutorials, *Br. J. Educ. Technol.*, **42**(3), 470–481. DOI: 10.1111/j.1467-8535.2009.01037.x.

Bonawitz E., Shafto P., Gweon H., Goodman N. D., Spelke E. and Schulz L., (2011), The double-edged sword of pedagogy: instruction limits spontaneous exploration and discovery, *Cognition*, **120**(3), 322–330. DOI: 10.1016/j.cognition.2010.10.001.

Bretz S. L., (2001), Novak's Theory of Education: Human Constructivism and Meaningful Learning, *J. Chem. Educ.*, **78**(8), 1107, DOI: 10.1021/ed078p1107.6.

Bretz S. L. and McClary L., (2015), Students' Understandings of Acid Strength: How Meaningful Is Reliability When Measuring Alternative Conceptions? *J. Chem. Educ.*, **92**(2), 212–219, DOI: 10.1021/ed5005195.

Chamberlain J. M., Lancaster K., Parson R. and Perkins K. K., (2014), How guidance affects student engagement with an interactive simulation, *Chem. Educ. Res. Pract.*, **15**(4), 628–638, DOI: 10.1039/C4RP00009A.

Cooper M. M. and Sandi Urena S., (2009), Design and Validation of an Instrument To Assess Metacognitive Skillfulness in Chemistry Problem Solving, *J. Chem. Educ.*, **86**(2), 240–246, DOI: 10.1021/ed086p240.

Flynn A. B. and Ogilvie W. W., (2015), Mechanisms before Reactions: A Mechanistic Approach to the Organic Chemistry Curriculum Based on Patterns of Electron Flow, *J. Chem. Educ.*, **92**(5), 803–810, DOI: 10.1021/ed500284d.

Flynn A. B., Caron J., Laroche J., Marcoux C. and Richard G., (2013), *Organic chemistry nomenclature: an online learning tool*, Presented at the Canadian Network for Innovation in Education.

Flynn A. B., Caron J., Laroche J., Daviau-Duguay M., Marcoux C. and Richard G., (2014), nomenclature101.com: A Free, Student-Driven Organic Chemistry Nomenclature Learning Tool, *J. Chem. Educ.*, **91**(11), 1855–1859, DOI: 10.1021/ed500353a.

González-Cruz J., Rodríguez-Sotres R. and Rodríguez-Penagos M., (2003), On the convenience of using a computer simulation to teach enzyme kinetics to undergraduate students with

biological chemistry-related curricula, *Biochem. Mol. Biol. Educ.*, **31**(2), 93–101, DOI: 10.1002/bmb.2003.494031020193.

Guadagnoli E. and Velicer W. F., (1988), Relation to sample size to the stability of component patterns, *Psychol. Bull.*, **103**(2), 265–275, DOI: 10.1037/0033-2909.103.2.265.

Guskey T. R., (2002), Does It Make a Difference? Evaluating Professional Development, *Educ. Leader.*, **59**(6), 45–51.

Guskey T. R., (2010), Professional Development and Teacher Change, *Teachers Teach.: Theor. Pract.*, **8**(3), 381–391, DOI: 10.1080/135406002100000512.

Harshman J. and Yezierski E., (2015), Guiding teaching with assessments: high school chemistry teachers' use of data-driven inquiry, *Chem. Educ. Res. Pract.*, **16**(1), 93–103, DOI: 10.1039/C4RP00188E.

Hawker M. J., Dysleski L. and Rickey D., (2016), Investigating General Chemistry Students' Metacognitive Monitoring of Their Exam Performance by Measuring Postdiction Accuracies over Time, *J. Chem. Educ.*, **93**(5), 832–840, DOI: 10.1021/acs.jchemed.5b00705.

Kaiser H. F., (1974), An index of factorial simplicity, *Psychometrika*, **39**(1), 31–36, DOI: 10.1007/BF02291575.

Karatjas A. G., (2013), Comparing College Students' Self-Assessment of Knowledge in Organic Chemistry to Their Actual Performance, *J. Chem. Educ.*, **90**(8), 1096–1099, DOI: 10.1021/ed400037p.

Kinnebrew J. and Biswas G., (2011), Modeling and Measuring Self-Regulated Learning in Teachable Agent Environments, *J. e-Learn. Knowl. Soc.*, **7**(2), 19–35.

Kornell N., (2009), Optimising learning using flashcards: spacing is more effective than cramming, *Appl. Cognit. Psychol.*, **23**(9), 1297–1317, DOI: 10.1002/acp.1537.

Krippendorff K., (2004), Reliability in Content Analysis, *Hum. Comm. Res.*, **30**(3), 411–433, DOI: 10.1111/j.1468-2958.2004.tb00738.x.

Kruger J. and Dunning D., (1999), Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments, *J. Pers. Soc. Psychol.*, **77**(6), 1121–1134, DOI: 10.1037/0022-3514.77.6.1121.

Larsen R. and Warne R. T., (2010), Estimating confidence intervals for eigenvalues in exploratory factor analysis, *Behav. Res. Meth.*, **42**(3), 871–876, DOI: 10.3758/BRM.42.3.871.

Novak J. D., (2011), A theory of education: meaningful learning underlies the constructive integration of thinking, feeling, and acting leading to empowerment for commitment and responsibility, *Meaningful Learning Review*, **1**(2), 1–14.

Oriol M. D., Tumulty G. and Snyder K., (2010), Cognitive apprenticeship as a framework for teaching online, *J. Online Learn. Teach.*, **6**(1), 210–217.

Patton M. Q., (2014), Qualitative Research and Evaluation Methods, 4 edn, Thousand Oaks, CA: SAGE Publications.

Revelle W. and Zinbarg R. E., (2008), Coefficients Alpha, Beta, Omega and the glb: Comments on Sijtsma, *Psychometrika*, **74**(1), 145–154, DOI: 10.1007/s11336-008-9102-z.

Top Hat, (2016), Top Hat. Tophat.com. Retrieved June 2016, from https://tophat.com/.

Vygotsky L. S., (1978), in Cole M. Souberman E. John-Steiner V. and Scribner, S. (ed.), *Mind in Society*, Cambridge, MA: Harvard University Press.

Chapter 3 Studying problem-solving skill development in organic synthesis through strategy-based workshops

Introduction

Synthesis-type problems are among the greatest challenges in organic chemistry courses; where students must draw upon their knowledge of chemical principles and reactions to propose a synthesis of a given target molecule from a given starting material. While educators design synthesis problems so that they may be solved using reactions that have been taught, students must still recall all of the reactions and skills from their previous courses or semesters. Our overall goal is to address the gap between educators' expectations for students' problem-solving skills and their current abilities and help identify a pathway for the development of expertise.

Prior research on problem solving for synthesis in organic chemistry has focused on students' reasoning processes when solving these types of problems.[1-4] In these studies, the researchers observed that undergraduate students primarily used rule-based reasoning to propose solutions. Case-based reasoning was the next most common type observed, where solutions are based on recalling specific learned material that a problem solver has judged to be sufficiently similar to the problem at hand. Model-based reasoning was the least common form observed in these studies.

Sevian and co-workers have previously described how these different modes of reasoning can be distinguished in terms of abstractness.[3] Within the context of problem-solving based on instructional activities, a representation would be defined as being more or less abstract based on the extent to which it incorporates elements from the students' prior knowledge in addition to specific elements from the relevant instructional activity.[5] Therefore, these forms of reasoning can be distinguished by how abstracted the student's connection between their existing knowledge of relevant reactions was from the problem at hand. A smaller degree of abstraction in how the student linked their existing knowledge to a given problem (*i.e.*, a surface-level analysis) would be considered either case-based or rule-based reasoning. While model-based reasoning was the least common, it was also the most fruitful. As such, students would likely benefit from learning how to use problem solving models (such as retrosynthetic

analysis) that allow them to make higher-level abstractions to their understanding of reactivity patterns when determining what reactions would best suit the construction of a desired bond.

The study outlined herein was informed by prior work on how students approach problem solving for organic synthesis. An initial study focused on scaffolding students' learning of retrosynthetic analysis skills in large classes using a classroom response system; the results of this study showed that the majority of students were able to identify a suitable forward reaction in association with a bond they were asked to disconnect, even using complex target molecules such as discodermolide and a Euphorbiaceae diterpene.[6]

A more in-depth qualitative study explored how undergraduate students worked through these scaffolded synthesis-type problems;[7] some major findings included:

- Students lacked defined problem-solving strategies when they could not remember a solution
- Students struggled to apply generalized reactions to a specific synthesis problem
- After the conventions of synthon construction were briefly explained to a subset of participants, these individuals were able to quickly apply this strategy to successfully identify a reaction that could make an indicated bond within a complex molecule and identify suitable starting materials

In a subsequent study, we analyzed students' responses to problems about organic synthesis on final exams and found that students who explicitly demonstrated certain well-defined strategies in association with one another were more successful in solving these problems than students who did not demonstrate a well-defined strategy.[8] These strategies included mapping, identifying atoms added/removed and bonds formed/broken, and retrosynthetic analysis.

A major implication of our prior work was that students need opportunities to explicitly practice using specific strategies to solve "authentic"[9] problems. This implication is supported by other work that has suggested that using abstract visual representations is an important part of the problem-solving process.[3,5,10,11] Accordingly, teaching and learning activities should include formative and summative assessments[12] on correctly applying these strategies as much as on the ability to actually propose a synthesis of a given target molecule.

To build on previous findings, this study's goals are to (1) compare the effects of different learning activities on synthesis skills: using synthesis problem solving strategies compared to learning activities using reaction types (as synthesis is commonly taught), and (2) determine students' progress toward achieving learning outcomes related to metacognitive and strategic knowledge. We hypothesized that strategy-based learning activities would be more effective than synthesis learning activities centered on reaction types. We investigated this hypothesis by offering three synthesis learning workshops, guided by constructivist theories of learning.

Research Questions

We investigated the following research questions:

RQ1 What differences in the use of strategies/degree of proficiency might be observed in Organic Chemistry 2 students' responses to synthesis-type problems following workshops focusing on different strategic approaches to solving synthesis problems?

RQ2 What differences in metacognitive skillfulness and attitudes toward chemistry might be observed between participants who provide successful and unsuccessful solutions to synthesis problems?

The data for this research was collected from four main sources: a synthesis learning workshop, a classroom observation protocol,[13] and two inventories: the Metacognitive Activities Inventory (MCAI),[14] and the revised Attitudes Toward the Subject of Chemistry Inventory (ASCIv2).[15]

Below we describe the theoretical framework used in the workshop's design, the intended learning outcomes, then the study itself.

Theoretical Framework

Constructivism is a theory maintaining that knowledge and understanding are constructed in the mind of the learner through experiences that can be meaningfully connected with their prior knowledge and understanding.[16,17] Social constructivists propose that learning occurs in the cognitive space between a person's individual capabilities and their capabilities when collaborating with others who possess greater knowledge and skills. This approach of

knowledge construction through collaboration facilitates the learning process; collaborative activities provide opportunities for students to link their current experiences to their prior knowledge at a skill level that is beyond what they are capable of by themselves.[11,18-20] Therefore, our focus when designing the workshops was to develop learning activities that allowed students to work in groups (with the assistance of a facilitator who had mastery of the intended learning outcomes) to solve problems that were constructively aligned with the learning outcomes we wanted participants to achieve.

Constructive alignment is achieved when students are explicitly told what they are expected to be able to achieve (using specific action verbs) following a learning activity, the learning activity gives students the opportunity to engage directly with the knowledge or skills described in the learning outcome, and the final assessment task directly assesses those knowledge and skills using the same action verbs as in the intended learning outcome(s) being assessed.[21,22]

While social constructivist theories of learning provided the basis for the pedagogy used in the workshops, the research design was also informed by theories of meaningful learning related to constructivist paradigms for how learning occurs. Novak described meaningful learning as being opposed to rote memorization and describes the following criteria for meaningful learning: first, one requires prerequisite knowledge that acts as an anchor for new information to be assimilated into those existing knowledge structures. This new information must be consciously perceived by the learner as relevant to their existing knowledge, and they must make an affective commitment to make meaningful connections between their prior knowledge and the new information.[23] As such, students' metacognitive skillfulness and belief attitudes toward the subject of chemistry are important mitigating factors to investigate.

Intended Learning Outcomes for the Workshops

We focused on the following learning outcomes related to organic synthesis,[7] many of which are reflected in the key strategies that have been observed in the analysis of successful solutions to synthesis problems.[8] Students who have achieved these learning outcomes can:

- Map atoms and electrons between starting materials and products
- Analyze regiochemical patterns in reaction products

- Analyze stereochemical patterns in reaction products
- Analyze bonds formed and broken for possible reaction types
- Analyze possible mechanisms (including using the electron-pushing formalism as a tool)
- Use retrosynthetic analysis (*e.g.*, synthon approaches) to identify possible synthetic pathways

Methods

The data for this research was collected from four main sources: a synthesis learning workshop, a classroom observation protocol,[13] and two inventories: the Metacognitive Activities Inventory (MCAI),[14] and the revised Attitudes Toward the Subject of Chemistry Inventory (ASCIv2). The synthesis learning workshop was the primary means by which we investigated our first research question. During the workshops, observers completed an observation protocol to measure the extent to which the workshops gave participants the opportunity for group problem solving activities and avoided a passive lecture format.

We administered the MCAI and ASCIv2 to investigate our second research question; based on Novak's theory of meaningful learning, we hypothesized that participants' metacognitive skillfulness and attitudes toward chemistry could be mitigating factors in terms of whether or not the workshop would be beneficial. The authors' Institutional Review Board (IRB) approved all stages of this project and participants provided informed consent.

Setting and Course

This research was conducted in Organic Chemistry 2 (OC2) courses taught at a large, research-intensive Canadian university. OC2 is the students' second semester of organic chemistry. OC1 is offered in the winter semester of students' first year of their studies, and OC2 is offered in the summer and fall. Students can take these courses in English or French.

OC2 is a 12-week course consisting of two weekly classes (200–400 students total, 1.5 hours each, mandatory, lecture or flipped format) and a tutorial session (typically attended by ~20–30 students,1.5 hours, optional, also called a recitation or discussion group). Assessment for the course is generally comprised of two midterms, a final exam, online homework

assignments, and class participation using a classroom response system. The course is composed of ~75% Faculty of Science students, ~17% Faculty of Health Sciences students, and ~8% students from other faculties. The general topics covered in OC2 include reactions with σ electrophiles (S_N1/S_N2/E1/E2, oxidations), introduction to ^{1}H NMR, additions to π electrophiles with leaving groups, and reactions with π nucleophiles (*e.g.*, aldol reaction). A full description of the patterns of mechanisms organic chemistry curriculum at this institution has been reported previously.[24]

Synthesis strategies are taught in the latter third of the course, including retrosynthetic analysis and a synthon approach. The aforementioned learning outcomes are taught, with opportunities to practice in class. Additional synthesis questions are practice in class an in a problem set in formats described previously.[6] The final exam contained a synthesis question that included proposing a synthesis of a target compound, with instructions to include analysis worth three points (identify the bonds broken in the starting material, bonds formed in the product, stereochemistry, and regiochemistry, as applicable) and a note that a retrosynthetic analysis, synthon approach, and mechanisms were not required but could be used nevertheless.

Participants

Participants for the workshop and inventory stages of the study were recruited from both sections of OC2 in the fall of 2016. An announcement about the workshop was posted on the course website, and students were also told about the workshops in class by their instructor. As part of the informed consent process, students enrolled in OC2 were explicitly told that the workshops would focus on helping us determine effective methods for teaching organic synthesis and that the workshop was designed to help them solve organic synthesis problems using strategies that past successful students used. Students were also permitted to attend the workshops without consenting for their data to be used for this research study (*i.e.*, they could just come for extra help with synthesis problem solving for the course, as they knew it was going to be assessed on their final exam).

In total, 44 workshop participants who completed both the pre-test and post-test consented to the use of their data; 10 completed workshop A, 20 completed workshop B, and 14 completed workshop C (Figure 3-1). The following demographic data were collected:

- 70% (*N* = 31) female, 30% (*N* = 13) male (institution: 59% female, 41% male)
- 89% (*N* = 39) second year students, 11% (*N* = 5) third year students
- 86% (*N* = 38) Faculty of Science students, 7% (*N* = 3) Faculty of Health Science students, 7% (*N* = 3) Faculty of Social Science students

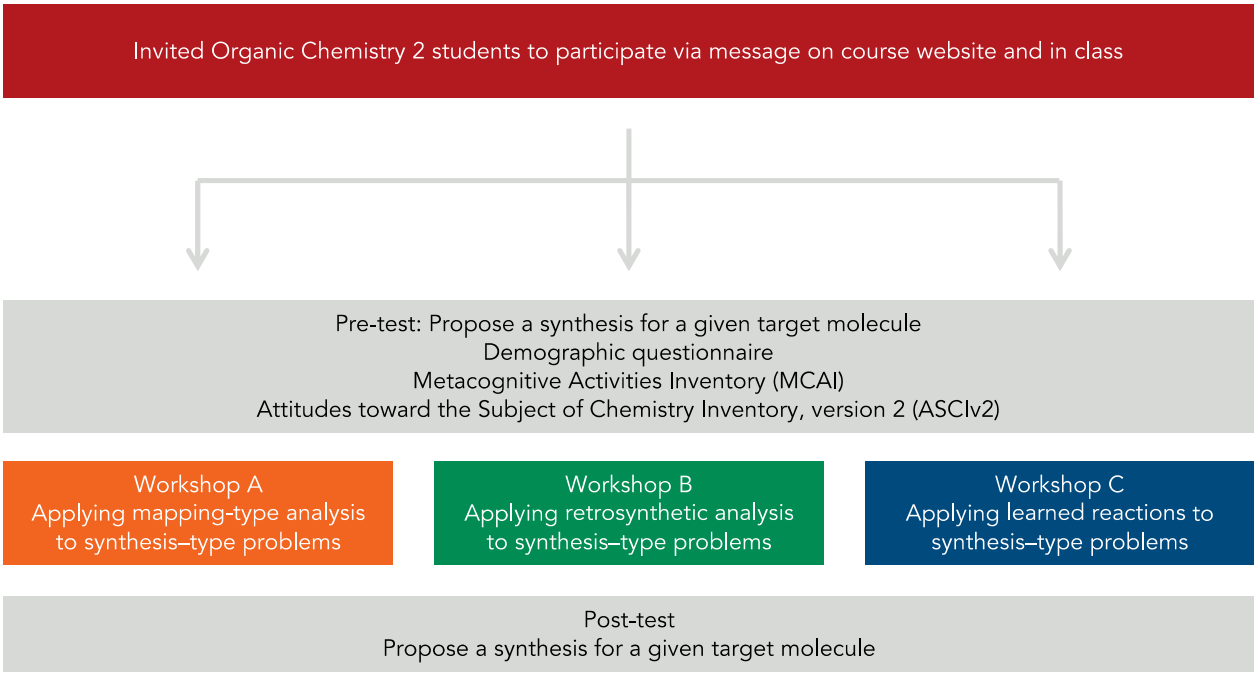


Figure 3-1. An overview of the study's data collection methods.

Data Collection

To evaluate the efficacy of the strategic models that we think successful problem solvers apply to "propose a synthesis"-type problems,[8] we offered three simultaneous organic synthesis workshops to OC2 students. An overview of the study and workshops is provided in Figure 3-1.

Participants were randomly placed into one of the three workshops, each of which provided them the opportunity to develop problem-solving skills for organic synthesis in a different setting. We thought addressing all skills in a single workshop would be too overwhelming for participants. All the key strategies were provided to participants in all three workshops as intended learning outcomes, and they were given these learning outcomes as a "checklist" as steps to complete when solving any synthesis problem.

In the first setting (Group A), participants were taught how to apply mapping-type analysis to synthesis problems, in which the solver identifies the atoms of a given starting material within a given target to identify bonds that were formed/broken and subsequently identify associated reactions, similar to the approach we have previously observed in successful answers to synthesis problems. In the second setting (Group B), participants were taught the principles of synthon–based retrosynthetic analysis, and practiced skills related to identifying useful retrosynthetic disconnections, constructing synthons based on those disconnections, and identifying related reagents, similar to how experts solve synthesis problems. In the final setting (Group C), participants completed a series of learning activities based on reactions-based analysis. These learning activities included identifying suitable reactions for forming specific bond types, predicting the outcomes of reactions given starting materials and reagents, and suggesting ideal reagents for accomplishing given chemical transformations. This is similar to how synthesis is often explained in undergraduate-level textbooks, which often contain sections regarding synthetic analysis, but do not always integrate those strategies into synthetic examples presented in other sections.[25-27]

The workshops were approximately two hours in duration, including the time given to complete the pre-test and post-test (20 minutes each). Workshop A was facilitated by the first author of this research, and workshops B and C were facilitated by post-doctoral fellows in chemistry education research, each of whom had prior experience as a teaching assistant in organic chemistry courses at the undergraduate level. The workshops followed a group active learning format; while active learning generally refers to any instructional method that engages students in the learning process (as opposed to a passive lecture), group active learning places a specific emphasis on group discussion during these activities.[28]

The choice of this learning format is aligned with constructivist theories of learning and supported by the success that has been observed using variations of this format such as process-oriented guided inquiry learning (POGIL)[29-31] and the flipped classroom[20,32,33] in organic chemistry. The workshop's learning activities were designed to be constructively aligned[22] with the intended learning outcomes for OC2[7] and with the results of our previous research on students' use of problem–solving strategies.[8] A more detailed description of the workshops is provided in the following sections.

Detailed Overview of the Workshops

In each workshop, participants were first provided with the intended learning outcomes (listed in the following section). The facilitator then led participants through a series of learning activities, which were specific to each workshop. The problems were designed to be unfamiliar to the participants, in that they were not given in previous classes, practice activities, or assessments.

The learning activities in each workshop consisted of formative assessment activities to allow participants to actively practice applying the material to authentic problems. Formative assessment refers to activities that provide immediate feedback to students about their skills and abilities with respect to the learning outcome currently being learned. Instructors can also use the outcomes of formative assessment tasks to plan their next steps, for instance, whether they should spend more time on a given learning outcome. Formative assessment is distinct from summative assessment, where feedback serves primarily as a gauge of how well a student has satisfied the intended learning outcomes of a course.[12] In our workshops, the goals when providing feedback were to:

- Provide participants with a correct answer (there may have been more than one correct answer) to each learning activity
- Explain the strategies the facilitator used (the conventions of how the strategy is meant to be applied, and the process by which the facilitator would apply it)
- Discuss incorrect answers, and possible common mistakes that would lead to those answers

The decision to use formative assessment tasks in the workshops was based in part on Bhattacharyya and Bodner's finding that in order for organic chemistry students to develop into knowledgeable practitioners of organic chemistry, "scaffolding in the form of feedback on [an] authentic activity" was necessary.[9]

Several decisions were made to make the material both useful and engaging to participants. Raker and Towns suggest that problems should be "authentic" with respect to the literature, and that the statement of a given problem should provide contextual information that includes the problem's broader applications outside the immediate scope of the course.[34] Authenticity describes how an activity reflects the day-to-day activities of practitioners. For a

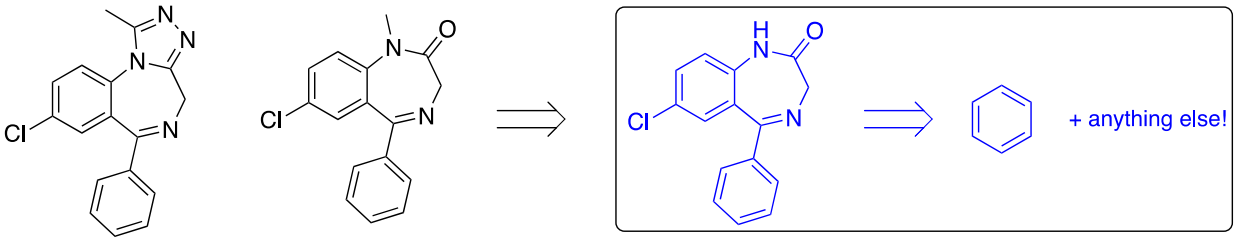
student to be able to properly learn the conceptual and practical tools of a discipline that are actively applied by practitioners in that discipline, they need to have the opportunity to engage in learning activities that accurately reflect the activities of practitioners.[35] For example, synthetic planning problems (retrosynthetic analysis and developing proposed routes) are a key part of practice for synthetic organic chemists,[34] so instruction on this skill would best serve students if it actively incorporated learning activities that gave them the opportunity to engage in these activities.

The notion of authenticity is supported by Bhattacharyya and Bodner, who suggest that when graduate students in organic chemistry develop a "perception of realness" of the synthetic problems they are faced with, they begin to approach the course material as something to be learned for their futures, and not just for their final exam. The consequence of this is a greater motivation to learn.[9] Following these suggestions, we based several of the problems in the workshops on synthesis examples from the research literature or on target molecules that we expected might hold some relevance to participants (*e.g.,* common pharmaceuticals).

Each workshop concluded with a problem-solving activity (Figure 3-2) where participants were given a synthetic problem to be completed as a group, so that they could practice applying what they learned in the workshop to an authentic problem. The facilitator then walked the group through a solution in a stepwise fashion, asking participants to contribute their ideas for how they approached the problem and what reagents they proposed for each synthetic step.

Diazepam (Valium) and alprazolam (Xanax) are both examples of benzodiazepine-class drugs, often prescribed to treat anxiety disorders. Despite being notorious as drugs of abuse, ~7.5M prescriptions were issued in Canada between 2012-2015 (for diazepam and alprazolam combined).

Both of these benzodiazepines are synthesized from the same precursor. Propose a synthesis of this precursor (on the far right), starting from benzene and using any other reagents you deem necessary.



alprazolam (Xanax) diazepam (Valium)

γ-aminobutyric acid (GABA) receptor agonists

Figure 3-2. The concluding problem-solving activity used in all three workshops.

Intended Learning Outcomes for the Workshops

Given a synthetic target and starting material(s):

- Map the atoms of the starting materials in the target
    - Identify the site(s) of bond formation and cleavage
    - Identify atoms added to/removed from the starting material(s) to form the target
    - Identify functional groups in the starting material(s) and target
    - Describe functional group transformations, oxidation levels
- Describe the regiochemical features of the starting material(s) and target
    - Compare and contrast those features between the starting material(s) and target
- Describe the stereochemical features of the starting material(s) and target
    - Compare and contrast those features between the starting material(s) and target
- Design a retrosynthetic analysis
    - Brainstorm reactions that could accomplish the transformations needed to reach the target
        - Associate functional group patterns (*e.g.,* 1,3–dicarbonyl, *syn/anti* vicinal dihalide, *ortho/meta/para* disubstituted aryl ring) with a suitable reaction
        - Associate bond types (C-O, C=C, C-X, Ar-R) with a suitable reaction
        - Recognize functional group patterns in a synthetic target
    - Given a list of reagent choices, find a suitable reaction to accomplish a necessary transformation
    - Decide on an appropriate order of reactions
    - Identify potentially competing reactions/side products
- Draw reaction mechanisms to explore options and/or provide support/rationale to a proposed synthetic step

Workshop A: Mapping-type analysis

Participants in Group A were given instruction that outlined key strategies associated with successful synthetic problem solving, which included mapping, identifying atoms added/removed, identifying bonds formed/broken, regiochemical analysis, and stereochemical analysis. The learning activities in strategy-based learning workshop were designed to explicitly ask participants to perform one or more of the activities described in the intended learning outcomes for the workshop, with the hypothesis that students would be able to better propose a synthesis after developing these skills from purposeful practice with the requisite problem-solving skills.

The first learning activity from this workshop is provided in Figure 3-3, where participants were shown the pinacol-like coupling step (an unfamiliar reaction to these

students) in Nicolaou and co-workers' total synthesis of paclitaxel (Taxol)[36] and asked to map the atoms so they could identify the site where the C–C bond formed in this reaction. Proposing reaction conditions for the forward reaction would have been beyond the scope of what participants had learned in their courses up to that point; however, they could still be reasonably expected to map the atoms and identify the site of the bond-forming reaction.

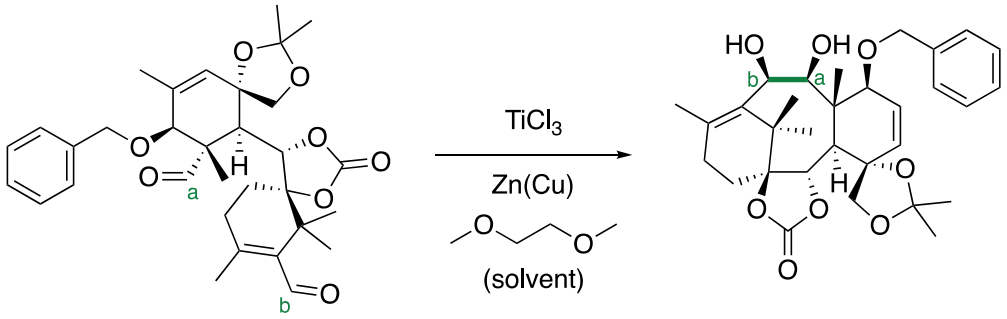


Figure 3-3. A mapping activity used in Workshop A. The site of the C–C bond forming reaction carried out in this synthetic step is indicated in green, with the atoms labeled "a" and "b".

This first learning activity was not assessed; the solution was discussed as a group. Participants were also asked to complete formative assessment tasks using a classroom response system (Top Hat). An example of one of these formative assessment tasks is provided in Figure 3-4, where participants were asked to map the atoms between geranyl pyrophosphate and (+)-α-pinene. This problem was based on the biosynthesis of (+)-α-pinene[37] and was chosen to give participants the opportunity to map between molecules with an all-hydrocarbon structure, requiring them to use more nuanced landmarks than heteroatoms. Participants were provided with immediate feedback, which discussed the correct answer, and the general strategy that the facilitator followed to reach that answer (finding a landmark functional group and basing the location of other atoms off that landmark).

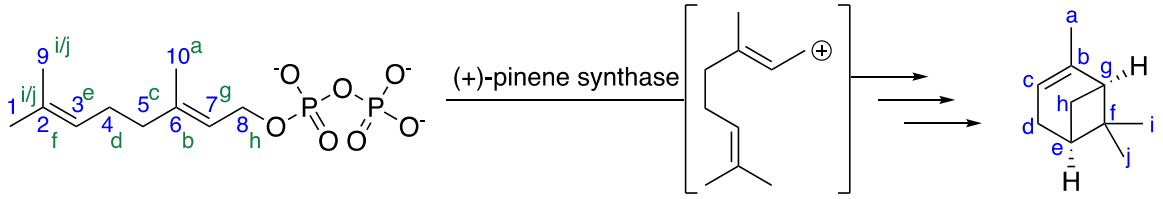


Figure 3-4. One of the mapping problems from workshop A. Participants were asked to map the atoms between geranyl pyrophosphate and (+)-a-pinene, then enter their answer in order from the lettered atom that corresponds to atom 1 to the lettered atom that corresponds to atom 10. Solution indicated in green (i/j were considered interchangeable).

Next came a different type of mapping problem, where they were asked to identify the bonds formed in the synthesis of azithromycin, given the indicated starting materials (Figure

3-5). This problem was based on Kim and Kang's total synthesis of azithromycin.[38] Participants were told that they could check the plausibility of their mapping by thinking about a plausible mechanism for the formation of each bond.

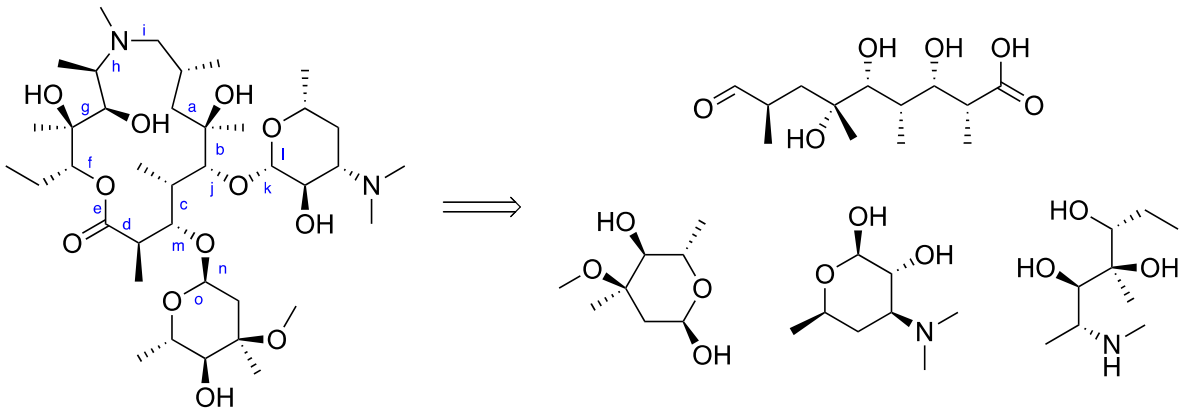


Figure 3-5. Another mapping problem from workshop A. Participants were asked to identify the bonds formed in the synthesis of azithromycin given the indicated starting materials. They were told that they should be able to propose a plausible reaction (or mechanism) for how those bonds were formed. Correct answer: ehkn.

Other activities in this workshop focused on identifying regiochemical and stereochemical features of a target molecule to propose associated reactions. For example, participants were shown that an inversion of configuration at a heteroatom-bearing stereocenter could be synthetically achieved using a S_N2 reaction, because these reactions proceed with inversion of configuration at a stereocenter. Due to time constraints, these skills were not formally assessed.

Workshop B: Retrosynthesis-based learning

Participants in Workshop B were given instruction that outlined the key features of a synthon-based retrosynthetic analysis. They were taught general strategies for making retrosynthetic disconnections that simplified the problem at hand and that were associated with known reactions. They were also taught how to create plausible synthons resulting from retrosynthetic disconnections. The strategies outlined in this workshop were based on Corey and Cheng's *The Logic of Chemical Synthesis.*[39] An example of a formative assessment task is provided in Figure 3-6, where participants were assessed on their ability to identify the most plausible synthon resulting from a given retrosynthetic disconnection. The electrons in the C–S bond were colour coded to ensure participants properly understood the conventions behind

synthon construction. Participants were only taught to construct “normal” polarity synthons; umpolung reactivity was not discussed.

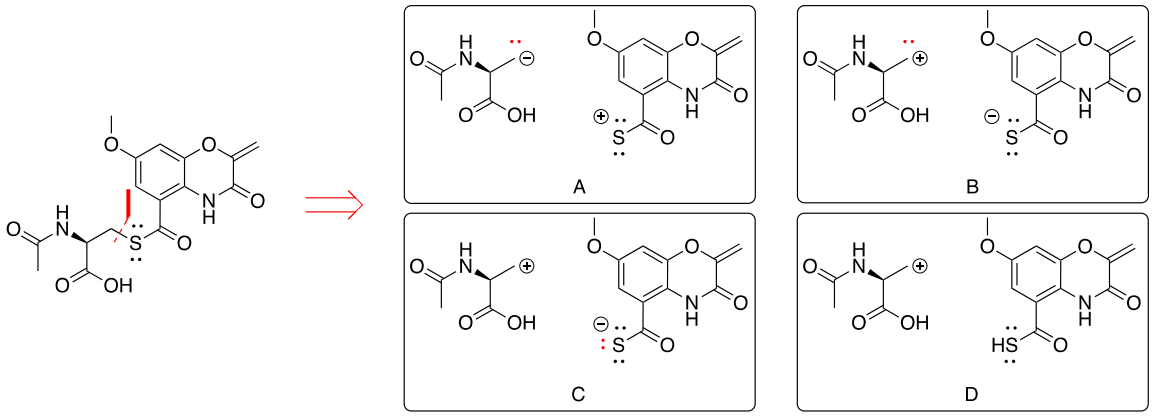


Figure 3-6. An example of an assessment task from Workshop B related to synthon construction. Participants were asked to “Identify the most plausible synthon pair that would result from disconnecting the C–S bond indicated in red.” Correct answer: C.

This problem allowed the facilitator to ensure participants could identify the conventions behind constructing synthons before moving onto a more challenging problem, where participants were asked to design a retrosynthetic analysis of azithromycin (described in the results and discussion). This problem was chosen to maximize alignment of this workshop with the mapping-based workshop. While participants in the mapping-based workshop were given starting materials and asked to identify sites of bond formation (Figure 3-5), participants in the retrosynthesis workshop were given the target and asked to disconnect bonds that would lead to plausible starting materials. After participants provided their proposals, the facilitator highlighted the bonds in azithromycin that would be suitable for disconnection, why that is the case (polar bonds associated with reactions that are familiar to the participants), and then highlighted bonds that would not be suitable for disconnection (not associated with familiar reactions).

After learning some fundamental rules behind the construction of synthons, participants were guided through retrosynthetic analyses that were designed based on “transform-goal” strategies, where retrosynthetic disconnects are made in the context of regiochemical patterns that allow the chemist to link those patterns to known reactions.[39] This procedure is analogous to the procedure participants in the mapping-based workshop learned for basing their choice of reactions on regiochemical patters in the target molecule. This analysis was discussed in the context of 1,1-disconnects and 1,2-disconnects, as participants

had not yet learned the reactions necessary to derive any meaningful information from identifying more distant disconnects. Participants had the opportunity to practice by evaluating three retrosynthetic analyses and deciding which followed the most plausible order of forward reactions.

The final portion of this workshop focused on identifying functional group transformations that simplify a retrosynthetic analysis. A sample problem is provided in Figure 3-7.

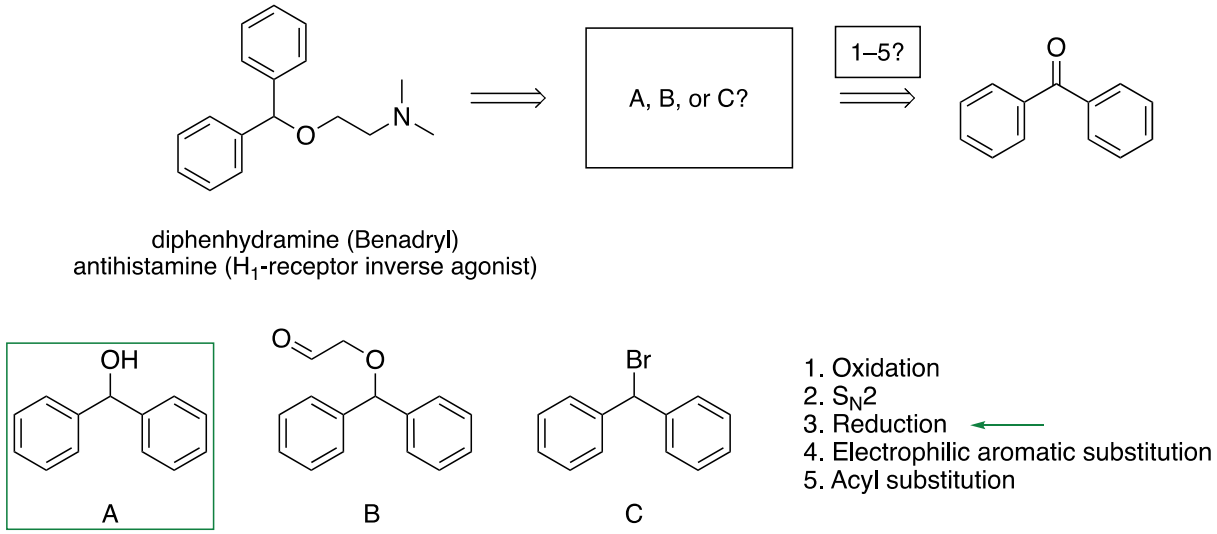


Figure 3-7. Sample practice problem for identifying functional group transformations in the retrosynthesis-based workshop. Correct answers indicated in green.

Workshop C: Reactions–based learning

The reactions-based workshop consisted entirely of clicker-based formative assessment tasks where participants completed different types of synthetic problems that focused on identifying either reagents or reactions that could be applied to the formation of an indicated bond. Participants in workshop C were not given any instruction related to strategies for problem-solving. An example is provided in Figure 3-8, where participants were shown an intermediate in a possible synthesis of sildenafil (Viagra) and asked to suggest reactions that could form indicated bonds. On the following slide (Figure 3-9), participants were given a synthetic scheme along with a bank of reagents and asked to select which reagents would accomplish each of the indicated synthetic transformations. This procedure is similar to how synthesis is commonly taught in textbooks and resembles what one might do when constructing a retrosynthetic analysis and using the information obtained from that analysis to propose a forward synthesis. However, we hypothesized that this learning approach would

not be as effective for helping students learn to approach novel synthetic problems as it does not place a strong emphasis on concrete problem-solving strategies that can provide clarity in unfamiliar problem situations; students would have to independently decide to use any strategies they thought might be appropriate on each problem.

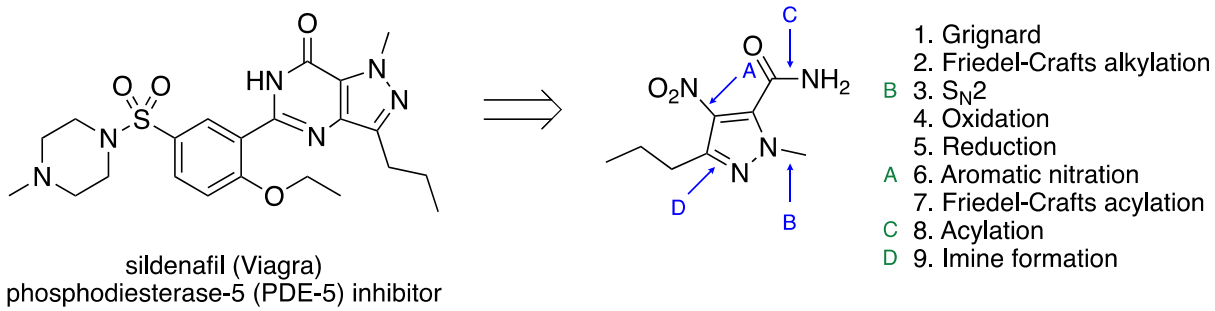


Figure 3-8. A clicker question from workshop C. Participants were given the following instructions: "Suggest reactions that could be used to form each of bonds A–D in this synthetic precursor to sildenafil. You may use the same reaction multiple times." Bond labels are matched to their respective reaction types in green.

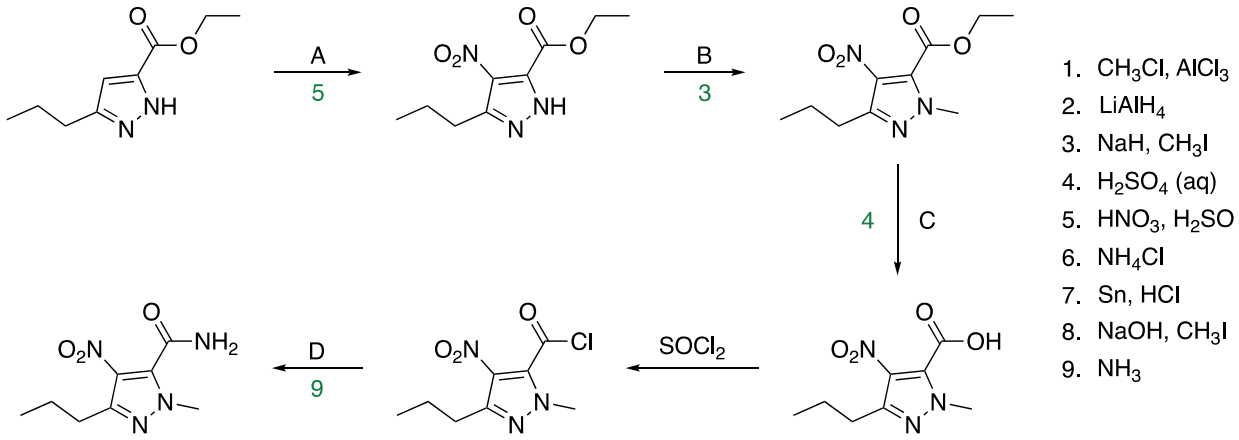


Figure 3-9. A clicker question from workshop C. Participants were given the following instructions: "Suggest reagents for each of reaction A-D. You may use the same reagents more than once." The fourth step was completed for participants because this reaction was still unfamiliar to them at the time of the workshop. Correct answers are indicated in green.

OC2 Final Exam and In-Class Learning Activities

Along with learning gains from the workshop, we wanted to determine if workshop participants were outperforming non-participants on learning activities that took place after the workshops, as well as the synthesis problem on the final exam (described fully in the data analysis section). We chose two learning activities from participants' OC2 class that focused explicitly on mapping (Figure 3-10). Students responded to these questions using Top Hat, a classroom response system. These questions were presented in two different classes that occurred roughly three weeks after the workshops.

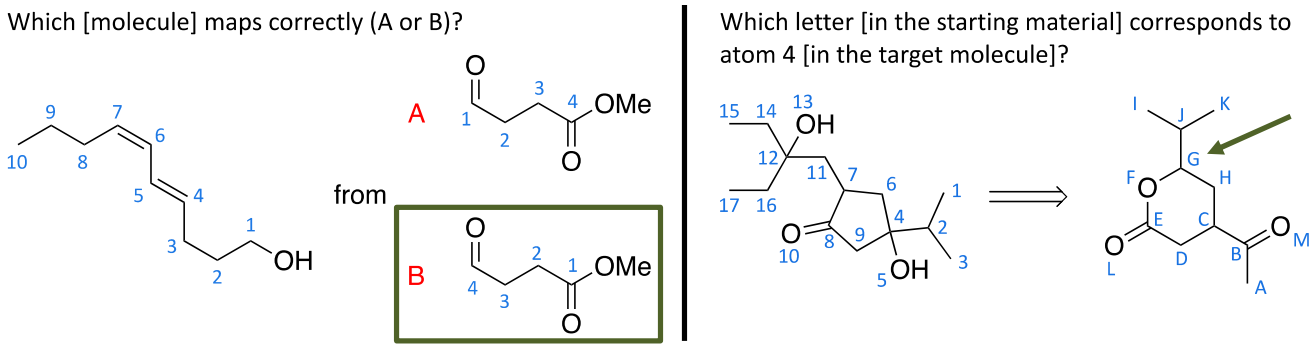


Figure 3-10. In-class learning activities used to assess students' mapping and synthesis analysis skills. Correct answers indicated in green.

Pre- and Post-tests

We used a pre-workshop/post-workshop test format to measure learning gains from the workshop. Each assessment (described fully in the Data Analysis section) consisted of a single synthesis problem, where participants were given starting materials and asked to synthesize a target molecule using any other reagents they deemed necessary. The problems were unfamiliar to the participants, but developed so that they could be solved using reactions that participants had been taught. Participants who did not complete both the pre-test and the post-test were excluded from the study.

Classroom Observation Protocol for Undergraduate STEM (COPUS)

We used the Classroom Observation Protocol for Undergraduate STEM (COPUS)[13] to measure the extent to which the workshops followed this group active learning format. There were three observers, one for each workshop, who were upper year or graduate students in chemistry; all had been previously trained in using the protocol. Using the protocol, observers describe the behaviours of the students and instructor. Each workshop had one observer, each of whom was trained how to apply the protocol to a flipped organic chemistry classroom by observing a session of the Organic Chemistry 2 course described herein (*vide supra*). A passive workshop or differences in workshop styles might have affected the results. The codes used in the COPUS protocol are outlined fully in Table 3-1.

Table 3-1. Description of the codes used in the COPUS observation protocol.[13]

Students are doing:	
L	Listening to instructor, taking notes, etc.
Ind	Individual thinking/problem solving. Only mark when an instructor explicitly asks students to think about a clicker question or another question/problem on their own.
CG	Discuss clicker question in groups or more students
WG	Working in groups on worksheet activity
OG	Other assigned group activity, such as responding to instructor question
AnQ	Student answering a question posed by the instructor with rest of class listening
SQ	Student asks question
WC	Engaged in whole class discussion by offering explanations, opinion, judgment, etc. to whole class, often facilitated by instructor
Prd	Making a prediction about the outcome of demo or experiment
SP	Presentation by student(s)
TQ	Test or quiz
W	Waiting (instructor late, working on fixing AV problems, instructor otherwise occupied, etc.)
O	Other – explain in comments
Instructor is doing:	
Lec	Lecturing (presenting content, deriving mathematical results, presenting a problem solution, etc.)
RtW	Real-time writing on board, document projector, etc. (often checked off along with Lec)
FUp	Follow-up/feedback on clicker question or activity to entire class
PQ	Posing non-clicker question to students (non-rhetorical)
CQ	Asking a clicker question (mark the entire time the instructor is using a clicker question, not just when first asked)
MG	Moving through class guiding ongoing student work during active learning task
1o1	One-on-one extended discussion with one or a few individuals, not paying attention to the rest of the class (can be along with MG or AnQ)
D/V	Showing or conducting a demo, experiment, simulation, video, or animation
Adm	Administration (assign homework, return tests, etc.)
W	Waiting when there is an opportunity to be interacting with or observing/listening to student or group activities and the instructor is not doing so
O	Other – explain in comments

Attitudes toward the Subject of Chemistry Inventory, version 2 (ASCIv2)

The ASCIv2 instrument addresses components of the affective domain of meaningful learning (*i.e.*, intellectual accessibility and emotional satisfaction) and allowed us to measure any changes in participants' attitudes toward chemistry following the workshop. In this inventory, attitude is defined as "a learned predisposition to respond favourably or unfavourably toward an attitude object."[40] The structure and items in the ASCIv2 are provided in Table 3-2.

Table 3-2. Subcategories of the ASCIv2 (with related domains of meaningful learning) and items pertaining to each subcategory. Items are phrased as, for example, "Chemistry is easy/hard", where 1 = hard and 7 = easy. When we implemented the ASCIv2, we made an oversight in one of the items: we used the item "Chemistry is unsatisfying/satisfying", while the ASCIv1 and v2 used the item "Chemistry is frustrating/satisfying".

	Intellectual accessibility (cognitive domain)	Emotional satisfaction (affective domain)
Chemistry is…	Hard/Easy Complicated/Simple Confusing/Clear Challenging/Unchallenging	Uncomfortable/Comfortable Unsatisfying/Satisfying Unpleasant/Pleasant Chaotic/organized

Metacognitive Activities Inventory (MCAI)

We used the MCAI to measure participants' metacognitive skillfulness. [14] The MCAI is a 5-point Likert-scale inventory that asks participants to reflect on the activities and actions they use to regulate their thought processes during problem-solving activities. The full MCAI is provided in Table 3-3. We anticipated that participants would be higher achievers in the course and have high metacognitive skill. We intended to explore whether the workshop would *only* be beneficial for students who were highly metacognitively skilled (RQ2).

Table 3-3. The Metacognitive Activities Inventory.[14] The left–hand column indicates whether the activity described in a given statement has a negative (–) or positive (+) influence on problem solving.

Coding (+/–)	Statement
+	I read the statement of a problem carefully to fully understand it and determine what the goal is.
+	When I do assigned problems, I try to learn more about the concepts so that I can apply this knowledge to test problems.
+	I sort the information in the statement and determine what is relevant.
+	Once a result is obtained, I check to see that it agrees with what I expected.
+	I try to relate unfamiliar problems with previous situations or problems solved.
+	I try to determine the form in which the answer or product will be expressed.
+	If a problem involves several calculations, I make those calculations separately and check the intermediate results.
+	I clearly identify the goal of a problem (unknown variable to solve for or the concept to be defined) before attempting a solution.
+	I consider what information needed might not be given in the statement of the problem.
+	I try to double-check everything: my understanding of the problem, calculations, units, etc.
+	I use graphic organizers (diagrams, flow-charts, etc.) to better understand problems.
+	I experience moments of insight or creativity while solving problems.
+	I jot down things I know that might help me solve a problem before attempting a solution.
+	I find important relations among the quantities, factors, or concepts involved before trying a solution.
+	I make sure that my solution actually answers the question.
+	I plan how to solve a problem before I actually start solving it (even if it is a brief mental plan).
+	I reflect upon things I know that are relevant to a problem.
+	I analyze the steps of my plan and the appropriateness of each step.
+	I attempt to break down the problem to find the starting point.

- I spend little time on problems for which I do not already have a set of solving rules or that I have not been taught before.
- When I solve problems, I omit thinking of concepts before attempting a solution.
- Once I know how to solve a type of problem, I put no more time in understanding the concepts involved.
- I do not check that the answer makes sense.
- If I do not know exactly how to solve a problem, I immediately try to guess the answer.
- I start solving problems without having to read all the details of the statement.
- I spend little time on problems I am not sure I can solve.
- When practicing, if a problem takes several attempts and I cannot get it right, I get someone to do it for me and I try to memorize the procedure.

Data analysis

Assessment of responses to synthesis problems

To facilitate our analysis of participants' responses to the synthesis problems on the pre- and post-test, we broke each synthesis problem's solution down into three key steps that would be required to solve the problem. We scored participants' responses to the pre-test and post-test questions out of six points, following the coding scheme summarized in Figure 3-11, which includes generalized pathways that would successfully answer each problem. We did not expect participants to be familiar with more advanced/novel synthesis approaches to solving these problems. Each solution that differed from the ones provided therein was assessed for its own merit, and points were still awarded where appropriate.

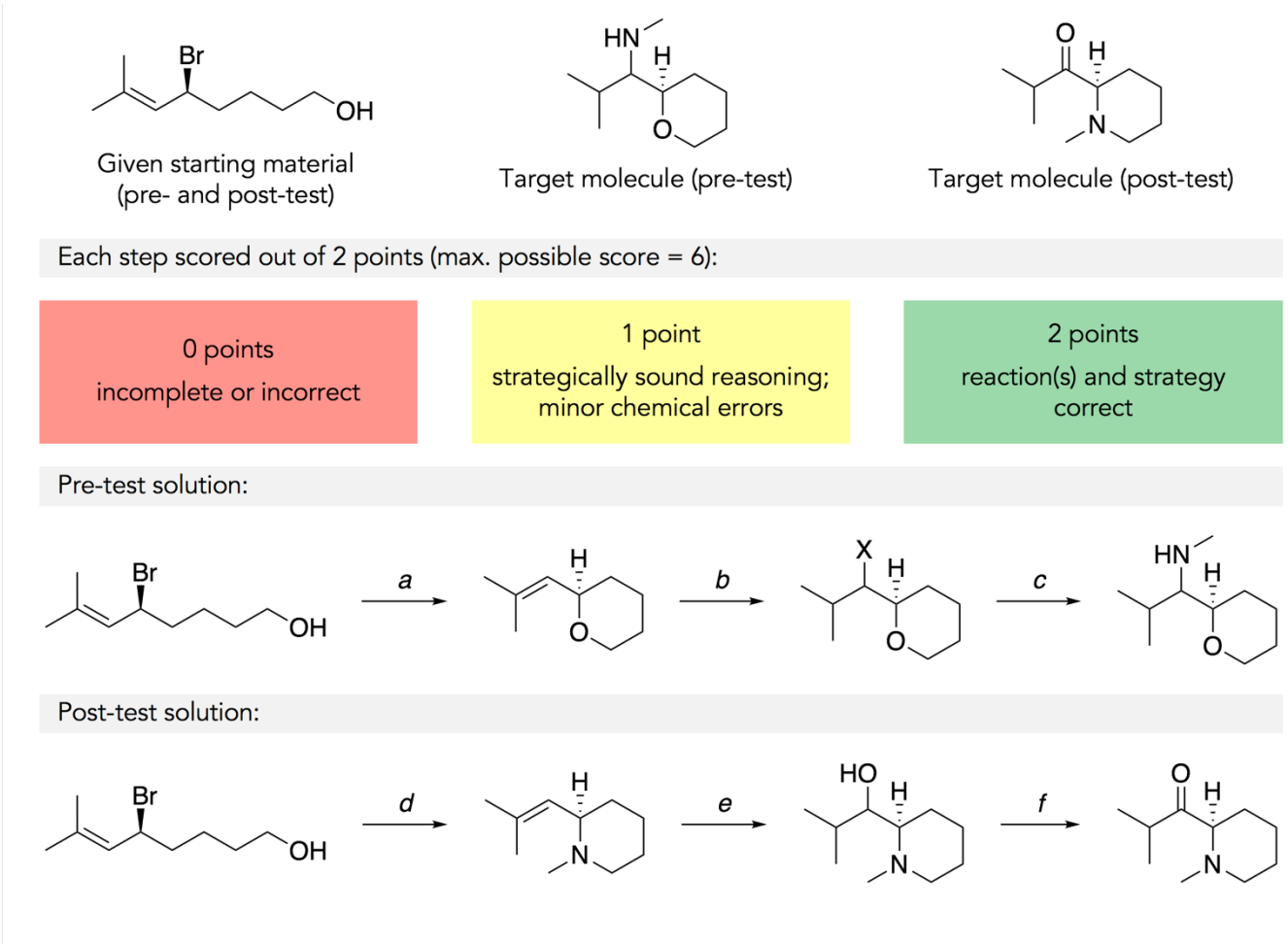


Figure 3-11. An overview of the assessment tasks used in the workshops. Pre-test solution (others were possible): a = NaOH; b = N-bromosuccinimide, H_2O_2 (X = Br); c = $MeNH_2$, Et_3N. Post-test solution (others were possible): a = 1. TsCl 2. CH_3NH_2, Et_3N; e = 1. BH_3 2. H_2O_2, NaOH; f = pyridinium chlorochromate (PCC).

We evaluated the equivalency of these tasks (content validity) by administering each problem on two separate versions of the Organic Chemistry 2 final exam in fall 2017 and found no significant difference in the degree of success on either problem. Full details of this evaluation are provided in the results and discussion.

We also wished to investigate for the presence of any delayed learning gains experienced by workshop participants; to this end, we analyzed participants' responses to the synthesis problem on the final exam for their OC2 course, then compared their performance to an equal number of responses by OC2 participants who did not participate in a workshop. Non-workshop participants were matched to workshop participants based on final exam score; every non-workshop participant who was matched to a workshop participant had a final exam score

within at least 2.5% of their match. The average exam score of workshop participants was the same as the matched group (71%, s.d. = 16%), and the mean difference in exam score between matched students was 1.2% (s.d. = 0.8%).

The synthesis problem on the final exam for participants' OC2 course can be found in Figure 3-12; answers were coded in the same fashion as the workshop assessment tasks. The definitions of each step that we used when coding participants' responses are in brackets. For an example, a common solution (Figure 3-12) provided access to **1** through a Wittig reaction, where 3-bromopentane (**2**) was converted to pentan-3-one (**5**), and 1-bromobutane (**3**) was converted to phosphonium ylide **7** to enable the Wittig reaction that provides access to **1**.

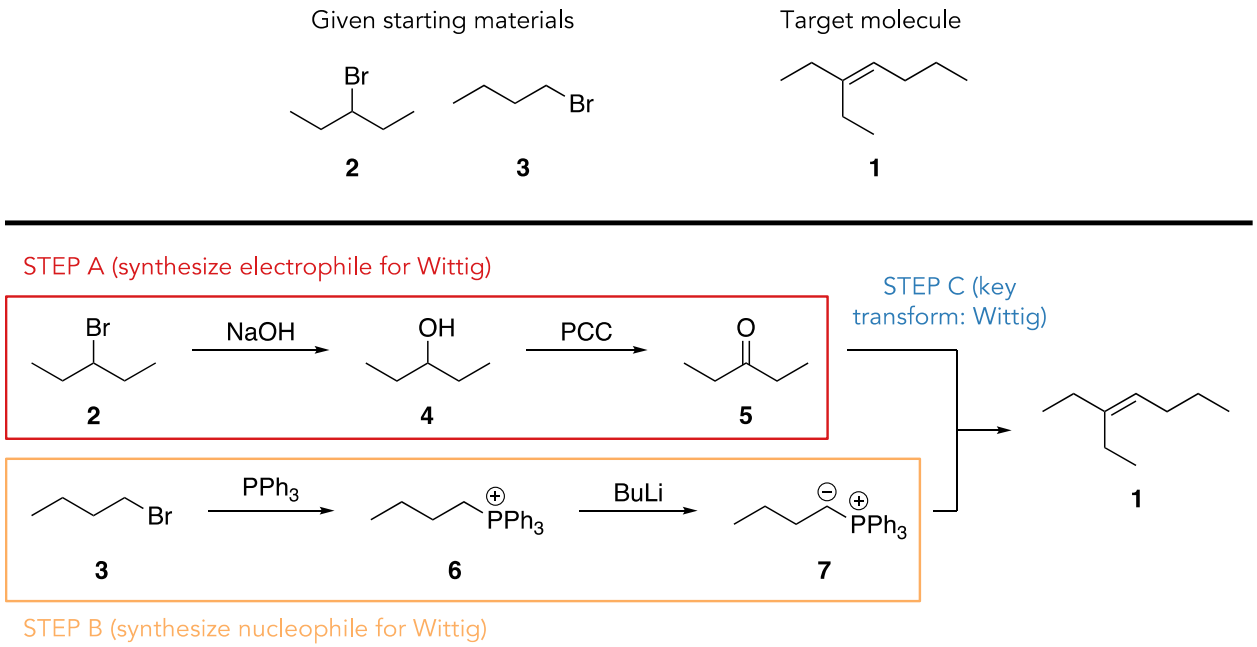


Figure 3-12. Key steps required for completing the synthesis problem on participants' OC2 final exam. One sample solution is shown; alternate answers were accepted. PCC = pyridinium chlorochromate. BuLi = n-butyllithium.

Analysis of MCAI responses

The MCAI data were normally distributed as indicated by a Shapiro-Wilk test, $W(43) = 0.977$, $p = .550$. Levene's test was non-significant ($p = .368$), indicating homogeneity of variance. Therefore, we analyzed these data using one-way analysis of variance (ANOVA) to compare MCAI scores between workshops and *t*-tests to compare MCAI scores between participants who provided successful and unsuccessful solutions to the post-test.

We used non-parametric tests (Mann-Whitney *U*) to analyze responses to individual MCAI items. Although participants' overall MCAI scores were normally distributed, responses to the individual items that we discuss in more detail were not; they had significant Shapiro-Wilk test results ($p < .001$). Also, individual responses to Likert items are considered to be ordinal rather than interval data, because while the responses have a rank order, the gaps between response categories cannot be presumed to be equal.[41] As such, it is inappropriate to use parametric tests to analyze these data.

Results and Discussion

Evaluating content validity of the pre-test and post-test questions

We carried out content validity evaluation by using two different versions of a synthesis question in the Organic Chemistry 2 final exam in the fall 2017 semester at the University of Ottawa. Version A contained the workshop pre-test question, and Version B contained the post-test question. We coded a subset of responses to each of these questions using the coding scheme outlined in Figure 3-11. This subset of responses was selected by taking the total final exam score of each student, and matching their score (± 2.5%) to a workshop participant from the fall 2016 OC2 cohort. This process helped ensure the responses we chose came from students who had comparable abilities in organic chemistry to the workshop participants. We were comparing workshop vs. non-workshop participants' synthesis problem-solving skills on the final exam, and therefore felt it best to choose the non-workshop participants' final exams to code based on overall score on the assessment that contained synthesis questions (earlier assessments did not).

In the subset of the fall 2017 cohort who wrote version A of the final exam, a suitable match was found for 37 workshop participants, and for those who wrote version B, a suitable match was found for 39 participants. We first compared the distribution of scores on each question. These score distributions deviated significantly from normal as indicated by the Shapiro-Wilk test, $W_{version\ A}(37) = 0.913$, $p = .007$; $W_{version\ B}(39) = 0.878$, $p = .001$; therefore, we chose non-parametric tests to analyze these data with greater statistical power.

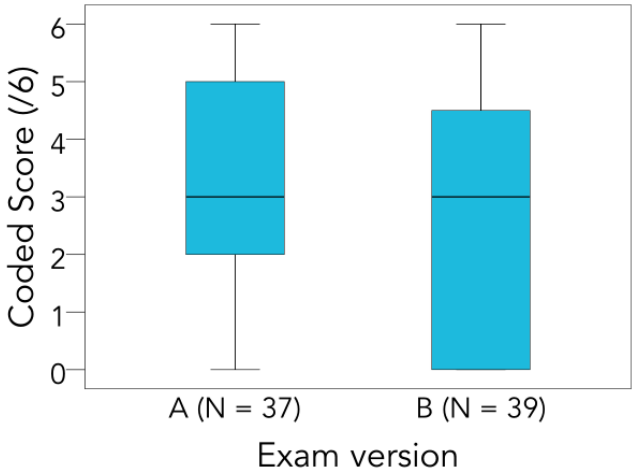
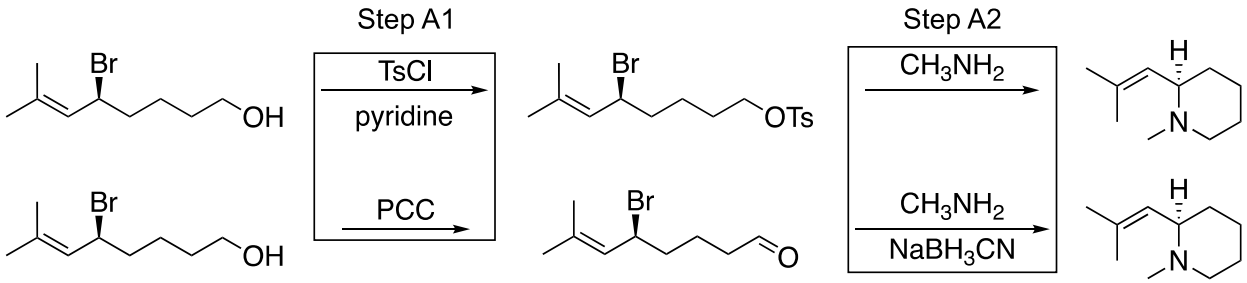


Figure 3-13. Distribution of coded scores of matched students on each version of the 2017 OC2 final exam. Exam version A = workshop pre-test; B = post-test.

We found no significant difference in the distribution of scores on these problems, $U(76) = 642.0$, $z = –0.837$, $p = .203$, with a small effect size, $r = –0.10$. The distribution of these scores is displayed in Figure 3-13; in each case, the median score was 3/6. This evidence supports that the problems were reasonably equal in difficulty.

Analyzing participants' success rates in a stepwise fashion

To further explore differences in difficulty between the pre- and post-test, we looked at their relative rates of success on each individual step of the pre-test and post-test problem, using the steps outlined in Figure 3-11. Since step A of the post-test required at minimum two different reactions to carry out (Scheme 3-1), we looked at whether participants struggled more with either of these "substeps" A1 and A2 as well. Figure 3-14 illustrates the rates of success of each step of the pre-test and post-test, with discussion of each test in the following sections.



Scheme 3-1. Two expected solutions to Step A of the post-test. PCC = pyridinium chlorochromate.

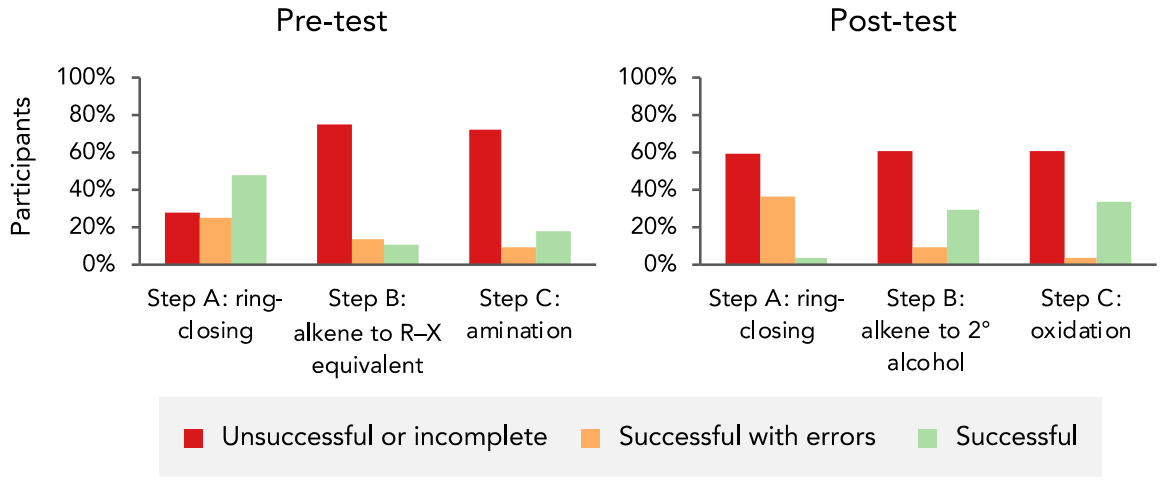


Figure 3-14. Success rates on each individual step of the pre-test and post-test. The steps are as described in Figure 3-11. N = 44 for both the pre- and post-test.

Stepwise analysis of responses to the pre-test problem

Following the application of the coding scheme in, we used chi-square analysis to investigate any differences in success rates on *individual steps* of the pre-test problem. The significant result of this analysis, $\chi^2(4) = 27.41$, $p < .001$, with a medium effect size, Cramèr's $V = 0.322$, indicates that participants experienced more difficulty with certain synthetic steps than others.

Step A of the pre-test seemed to be the easiest for participants; only 27% (N = 12) of participants were unable to propose a set of conditions for the synthesis of the tetrahydrofuranyl portion of the product (Step A; Figure 3-11). The remainder of participants recognized the need for a ring-closing S_N2 reaction. There were 48% (N = 21) of participants who were successful in proposing a plausible set of conditions for this reaction (typically some form of base catalysis); the remaining 25% (N = 11) of participants recognized the need for a ring-closing S_N2 reaction, but proposed a set of conditions with errors (either failing to include a base, suggesting a base that would be unsuitable, or drawing an implausible reaction mechanism).

The majority of errors were observed in steps B and C. Step B required participants to convert the trisubstituted alkene in the given starting material to either a secondary alkyl halide or some other R–X equivalent via an anti-Markovnikov type addition. Many participants (56%, N = 26) did not provide conditions for this step at all; they either did not propose a complete synthesis (11% of participants, N = 5), or attempted to aminate the alkene directly (step C) through implausible mechanisms (48%, N = 21).

Stepwise analysis of responses to the post-test problem

Stepwise analysis of participants' responses to the post-test problem showed a significant difference in the steps for which participants were able to successfully propose a set of conditions that would accomplish the required synthetic transformation without errors in chemistry, $\chi^2(4) = 25.47$, $p < .001$, with a medium effect size, Cramèr's $V = 0.311$. However, this analysis was only significant when it distinguished between participants who were successful with/without errors. When participants were treated simply as successful or unsuccessful on a given step, no significant difference was observed, $\chi^2(2) = 0.063$, $p < .969$, with a small effect size, $\phi = 0.022$. Therefore, it seems that many of these participants could successfully identify the synthetic transformations that were required, and struggled mainly with coming up with the correct reagents for those synthetic transformations.

Participants struggled the most with step A of the post-test—synthesis of the piperidinyl moiety in the target molecule (Figure 3-15). Chi-square analysis showed a significant difference in the degree to which participants were able to successfully complete step A of the pre-test compared to step A of the post-test, $\chi^2(2) = 21.78$, $p < 0.001$, with a large effect size, $\phi = 0.50$. Compared with 27% (N = 12) of participants who were unable to propose a synthesis of the tetrahydropyranyl ring portion of the pre-test target molecule, 59% (N = 26) were unable to synthesize the piperidinyl ring portion of the post-test target molecule, and only two participants were able to complete this step without any errors in chemistry. Furthermore, we observed that the majority of errors made by successful problem solvers were made on this step. Therefore, we chose to look at this step more closely.

Post-test Step A

Given the starting materials, synthesizing the piperidinyl moiety (post-test) required at least one extra step compared to the tetrahydropyranyl moiety (pre-test; Figure 3-11), so we investigated if this extra step introduced a significant barrier to success. Our analysis suggests that participants did not struggle with one step of the piperidinyl moiety (post-test step A) synthesis more than the other; chi-square analysis indicated that the difference between participants' success on steps A1 and A2 was not significant, $\chi^2(2) = 5.555$, $p = .073$, with a small effect size, $\phi = 0.251$. The result of this analysis approaches a significant result, but based on the data presented in Figure 3-15, this is due to the distinction made between participants who were successful with/without errors. When "successful" and "successful with errors" are treated as equivalent, chi-square analysis indicated that the difference between participants'

success on steps A1 and A2 was not significant, $\chi^2(1) = 0.046$, $p = .831$, with a much smaller small effect size, $\phi = 0.023$.

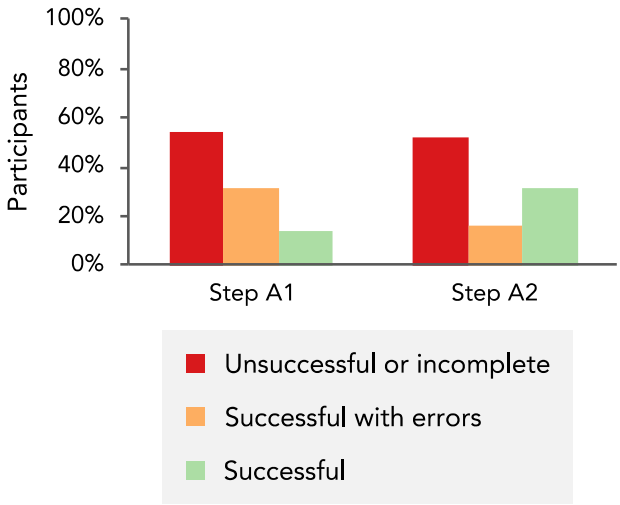


Figure 3-15. Success rates on each sub-step of post-test step A (as described in Figure 3-11; examples in Scheme 3-1). N = 44.

This observed discrepancy between successful with and without errors on step A1 is primarily due to participants choosing to protonate the alcohol in the starting material, an unsuitable strategy for a substitution reaction with an amine. However, this indicates that participants knew what *general strategy* to use (convert the alcohol to a leaving group and substitute with methylamine), and their errors were in choice of reagents.

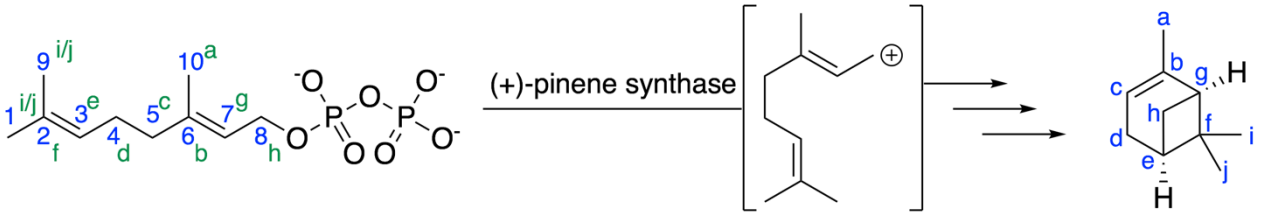
Based on this in-depth analysis into the relationships between participants' responses to the pre-test and post-test and the content validity investigation of these items we performed using the 2017 OC2 cohort's final exam (*vide supra*), the lack of observed learning gains does not seem to be attributable to any difference in experienced difficulty between the pre-test and post-test.

Workshop: Formative assessment showed skill using strategies

Throughout workshops A (mapping) and B (retrosynthesis), the formative assessment opportunities showed that the majority could provide correct/acceptable answers to these questions, suggesting that students were able to quickly learn the conventions of the key problem-solving strategies addressed in the workshops.

An early mapping activity in Workshop A is presented in Figure 3-16; although few participants chose to respond to this item (participation was optional), we saw that some students were able to map this relatively challenging situation without any further information. We chose a more challenging problem because participants had the opportunity to work in groups and seek the assistance of the expert facilitator; therefore, we expected they would be able to solve problems beyond the scope of what they would be able to do on their own.

Map the atoms between the starting material and product.



geranyl pyrophosphate

(+)-α-pinene

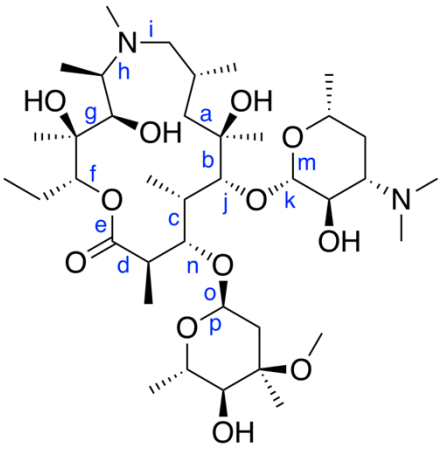
Participant	Response	
1	jfedcbghia	✓
2	ifedcbghja	✓
3	ifedcbghja	✓
4	jhefgbcdja	✗
5	adghghbhhg	✗

Figure 3-16. A formative assessment task from the mapping workshop, where participants were asked to map the atoms between the two indicated molecules. Checkmarks and cross-marks indicate if a given response was correct or not. The green letters (not provided to participants) indicate the correct answers; students were asked to input their answer as the letters corresponding to the numbers in order from 1-10.

The results of a learning activity from Workshop B are presented in Figure 3-17; participants were shown the macrolide antibiotic azithromycin and instructed to propose a plausible set of bonds to disconnect. They were instructed that an ideal response would involve 4–5 disconnections resulting in fragments that were at most ~200 g/mol in size, but due to the complexity of the task, any suitable polar bond disconnections were accepted. Of the 14 participants in this workshop who provided a response to this question, 10 provided a reasonable set of bonds to disconnect. Responses were considered reasonable if the bonds suggested for disconnection were associated with a transform that students were familiar with (*e.g.*, bond *e* via acyl substitution, bond *h* via epoxide opening). Due to the high degree of

complexity in the molecule used for this example, the high number of reasonable responses was very encouraging.

Propose a plausible set of bonds to disconnect that would produce reasonable starting materials.



Participant		Response	Participant		Response
1	✓	e i k o	8	✓	n j f h
2	✓	j/k, n/o, f, h/i	9	✓	e h k o
3	✓	e k o	10	✓	f i j n
4	✓	e j n	11	✗	d e f
5	✓	i j n f	12	✗	j n f h a
6	✓	n j	13	✗	d e f g
7	✓	n j f h	14	✗	a e m p

Accepted answers: e or f, h or i, j or k, n or o

Figure 3-17. A formative assessment task from the retrosynthesis workshop, where participants were asked to suggest which bonds in azithromycin could be suitably disconnected. Each response we received is shown organized by participant, edited for clarity but not for content. Checkmarks and cross-marks indicate if a given response was considered reasonable or not.

Success on Synthesis Problems

We found no significant differences in the proportion of successful solutions following the intervention, nor did we observe any difference in proportion of successful solutions between workshop groups (Figure 3-18 and Figure 3-19). Two-way repeated measures ANOVA showed a non-significant main effect of time, $\Lambda = 0.998$, F (1, 41) = 0.081, p = .777, with a very small effect size, partial η^2 = 0.002, as well as a non-significant interaction effect, $\Lambda = 0.976$, F (2, 41) = 0.495, p = .613, with a very small effect size, partial η^2 = 0.024. The following sections contain analyses of the strategies participants used to solve these problems. A finer-grained analysis of participants' responses to these problems can be found in our content validity evaluation (*vide supra*), in which each problem is broken down by synthetic step.

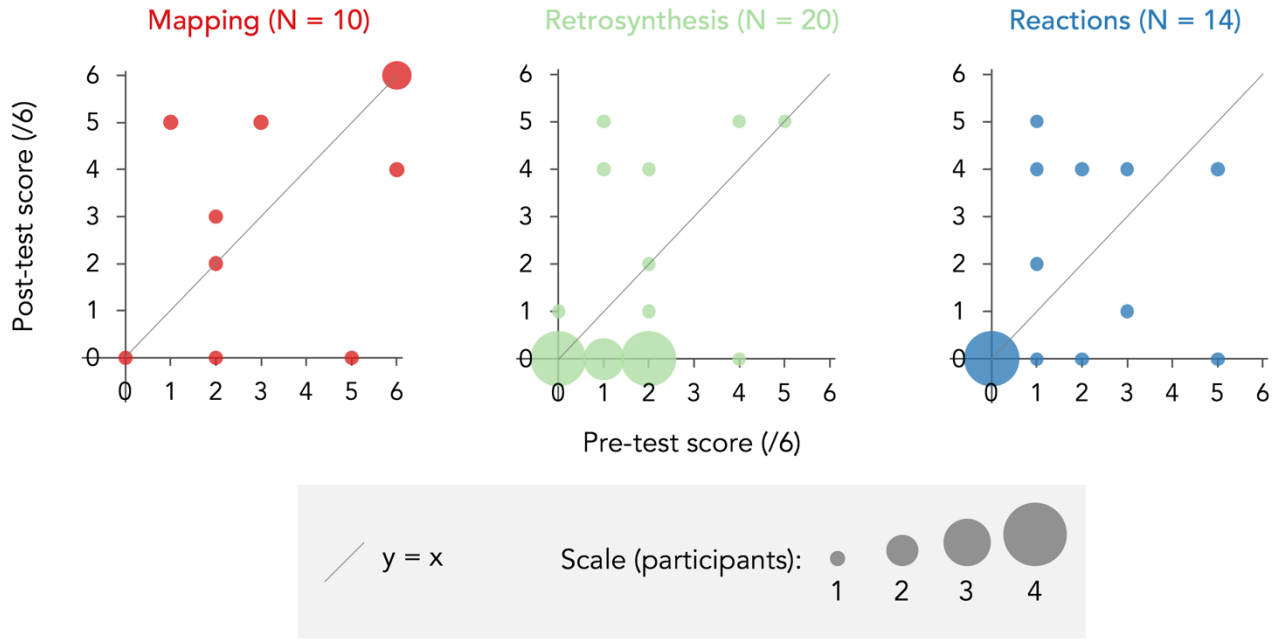


Figure 3-18. Participants' post-test scores as a function of their pre-test scores. Participants above the y = x lines appeared to improve following their respective workshop, while those below the line appeared to regress.

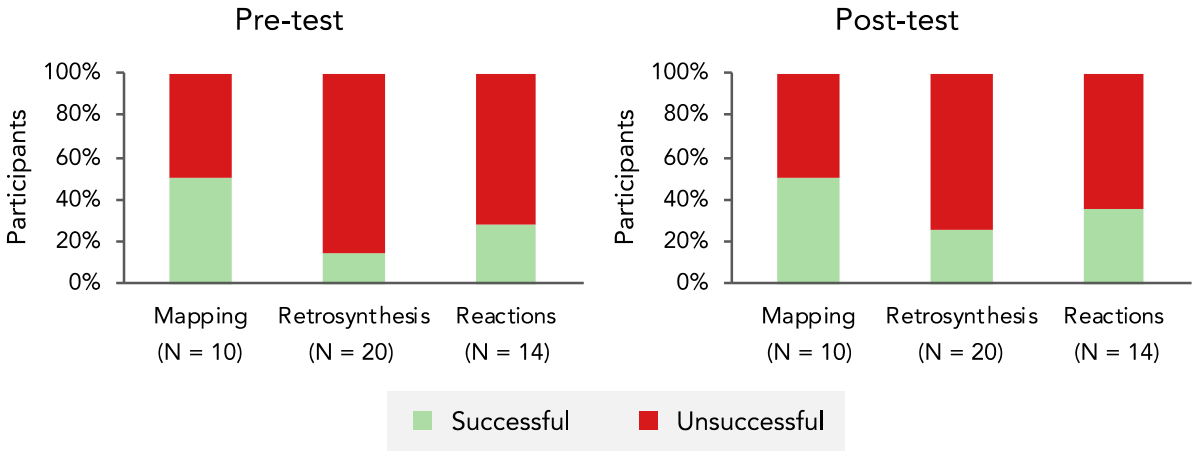


Figure 3-19. Proportions of successful and unsuccessful solutions on the pre-workshop and post-workshop synthesis problems.

Pre-test: High rates of mapping and mechanism use, but not success

The key problem-solving strategies for which we coded participants' pre-test responses were frequently found to be used correctly in successful solutions, and incorrectly in

unsuccessful solutions; this is consistent with our prior analysis of students' responses to synthesis problems.[8]

Reaction mechanisms and mapping were the most commonly used strategies

Reaction mechanisms were the most commonly used problem-solving tool on the pre-test (93% of participants). We observed a significant association between the use of reaction mechanisms within successful solutions to the pre-test, $\chi^2(2) = 26.65$, $p < .001$, with a large effect size, $\phi = 0.778$ (Figure 3-20); 82% of participants with unsuccessful solutions who used mechanisms on the pre-test did so incorrectly.

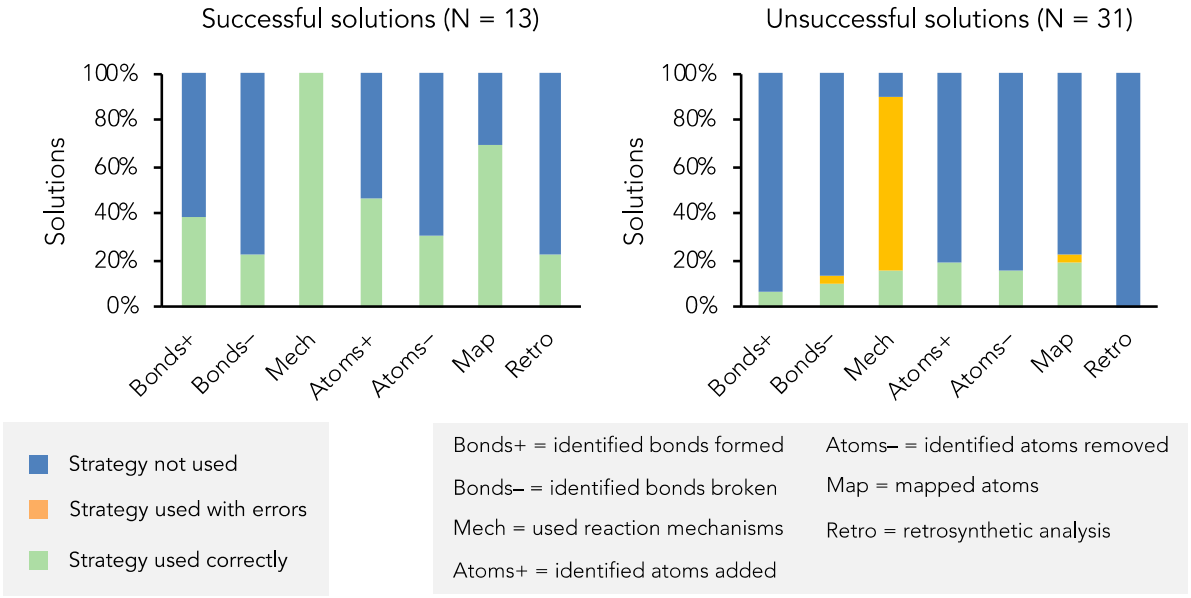


Figure 3-20. Key problem-solving strategies found in responses to the synthesis problem on the pre-test.

A possible reason for the significant association between the incorrect use of reaction mechanisms and unsuccessful solutions to the pre-test is that the participants who provided these unsuccessful solutions were using mechanisms in a non-meaningful fashion. The use of mechanisms would be considered meaningful when participants proposed plausible mechanisms in their synthesis proposals or if it was clear that they used the electron-pushing formalism to reinforce their mental models about the reaction outcome they were trying to rationalize (*i.e.*, check their work). We did not investigate participants' mental models explicitly but hypothesize that using mechanisms to double-check the plausibility of a proposed synthesis would be a common reason for using the electron-pushing formalism.

Conversely, the use of mechanisms was considered non-meaningful when there did not appear to be any logical approach for accessing the desired target molecule and instead drew arrows between starting materials that would get them to the product without any regard for the plausibility of the associated mechanism. This non-meaningful approach to proposing reaction mechanisms is well-documented;[42-44] to illustrate the difference between the correct and incorrect use of mechanisms in this sample, an example of a participant's response to the post-test problem is provided in Figure 3-21.

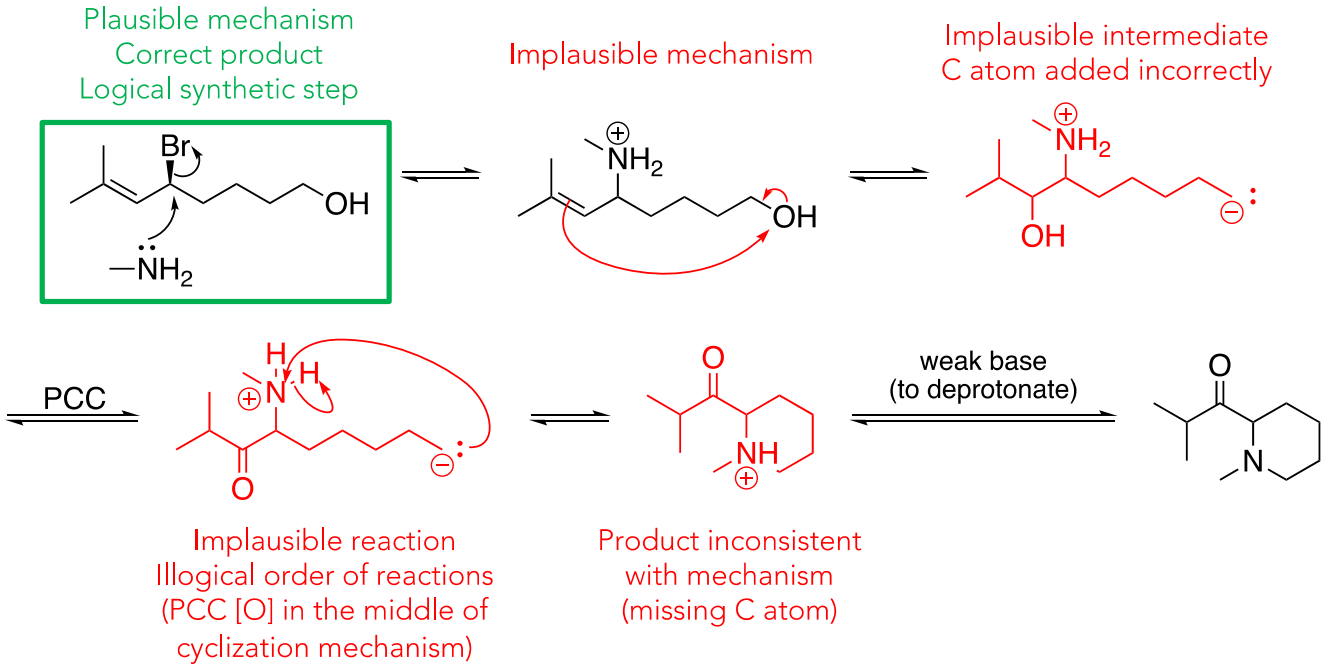


Figure 3-21. A participant's proposed solution to the post-test problem (edited for clarity). We have annotated their solution to indicate correct and incorrect usage of the electron-pushing formalism.

Mapping was the only other strategy for which we observed a significant association with successful problem solving, $\chi^2(2) = 10.24$, $p = .006$, with a medium effect size, $\phi = 0.482$. This result helps to support previous work that suggested the importance of this strategy as a determinant of success in synthesis problem solving. No significant association was observed between successful problem solving and the other strategies for which we coded participants' pre-test responses (*i.e.*, identifying bonds formed and broken, identifying atoms added and removed, retrosynthetic analysis).

Post-test: No increase in success rates, large decrease in use of strategies

There was no overall improvement in success rates by participants between the pre-test and post-test in any workshop. Although the relative difficulty of the pre-test and post-test does not appear to have influenced the results of this study, we observed a large decrease in evidence for problem-solving strategies on the post-test, which we believe may have been a mitigating factor. In the case of mapping, identifying atoms added/removed, identifying bonds formed/broken, and retrosynthetic analysis, greater than 65% of participants overall showed no evidence of using these strategies (the majority of participants who used a retrosynthetic analysis, correctly or with errors, were in workshop B, which focused explicitly on this strategy). Figure 3-22 illustrates how strategies were used on the post-test.

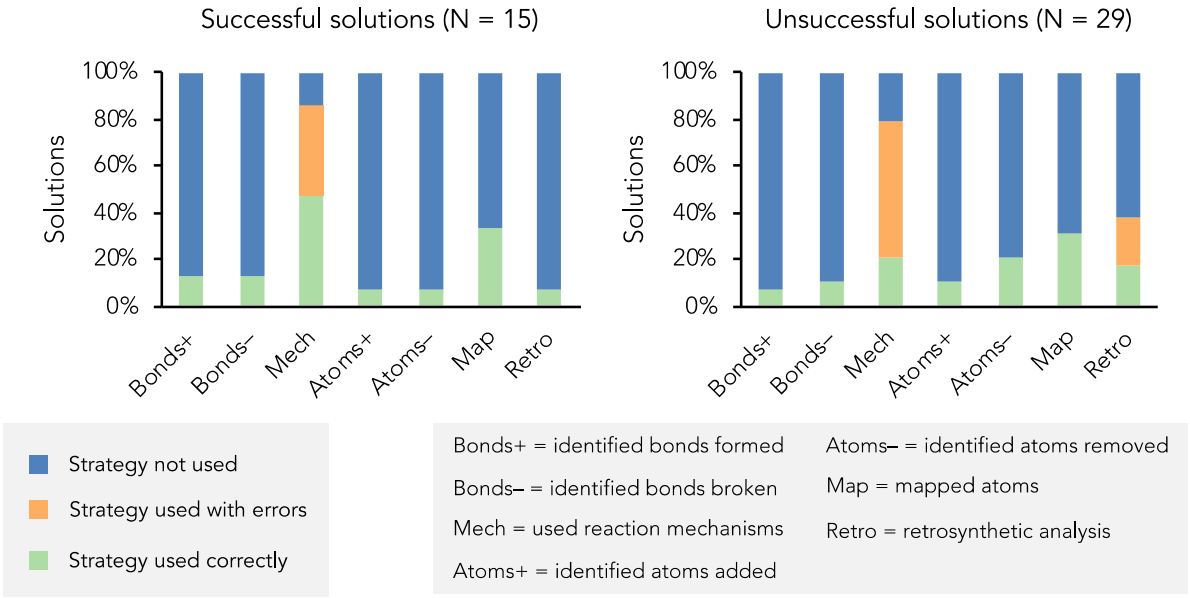


Figure 3-22. Key problem-solving strategies found in responses to the synthesis problem on the post-test.

When looking at the use of these strategies in individual workshops, it was often the case that very few (< 5) or no participants in a given workshop used certain strategies. Reaction mechanisms were still the most commonly used tool, possibly due to the strong emphasis on mechanisms in their course, but we observed no significant association between the use of reaction mechanisms and successful problem solving, $\chi^2(2) = 3.21$, $p = .228$, with a small effect size, $\phi = 0.270$.

In light of these findings, the interventions may have lacked any significant impact on synthesis problem-solving skills, but the participants may have also experienced more difficulty

with the post-test due to a less effortful analysis of the problem, perhaps due to fatigue. Participants might have been using these strategies mentally rather than writing down representations, but our previous findings have shown that successful solutions generally have several strategies committed to paper.[8] In addition, writing out strategic analysis of a synthesis problem was not an explicit learning outcome for the workshops; however, we expected that the learning activities in the workshop would model the value of explicitly writing out the analysis. As such, greater emphasis on the importance of applying these strategies explicitly on paper is needed. A single intervention focusing on problem-solving strategies is unlikely to be sufficient, due to the inherent complexity of synthesis problem-solving; future work could include instruction on each of the strategies, then on using them together.

Retrosynthetic analysis amongst Workshop B participants

We used chi-square analysis to measure any association between the use of retrosynthetic analysis on the post-test and successful problem solving. While retrosynthetic analysis was not a commonly used strategy, we did not expect it to be for participants in the workshops that did not explicitly focus on its use; as such, this analysis was only performed for Workshop B (retrosynthetic analysis) participants because we were primarily interested in determining whether this specific intervention was successful in teaching participants how to use retrosynthetic analysis to solve synthesis problems of appropriate difficulty. No significant association was observed between the correct use of a retrosynthetic analysis and success in solving the post-test synthesis problem, $\chi^2(2) = 2.42$, $p = .326$, with a medium effect size, $\phi = 0.348$ (Figure 3-23). This result was not surprising, as participants had no prior experience with this expert-like problem solving strategy.

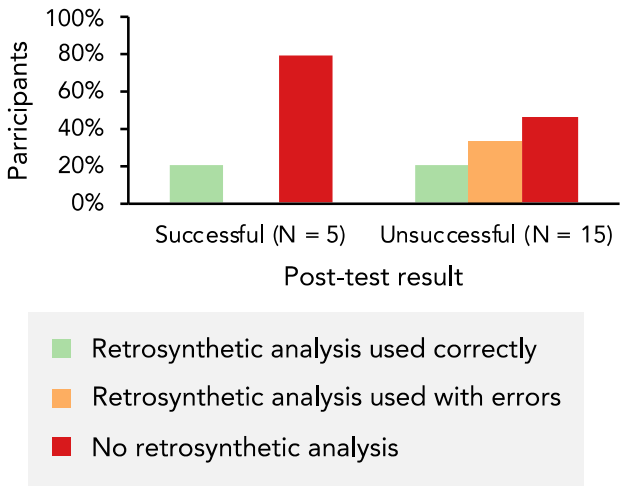


Figure 3-23. Frequency of the use of retrosynthetic analysis on the post-test by workshop B (retrosynthesis workshop) participants (N = 20). These data show that majority of participants did not use retrosynthetic analysis in their approach.

Only four out of 20 participants were able to use the conventions of a retrosynthetic analysis correctly on the post-test (*i.e.*, the transforms they proposed or disconnections they made were plausible), and only one of those four participants proposed a plausible synthesis of the target molecule from the given starting material. Since our analysis of participants' responses to formative assessment tasks during the workshop indicated that they were generally skilled with the conventions of synthon–based retrosynthetic analysis, and with making polar bond disconnections (Figure 3-17), these results suggest may need more than a single intervention before they are capable of using this strategy to solve full synthesis problems.

Class: Many students demonstrated competence with strategies, but some may have struggled to apply them

As we saw in the workshop (Figure 3-17), students were generally proficient at mapping relatively complex molecules in class (Figure 3-24). While non-workshop participants were not significantly less successful than workshop participants on this question, $\chi^2(1) = 1.545$, $p = .276$, with a small effect size, $\phi = 0.073$, the majority of participants in all groups correctly answered that atom 4 in the product shown in Figure 3-24 corresponds to atom G in the starting material shown. We still saw that some students (~40%) were unable to successfully use the isopropyl group as a mapping landmark, although using landmarks for facilitating mapping was explicitly discussed in the mapping workshop.

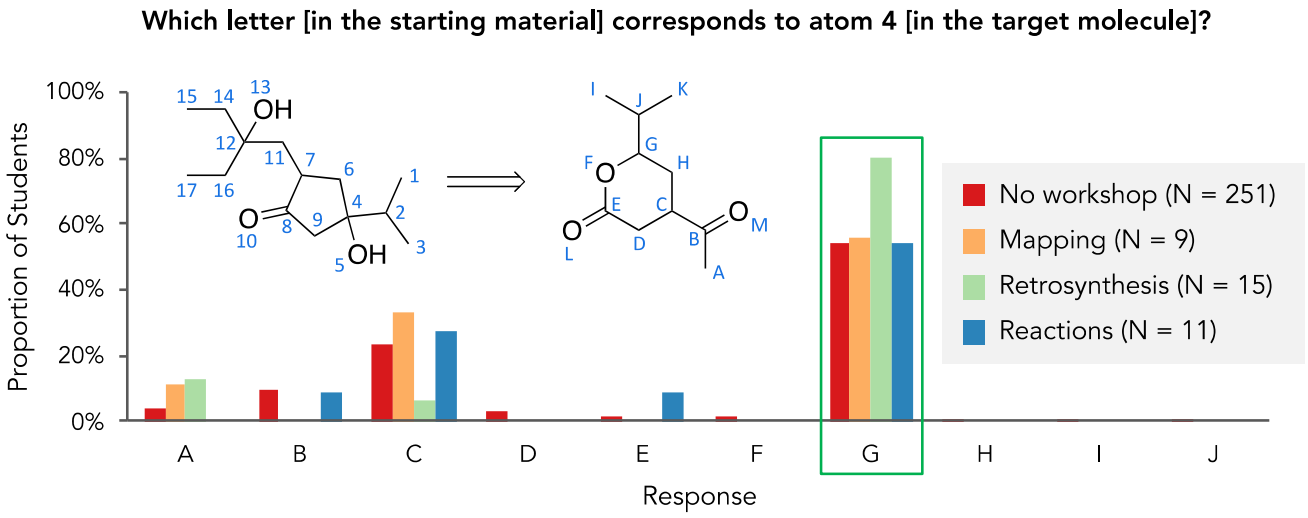


Figure 3-24. Students' responses to a formative assessment task from the participants' OC2 course, answered using a classroom response system. The correct response, G, is highlighted on the bar chart with a green box.

Students also had more difficulty when given a problem where they must map in the context of retrosynthetic analysis. Figure 3-25 contains students' responses to a formative assessment task where participants were asked to decide which of A or B maps more suitably onto the indicated alcohol. Once again, there was no significant difference between workshop participants and non-participants, $\chi^2(1) = 1.787$, $p = .198$, with a small effect size, $\phi = 0.077$. The correct answer, B, implies the C4–C5 bond was formed via aldol condensation, which is a direct route to the alkene. Although the majority of students did select the correct answer, a substantial proportion of students (~40%) selected the incorrect answer. These results are consistent with what we observed during the workshops; many students were capable of applying the strategies in isolation, but some may have required more scaffolding than what was already provided to be able to apply them to solving synthesis problems.

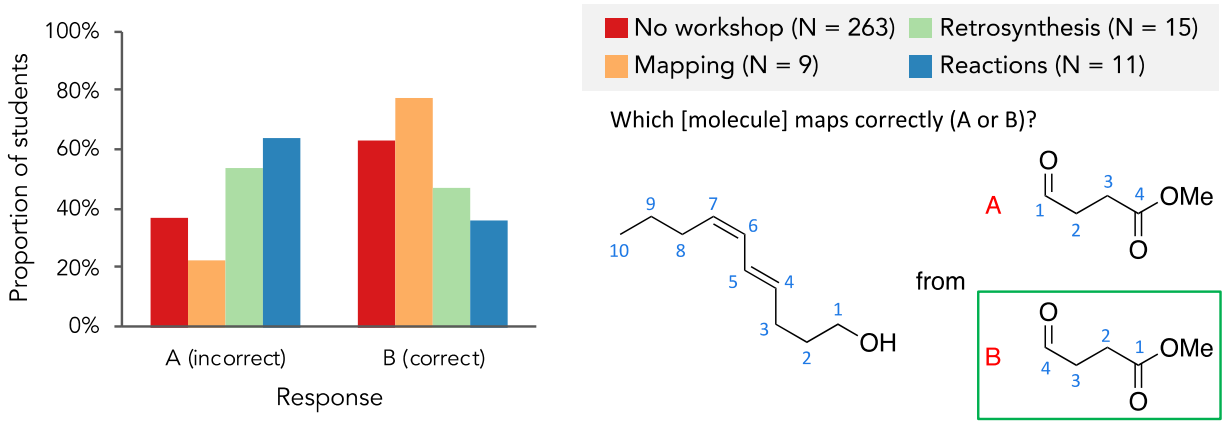


Figure 3-25. Students' responses to a formative assessment task from the participants' OC2 course, answered using a classroom response system. The correct answer, B, is indicated with a green box; mapping the atoms in this fashion implies an aldol reaction (with the aldehyde as the acceptor synthon) to form the C4-C5 bond; condensation gives the alkene directly.

Final exam: Workshop participants did not outperform students of similar ability, but overall success rates were high

Although students' performance on the workshop pre- and post-test was generally low, we saw a significant improvement overall in their performance on the synthesis problem on their final exam (Figure 3-12), $\Lambda = 0.470$, $F(2, 39) = 22.029$, $p < .001$, with a large effect size, partial $\eta^2 = 0.530$. Post hoc pairwise comparisons of final exam score with pre-test score and with post-test score were both significant, $p < .001$; the scores from the pre-test and post-test were not significantly different, $p = .737$, which is consistent with our prior analysis. In spite of this

improvement, we did not find a significant difference in scores on this problem between our workshop participant and the group of non-workshop participants with matched exam scores, $U(82) = 827.0$, $z = –0.129$, $p = .899$, with a small effect size, $r = –0.002$.

There are many possibilities for why participants experienced this significant improvement relative to the workshop, but not relative to a comparable subset of their OC2 cohort. There was much greater incentive to do well on the final exam compared to the workshop, and the workshop presented a large amount of information in a short time interval, which may have overwhelmed participants at the time.

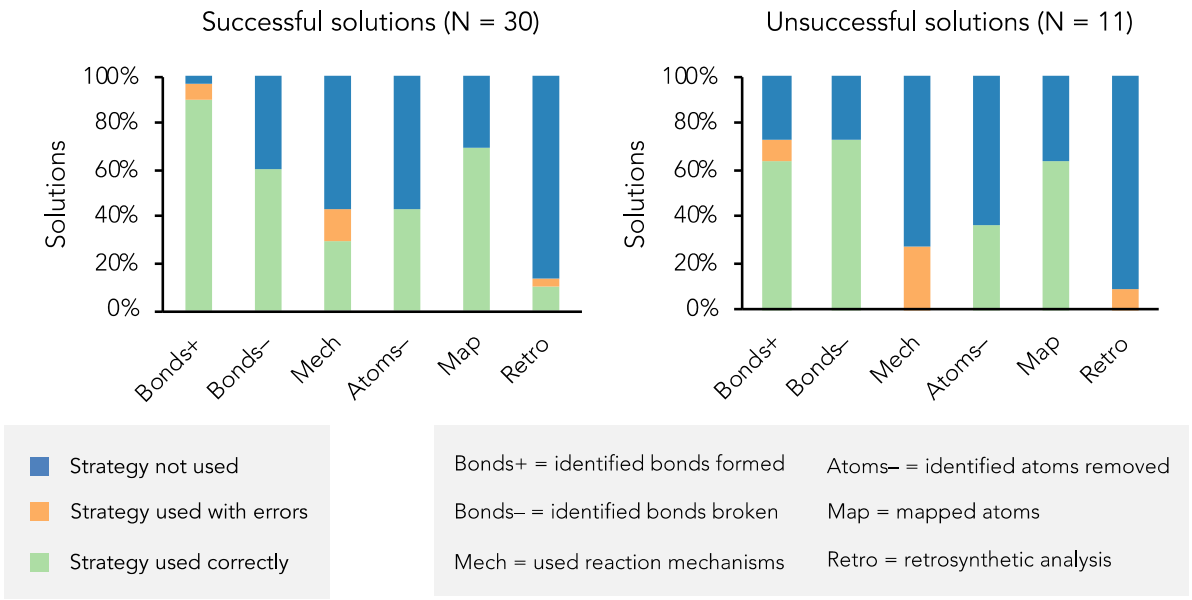


Figure 3-26. Key problem-solving strategies found in responses to the synthesis problem on the final exam.

A large increase in the use of key problem-solving strategies accompanied the increased rate of success, with the exception of retrosynthetic analysis; this exception could be because teaching and learning activities in class focused more heavily on mapping and related strategies, while retrosynthetic analysis using a synthon approach was not emphasized as strongly in class. Students proficiently answered the retrosynthetic analysis assessment tasks during the workshop including making polar bond disconnections (Figure 3-17), and constructing plausible synthons; therefore, if retrosynthesis were taught, practiced, and assessed throughout the entire semester like mapping has been, higher use of this strategy may be observed.

Since marks were awarded to students simply for showing their brainstorming and analysis (graded for completeness, not correctness), it is not surprising that more brainstorming was observed compared to the post-test, where participants were possibly fatigued, were not extrinsically motivated by grades, and were not explicitly prompted to use strategies. In this particular case, there was no significant association with the use of any particular strategy and whether or not students were successful in solving this problem. However, responses to this specific synthesis problem from a different cohort have been analyzed in our prior work; [8] we observed an 80% success rate on this problem among students who proposed a complete pathway (N = 89), despite the tendency that even students who provided successful solutions would only use one strategy (mean = 1.30, s.d. = 0.95). This suggests that this is a relatively easy problem, as other problems we analyzed in this study had success rates below 50% for our workshop participants as well as the fall 2017 cohort of Organic Chemistry 2 students. Our results suggest that awarding marks to students for explicitly providing their brainstorming and analysis may encourage them to be more methodical in their problem-solving approach, which could result in higher success rates than if students are not prompted to provide their explicit analysis.

Classroom Observation Protocol for Undergraduate STEM (COPUS)

The design and implementation of an active learning format for each workshop was carried out successfully. Analysis of the COPUS data collected from each workshop showed that there was a large proportion of active behavior among workshop participants (67% or greater), and that passive listening only accounted for roughly 33% (or less) of participants' behavior. Furthermore, the data also show that each facilitator spent the vast majority of the workshop on problem-solving activities, both clicker and non-clicker questions (74% or greater); lecturing (as defined by the COPUS protocol) accounted for less than 20% of facilitator codes for workshops A and C. Lecturing accounted for slightly more time in workshop B (26%), but this is likely because this workshop contained far more background information that was unfamiliar to participants (rules and conventions for retrosynthetic analysis). The data for each workshop are displayed in Figure 3-27 (Workshop A), Figure 3-28 (Workshop B), and Figure 3-29 (Workshop C).

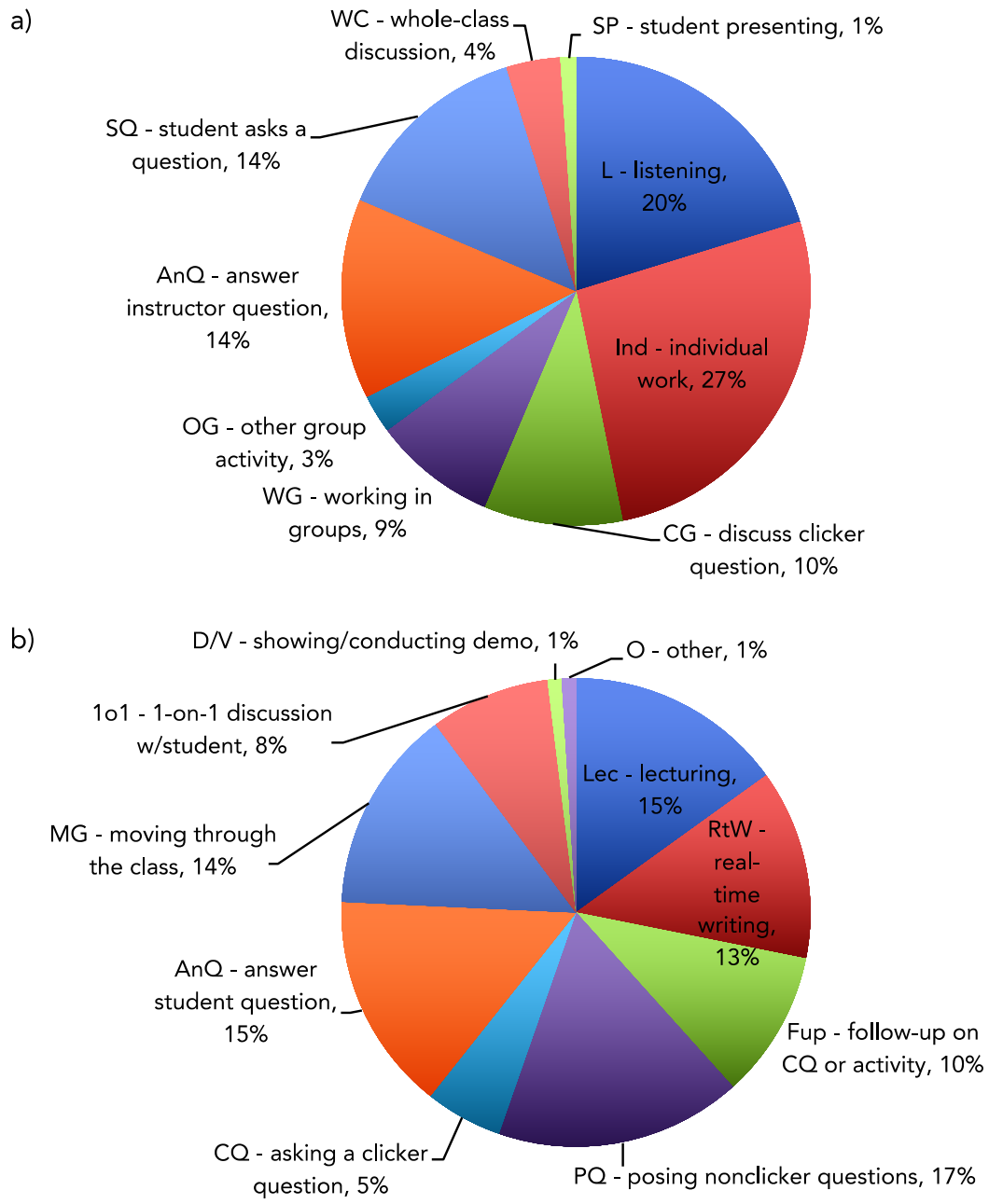


Figure 3-27. COPUS results from Workshop A (mapping-based analysis). (a) Relative proportions of each student code used. (b) Relative proportions of each instructor code used.

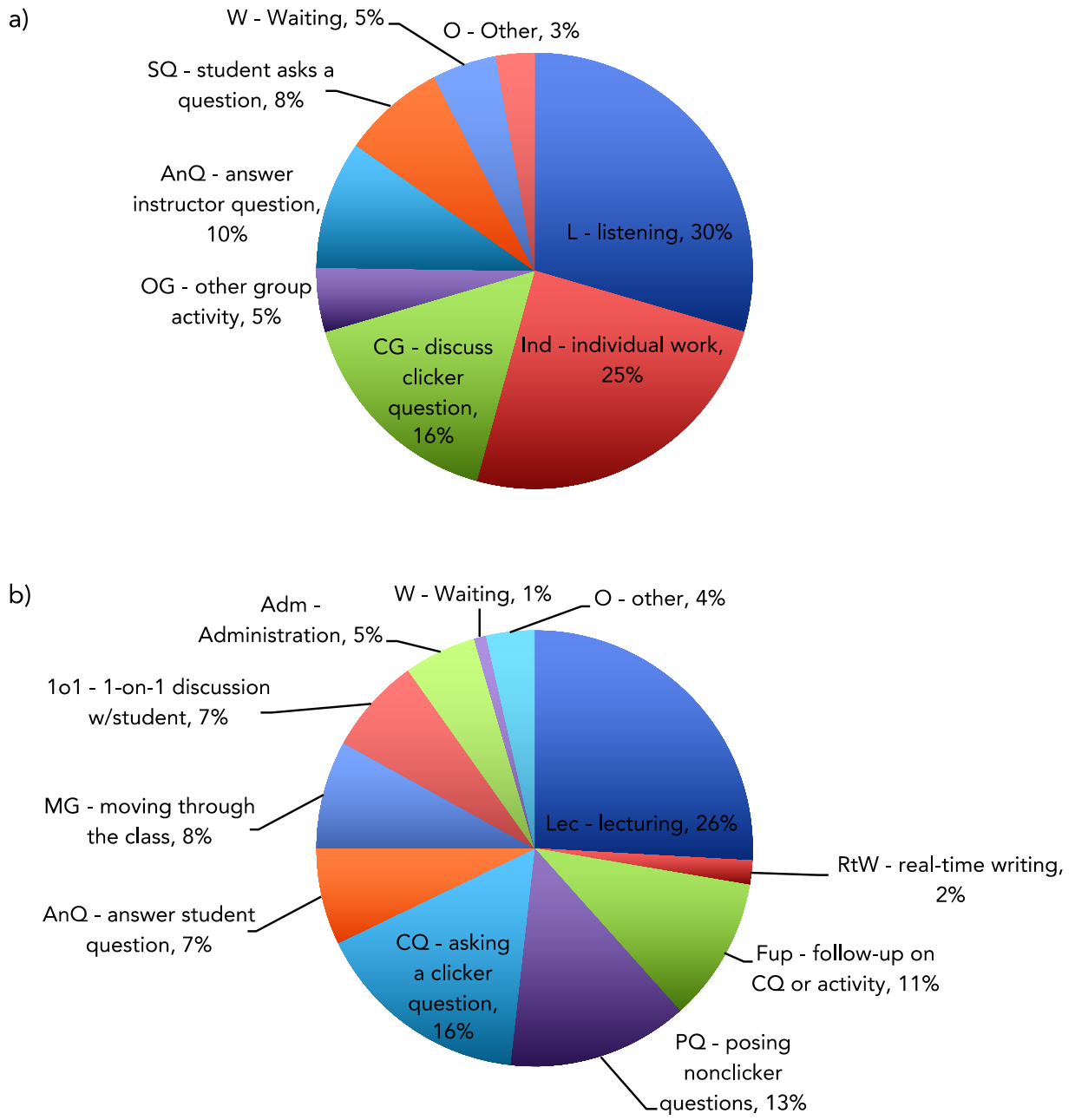


Figure 3-28. COPUS results from Workshop B (retrosynthetic analysis). (a) Relative proportions of each student code used. (b) Relative proportions of each instructor code used.

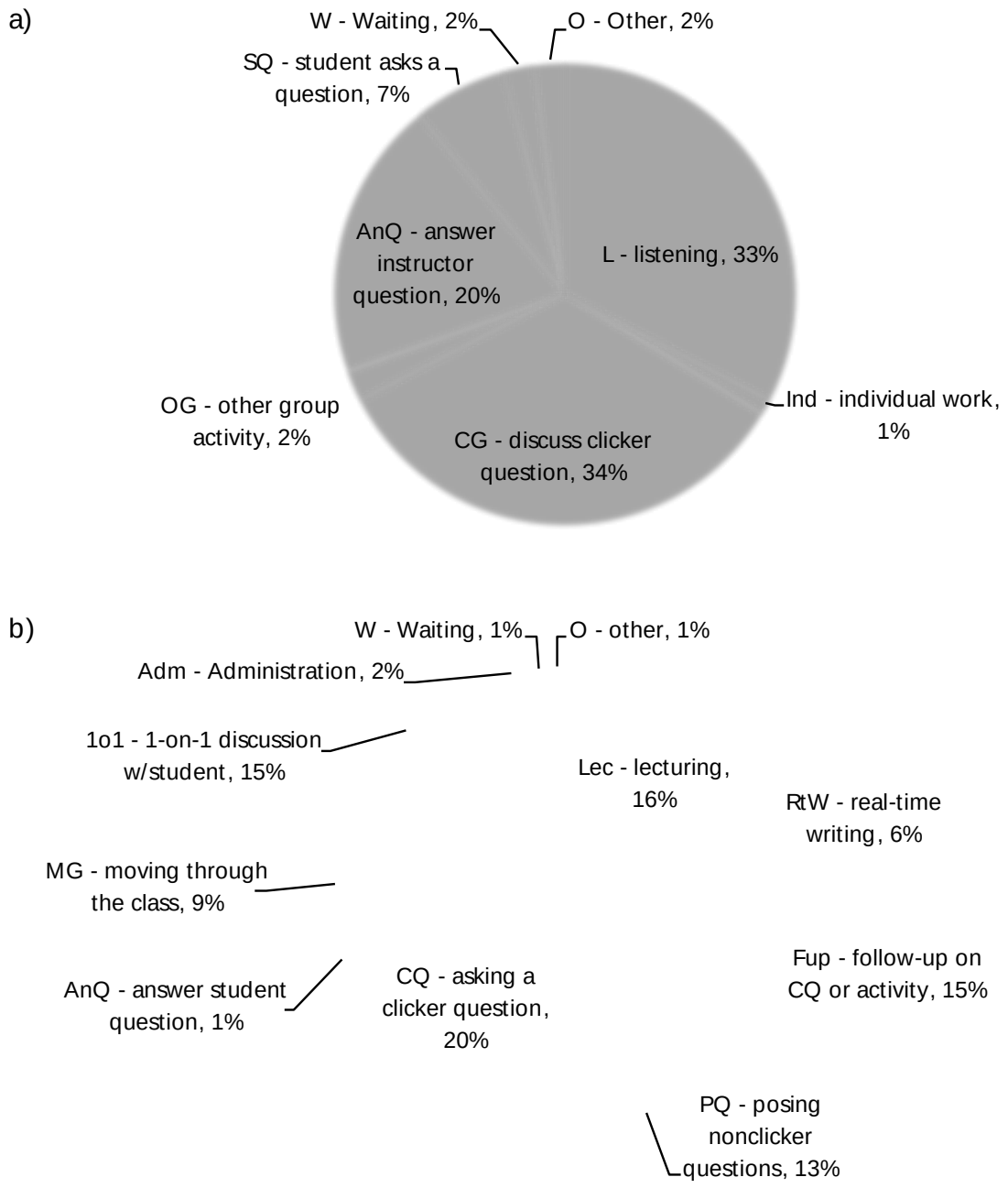


Figure 3-29. COPUS results from Workshop C (applying reactions to problems). (a) Relative proportions of each student code used. (b) Relative proportions of each instructor code used.

Metacognitive Activities Inventory (MCAI)

We analyzed participants' MCAI scores to investigate the presence of any mediating effects of metacognitive skillfulness on the effect of the workshops to help participants provide successful solutions to synthesis problems. There was good internal consistency between

participants' responses to the MCAI items (Cronbach's α value of 0.803), indicating good reliability in the data.

We found similar overall metacognitive skillfulness between the participants in each group. One-way ANOVA indicated the difference in MCAI scores between workshop groups was non-significant, $F(2,40) = 0.433$, $p = .652$, with a small effect size, $\eta^2 = 0.02$. The distribution of MCAI scores for each workshop is visualized in Figure 3-30. Pairwise comparisons are omitted due to the non-significant result of the ANOVA. We had originally hypothesized that metacognitive ability would be correlated with larger learning gains and that we might especially see differences in the first two workshops (more targeted strategies) than the broader workshop on reactions.

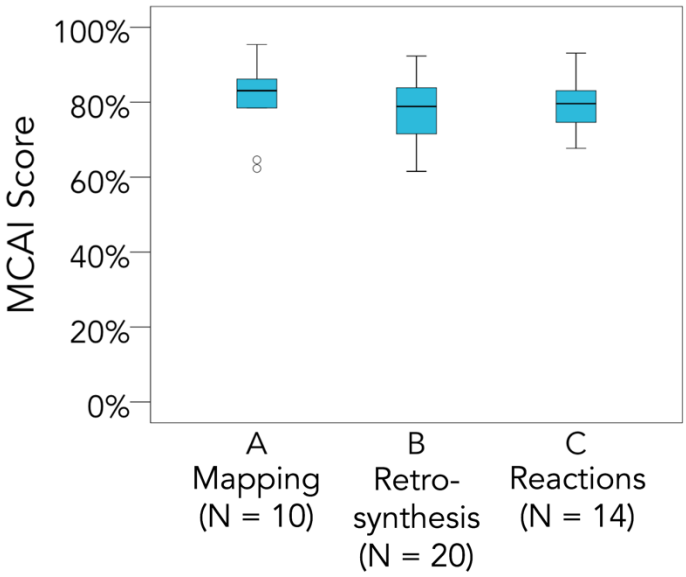


Figure 3-30. The distribution of MCAI scores was not significantly different across workshops, indicating that metacognitive skillfulness likely did not affect differences in learning gains.

We also found no significant difference in MCAI scores between participants who provided successful and unsuccessful solutions to the post-test (Figure 3-31), $t(41) = 0.033$, $p = .974$, with a small effect size, $d = 0.01$. Overall, we did not detect an effect of general metacognitive skillfulness on participants' synthesis problem solving skills during the intervention. This was surprising, because a previous application of the MCAI found that when differentiating by letter grade (A, B, C, or D) between different groups of students in a general chemistry course, an increase in letter grade was accompanied by a roughly 1–2% increase in corresponding mean MCAI score.[14]

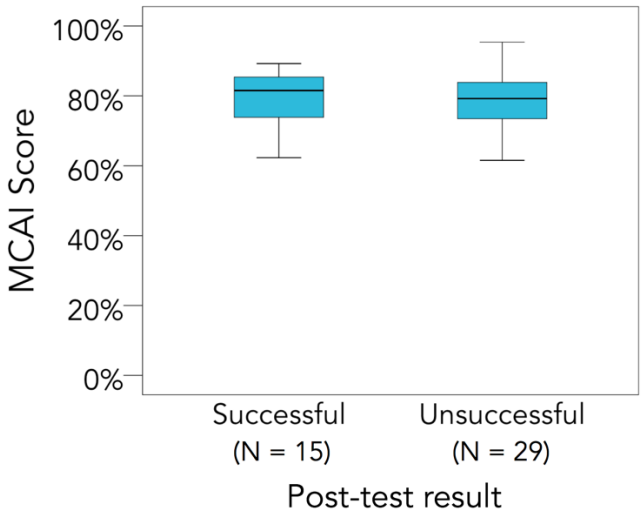


Figure 3-31. The distribution of MCAI scores was not significantly different between participants who provided successful post-test solutions and those who provided unsuccessful solutions. N = 43.

Participants' responses to two specific MCAI items are of note. The first is item 11: "I use graphic organizers (diagrams, flow-charts, *etc*.) to better understand problems." There was no significant difference in the distribution of responses to this item between participants who provided successful and unsuccessful post-test solutions, $U(44) = 173.5$, $z = -1.196$, $p = .245$, with a small effect size, $r = -0.18$, but in general, participants tended to disagree with this item (grand median = 2.00; Figure 3-32). One of our hopes for these workshops was that engaging in strategy-based learning activities would indicate to participants that committing key strategies to paper was an important part of successful synthesis problem solving. In future instruction, we will directly communicate the importance of committing strategies to paper. A possible limitation of this item is that the participants may have disagreed with this statement because it specifically refers to diagrams and flow-charts as examples of graphical organizers, and as a result, participants did not consider these strategies as graphic organizers. However, participants may have meant they were cognitively aware that they do not explicitly use strategies, so this item may indicate reason for the low success rates observed on the post-test.

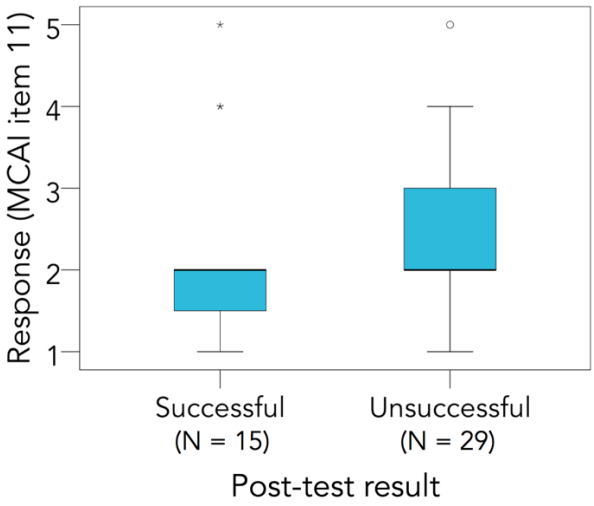


Figure 3-32. Distribution of participants' (distinguished by whether they provided successful or unsuccessful post-test solutions) responses to MCAI item 11: "I use graphic organizers (diagrams, flow-charts, etc.) to better understand problems." Responses: 1 = strongly disagree, 5 = strongly agree.

The second item of interest is item 27: "When practicing, if a problem takes several attempts and I cannot get it right, I get someone to do it for me and I try to memorize the procedure." This is the only item for which a significant difference in the distribution of responses was observed between participants who provided successful and unsuccessful post-test solutions, $U(44) = 130.5$, $z = -2.233$, $p = .026$, with a medium effect size, $r = -0.337$; the tendency was that participants who provided successful post-test solutions strongly disagreed with this statement, while participants who provided unsuccessful post-test solutions neither agreed nor disagreed with this statement. We believe an explanation for this result lies within the context of our working definition of a problem, where it is the condition of *unfamiliarity* that distinguishes a problem from an exercise.[45] Simply memorizing the procedure by which a given synthesis problem was solved (for instance, the synthesis problems presented as formative assessment tasks in the workshop) would not necessarily allow a student to provide a successful solution to a new problem, as the requisite procedural considerations for the new problem would not likely be the same as the procedure that student memorized. The implication is that organic chemistry students should be taught to focus on learning how to apply general strategies (such as the ones discussed in this manuscript) rather than memorizing how a given synthesis problem was solved and trying to follow that exact procedure in solving a novel problem.

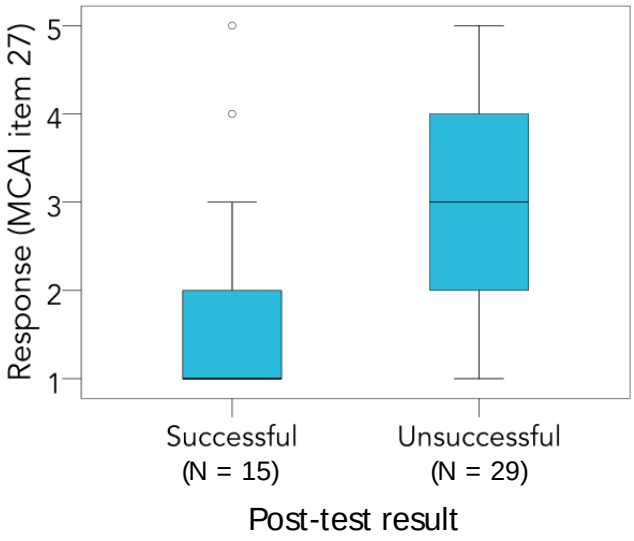


Figure 3-33. Distribution of participants' (distinguished by whether they provided successful or unsuccessful post-test solutions) responses to MCAI item 27: "When practicing, if a problem takes several attempts and I cannot get it right, I get someone to do it for me and I try to memorize the procedure." Responses: 1 = strongly disagree, 5 = strongly agree).

Attitudes Toward the Subject of Chemistry

Content validity of the ASCIv2

To evaluate the validity of the two-factor structure of the ASCIv2 in the study's setting, we performed confirmatory factor analysis (CFA) through structural equation modeling (SEM), using IBM SPSS AMOS software (version 24). Based on the CFA data, the estimation of the two-factor model fit is as follows: χ^2 (19) = 26.646, p = .113; CFI = 0.905; RMSEA = 0.097. The two-factor model does not fit our study's data exceptionally well, with a CFI value greater than 0.95 and a RMSEA value smaller than 0.08 being the accepted criteria.[46] However, we observed a high degree of correlation (0.70) between the latent factors of the ASCIv2, consistent with Xu and Lewis' findings; these factors should be related to some extent, because while they are not redundant, both pertain to the latent factor of attitude.[15] Cronbach's α values were 0.712 and 0.486 for the subscales pertaining to intellectual accessibility and emotional satisfaction, respectively. The first value is considered acceptable, but the latter value indicates low consistency in participants' responses to each item on the emotional satisfaction subscale, suggesting that the ASCIv2 did not reliably measure participants' emotional satisfaction with the subject of chemistry.

Due to less-than-desirable fit for the two-factor model, we explored an alternative one-factor model fit to see if it would better suit our data. However, the estimation of the one-factor model fit was worse than the two-factor model, based on the following fit statistics: χ^2 (20) = 36.83, p = .014; CFI = 0.796; RMSEA = 0.138. Therefore, we found the two-factor model still had the best fit for our data, and chose to continue our analysis of the ASCIv2 data we obtained using the two-factor model of the instrument. However, we do so with caution due to the low internal consistency in how participants responded to items pertaining to emotional satisfaction and to model fit statistics that fell just short of the accepted criteria.

A Shapiro-Wilk test indicated that the ASCIv2 data deviated significantly from a normal distribution ($p < 0.05$) for several items (Table 3-4. Shapiro-Wilk test results for each ASCIv2 item. W *refers to the test statistic.* N = 44.). Because the ASCIv2 data were not normally distributed, we analyzed these data using non-parametric tests (Kruskal-Wallis ANOVA and Mann-Whitney *U* tests).

Table 3-4. Shapiro-Wilk test results for each ASCIv2 item. W refers to the test statistic. *N = 44.*

ASCIv2 item (Chemistry is…)	Mean	Std. dev.	*W*	*p*
Hard/easy	3.25	1.22	0.926	.007
Complicated/simple	2.82	1.33	0.881	<.001
Confusing/clear	3.86	1.29	0.924	.007
Uncomfortable/comfortable	3.95	1.38	0.947	.043
Unsatisfying/satisfying	5.09	1.34	0.921	.005
Challenging/unchallenging	2.80	1.47	0.888	<.001
Unpleasant/pleasant	4.27	1.37	0.937	.019
Chaotic/organized	4.64	1.60	0.924	.006

ASCIv2: Results and Discussion

We did not observe any significant difference in attitudes toward the subject of chemistry across workshops with respect to either intellectual accessibility, $\chi^2(2) = 4.317$, $p = .115$, or emotional satisfaction, $\chi^2(2) = 2.135$, $p = .344$.[a] Pairwise comparisons are omitted due to the non-significant results of these Kruskal–Wallis tests; the distribution of ASCIv2 scores across workshops for each latent factor is displayed in Figure 3-34. Based on this result, it is unlikely that attitude had a major influence on learning gains.

[a] Effect sizes cannot be directly calculated for Kruskal-Wallis tests.

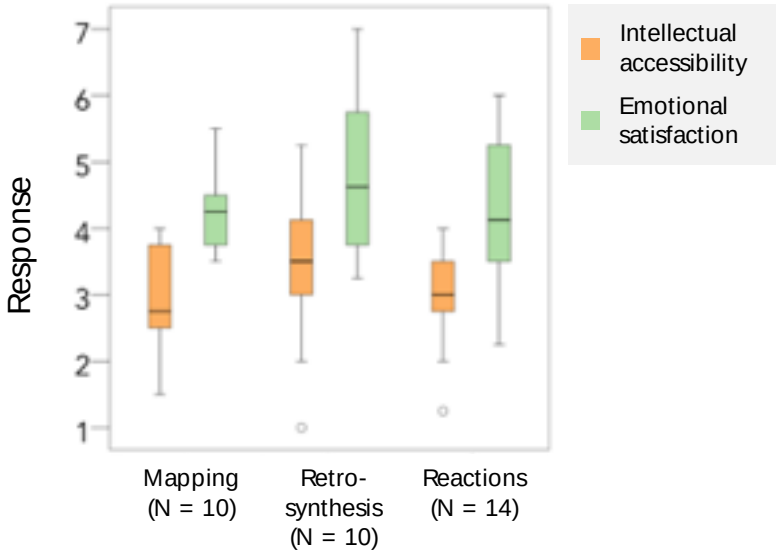


Figure 3-34. Distribution of ASCIv2 scores on each latent factor, organized by workshop. Workshop A = mapping (N = 10), Workshop B = retrosynthetic analysis (N = 20), Workshop C = reactions (N = 14).

Additionally, we did not observe any significant difference in attitudes toward the subject of chemistry between unsuccessful and successful problem solvers on the post-test with respect to either intellectual accessibility, $U(44) = 200.5$, $z = -0.423$, $p = .672$, $r = -0.004$, or emotional satisfaction, $U(44) = 246.0$, $z = 0.710$, $p = .478$, $r = 0.107$ (Figure 3-35).

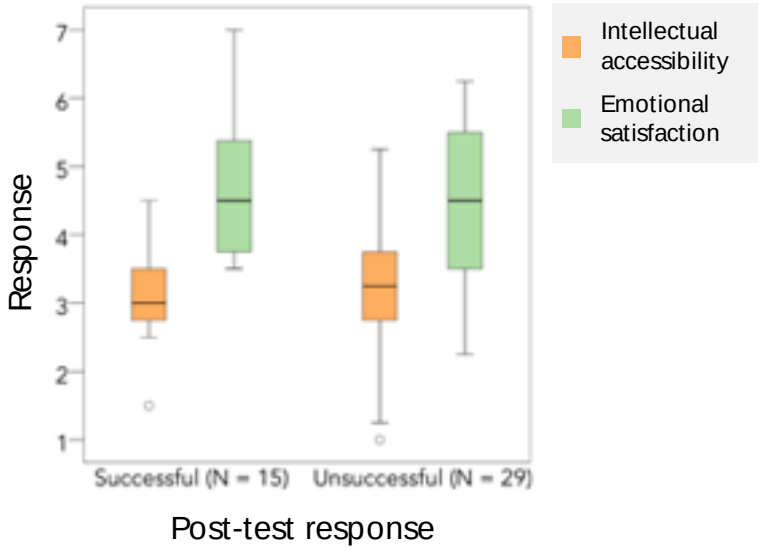


Figure 3-35. Distribution of participants' mean scores on the intellectual accessibility and emotional satisfaction subscales of the ASCIv2, organized by whether participants were successful or unsuccessful in solving the post-test problem. N (successful) = 15, N (unsuccessful) = 29.

However, participants' responses to the item "Chemistry is challenging/unchallenging" are of note. Although there was no significant difference between groups in the distribution of responses to this item, $U(44) = 190.0$, $z = -0.689$, $p = .491$, $r = -0.104$, overall, both groups rated

chemistry as challenging, suggesting that there may be a difference in how well successful and unsuccessful students perform in a situation they perceive to be challenging. This notion is related to self-efficacy beliefs, which refer to the extent to which one believes themselves capable of completing tasks and reaching goals.[47] Raker and co-workers proposed a reciprocal causation model for organic chemistry self-efficacy, where self-efficacy and academic performance influence one another in an ongoing fashion over the course of a semester. While we did not follow this study's participants' performance on problems related to organic synthesis in their OC2 course over the full semester, eight participants were successful on all three of the pre-test, post-test, and final exam synthesis problem, and each of these participants rated chemistry as challenging to some degree. In light of a reciprocal causation model, these participants may have been motivated by their past successes to persevere in a subject they find challenging.[48] Additionally, Dweck and co-workers have proposed that it is the students who believe intelligence is "malleable" who are more likely to persevere in spite of occasional failure.[49] Therefore, we hypothesize that successful participants may have a greater sense of self-efficacy, and propose that future studies of this nature may include a means for measuring student self-efficacy over the course of a semester to investigate the degree to which this relationship is present.

Conclusions & Implications for Instruction

The workshops and regular classes provided evidence that students can demonstrate proficiency in using strategies such as mapping and retrosynthetic analysis in isolation, they often struggle to, choose not to, or do not know to apply these strategies to solving synthesis problems, even those that can be solved using reactions they have already learned. We also observed no significant improvement in the ability of the workshop participants to solve synthesis problems following interventions focusing on strategies commonly used by undergraduate students who provided successful solutions to these types of problems. Our analysis provided some evidence that the synthesis problems we used for the pre- and post-test were not significantly different in degree of difficulty, but there were limitations to our evaluation.

Although we did not observe learning gains between the pre-test and the post-test, we still observed some positive outcomes; our analysis of students' responses to in-class learning activities showed that a large proportion of students were successful in solving formative assessment tasks focusing on the strategies taught in the mapping workshop. We therefore

hypothesize that since our prior work indicates that students can quickly pick up on the conventions of synthon-based retrosynthetic analysis, [7] they may be capable of applying a synthon approach to retrosynthetic analysis to synthesis problem solving if it were integrated into a semester-long curriculum. We believe that our findings demonstrate that a single intervention focusing on problem-solving strategies is unlikely to be sufficient, due to the inherent complexity of synthesis problem-solving.

We observed significant associations with medium to large effect sizes between the use of reaction mechanisms and atom mapping within successful solutions to the workshop pre-test, which supports our previous findings from exam analysis regarding the utility of these strategies.[8] However, we also observed that the vast majority of unsuccessful solutions did not include the correct use of key problem-solving strategies, and that several strategies required for a complete representation of a synthesis problem (*e.g.*, atoms added/removed, bonds formed/broken) were rarely applied. This finding is consistent with prior work highlighting completeness and correctness of representations (along with abstractness, which this study did not measure) as key determinants for successful problem-solving.[5]

Metacognitive skillfulness did not differ significantly between participants who provided successful and unsuccessful post-test solutions. Two statements from the MCAI yielded noteworthy results; students report that they do not tend to use graphic organizers, which was generally consistent with what we observed on the pre/post-test. Additionally, participants who provided unsuccessful post-test solutions were significantly distinct from those who provided successful solutions in reporting that they tend to try and memorize another student's problem-solving procedures when a situation is unfamiliar. This difference in memorization strategy suggests that some students may not have a useful approach to problem-solving and that the necessity of having explicit strategies to be able to solve unfamiliar problems should be emphasized to students, while discouraging memorization. We are unaware of any other use of this instrument in the organic CER literature; therefore, this study also provides new evidence of the MCAI's ability to reliably measure students' metacognitive skillfulness in a setting outside of a general chemistry course.

Attitudes toward chemistry did not differ significantly between participants who provided successful and unsuccessful solutions; however, in both populations, most participants rated chemistry as challenging (as opposed to unchallenging). This similarity in attitudes suggests there may be a difference in how more/less skilled or knowledgeable students perform

differently when they perceive a situation to be challenging, which would be consistent with literature regarding how self-efficacy and malleable intelligence beliefs influence learning.[7] Future studies could investigate how these factors might explain why both participants who provided successful and unsuccessful solutions alike believe organic chemistry in particular is challenging.

Although we did not administer the ASCIv2 to all OC2 students (only those who participated in the workshops) we believe that there is a fair chance that students who chose to participate in the workshops were more motivated to succeed in chemistry than the average OC2 student, and as such, the ASCIv2 scores we observed could be higher than what may be expected from a complete sample of OC2 students.

Based on our findings, we suggest that instructors should use learning activities related to synthesis learning outcomes are integrated throughout organic chemistry courses in order for students to develop these important problem-solving skills and that treating synthesis as a separate skill (with only a single isolated class period devoted to it) is unlikely to foster any significant learning gains.

Potential limitations

Despite our efforts to determine if the pre- and post-test are equal in difficulty (discussed in our content validity analysis), the evidence we obtained is not fully conclusive; a difference in difficulty may have contributed to the lack of observed learning gains in our workshops. Future studies of this nature may avoid this potential limitation by using identical instruments for the pre-test and post-test. Additionally, the majority of the workshops focused on the strategies themselves, and less time was spent applying them to authentic synthesis problems. For a more direct measurement of students' proficiency in using these strategies, the pre- and post-tests used in future work could also assess students' skills using the strategies without asking them to solve a synthesis problem. Slight differences in facilitators' styles or experience levels could have affected the outcomes of the study.

A possible limitation that we have addressed in the discussion of our findings was that participants may have been fatigued and therefore put forth less effort on the post-test than the pre-test, based on the presence of less overall use of strategies on the post-test. The participants were not compensated for attending the workshops and had no performance-based

incentives for successfully solving the post-test problem unless they were intrinsically motivated to succeed. Future studies of this nature may consider including some form of performance-based incentive to avoid this limitation. Lastly, for future work, we would ask demographic information at the end of the study, *i.e.*, after the post-test, to avoid the potential for stereotype threat. Asking demographic information last would be particularly important if research questions were related to the demographic information collected (*e.g.*, gender). Future studies may also explore the extent to which the MCAI can be used in different settings or what changes would need to be made.

References

(1) Kraft, A.; Strickland, A. M.; Bhattacharyya, G. Reasonable reasoning: multi-variate problem-solving in organic chemistry. *Chem. Educ. Res. Pract.* **2010**, *11* (4), 281–292 DOI: 10.1039/C0RP90003F.

(2) Christian, K.; Talanquer, V. Modes of reasoning in self-initiated study groups in chemistry. *Chem. Educ. Res. Pract.* **2012**, *13* (3), 286–295 DOI: 10.1039/C2RP20010D.

(3) Sevian, H.; Bernholt, S.; Szteinberg, G. A.; Auguste, S.; Pérez, L. C. Use of representation mapping to capture abstraction in problem solving in different courses in chemistry. *Chem. Educ. Res. Pract.* **2015**, *16* (3), 429–446 DOI: 10.1039/C5RP00030K.

(4) Weinrich, M. L.; Talanquer, V. Mapping students' modes of reasoning when thinking about chemical reactions used to make a desired product. *Chem. Educ. Res. Pract.* **2016**, *17* (2), 394–406 DOI: 10.1039/C5RP00208G.

(5) Domin, D.; Bodner, G. Using Students' Representations Constructed during Problem Solving To Infer Conceptual Understanding. *J. Chem. Educ.* **2012**, *89* (7), 837–843 DOI: 10.1021/ed1006037.

(6) Flynn, A. B. Developing Problem-Solving Skills through Retrosynthetic Analysis and Clickers in Organic Chemistry. *J. Chem. Educ.* **2011**, *88* (11), 1496–1500 DOI: 10.1021/ed200143k.

(7) Flynn, A. B. How do students work through organic synthesis learning activities? *Chem. Educ. Res. Pract.* **2014**, *15* (4), 747–762 DOI: 10.1039/C4RP00143E.

(8) Bodé, N. E.; Flynn, A. B. Strategies of Successful Synthesis Solutions: Mapping, Mechanisms, and More. *J. Chem. Educ.* **2016**, *93* (4), 593–604 DOI: 10.1021/acs.jchemed.5b00900.

(9) Bhattacharyya, G.; Bodner, G. M. Culturing reality: How organic chemistry graduate students develop into practitioners. *J. Res. Sci. Teach.* **2014**, *51* (6), 694–713 DOI: 10.1002/tea.21157.

(10) Bowen, C. W.; Bodner, G. M. Problem-solving processes used by students in organic synthesis. *Int. J. Sci. Educ.* **1991**, *13* (2), 143–158 DOI: 10.1080/0950069910130202.

(11) Moreno, R.; Ozogul, G.; Reisslein, M. Teaching with concrete and abstract visual representations: Effects on students' problem solving, problem representations, and learning perceptions. *J. Educ. Psychol.* **2011**, *103* (1), 32–47 DOI: 10.1037/a0021995.

(12) Nicol, D. J.; Macfarlane Dick, D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Stud. High. Educ.* **2006**, *31* (2), 199–218 DOI: 10.1080/03075070600572090.

(13) Smith, M. K.; Jones, F. H. M.; Gilbert, S. L.; Wieman, C. E. The Classroom Observation Protocol for Undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices. *CBE Life Sci. Educ.* **2013**, *12* (4), 618–627 DOI: 10.1187/cbe.13-08-0154.

(14) Cooper, M. M.; Sandi Urena, S. Design and Validation of an Instrument To Assess Metacognitive Skillfulness in Chemistry Problem Solving. *J. Chem. Educ.* **2009**, *86* (2), 240–246 DOI: 10.1021/ed086p240.

(15) Xu, X.; Lewis, J. E. Refinement of a Chemistry Attitude Measure for College Students. *J. Chem. Educ.* **2011**, *88* (5), 561–568 DOI: 10.1021/ed900071q.

(16) Bodner, G. M. Constructivism: A theory of knowledge. *J. Chem. Educ.* **1986**, *63* (10), 873 DOI: 10.1021/ed063p873.

(17) Bodner, G. M.; Orgill, M. *Theoretical Foundations of Learning Environments*; Routledge: Upper Saddle River, NJ, 2007.

(18) Vygotsky, L. S. *Mind in Society*; Cole, M., Souberman, E., John-Steiner, V., Scribner, S., Eds.; Harvard University Press: Cambridge, MA, 1978.

(19) Solomon, J. Social Influences on the Construction of Pupils' Understanding of Science. *Stud. Sci. Educ.* **1987**, *14* (1), 63–82 DOI: 10.1080/03057268708559939.

(20) DeLozier, S. J.; Rhodes, M. G. Flipped Classrooms: a Review of Key Ideas and Recommendations for Practice. *Educ. Psychol. Rev.* **2016**, *29* (1), 1–11 DOI: 10.1007/s10648-015-9356-9.

(21) Biggs, J. B.; Collis, K. F. Origin and Description of the SOLO Taxonomy. In *Evaluating the Quality of Learning*; Elsevier, 1982; pp 17–31.

(22) Biggs, J. B.; Tang, C. *Teaching for Quality Learning at University*, 3rd ed.; Society for Research into Higher Education: Maidenhead, UK, 2007.

(23) Novak, J. D. Human constructivism: A unification of psychological and epistemological phenomena in meaning making. *Int. J. Pers. Construct Psychol.* **1993**, *6* (2), 167–193 DOI: 10.1080/08936039308404338.

(24) Flynn, A. B.; Ogilvie, W. W. Mechanisms before Reactions: A Mechanistic Approach to the Organic Chemistry Curriculum Based on Patterns of Electron Flow. *J. Chem. Educ.* **2015**, *92* (5), 803–810 DOI: 10.1021/ed500284d.

(25) Bruice, P. Y. *Organic Chemistry*, 7 ed.; Pearson: Upper Saddle River, NJ, 2014.

(26) Clayden, J.; Greeves, N.; Warren, S. *Organic Chemistry*, 2nd ed.; Oxford UP: New York, NY, 2012.

(27) Klein, D. *Organic Chemistry*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, 2012; Vol. 127, pp 14952–14952.

(28) Prince, M. Does active learning work? A review of the research. *J. Eng. Educ*. **2013**, *93 (3)*, 223–231. DOI: 10.1002/j.2168-9830.2004.tb00809.x.

(29) Moog, R. S.; Spencer, J. N. POGIL: An Overview. In *Process Oriented Guided Inquiry Learning (POGIL)*; Moog, R. S., Spencer, J. N., Eds.; ACS Symposium Series; American Chemical Society: Washington, DC, 2009; Vol. 994, pp 1–13.

(30) Hein, S. M. Positive Impacts Using POGIL in Organic Chemistry. *J. Chem. Educ.* **2012**, *89* (7), 860–864 DOI: 10.1021/ed100217v.

(31) Luxford, C. J.; Crowder, M. W.; Bretz, S. L. A Symmetry POGIL Activity for Inorganic Chemistry. *J. Chem. Educ.* **2012**, *89* (2), 211–214 DOI: 10.1021/ed1007487.

(32) Flynn, A. B. Structure and evaluation of flipped chemistry courses: organic & spectroscopy, large and small, first to third year, English and French. *Chem. Educ. Res. Pract.* **2015**, *16* (2), 198–211 DOI: 10.1039/C4RP00224E.

(33) Flynn, A. B. Flipped Chemistry Courses: Structure, Aligning Learning Outcomes, and Evaluation. In *Tools of Chemistry Education Research*; ACS Symposium Series; American Chemical Society: Washington, DC, 2017; Vol. 1261, pp 151–164.

(34) Raker, J. R.; Towns, M. H. Designing undergraduate-level organic chemistry instructional problems: Seven ideas from a problem-solving study of practicing synthetic organic chemists. *Chem. Educ. Res. Pract.* **2012**, *13* (3), 277–285 DOI: 10.1039/C1RP90073K.

(35) Watkins, J.; Coffey, J. E.; Redish, E. F.; Cooke, T. Disciplinary authenticity: Enriching the reforms of introductory physics courses for life-science students. *J. Phys. Rev. ST Phys. Educ. Res*. **2012**, *8 (1)*, 010112. DOI: 10.1103/PhysRevSTPER.8.010112.

(36) Nicolaou, K. C.; Yang, Z.; Liu, J. J.; Ueno, H.; Nantermet, P. G.; Guy, R. K.; Claiborne, C. F.; Renaud, J.; Couladouros, E. A.; Paulvannan, K. Total synthesis of taxol. *Nature* **1994**, *367* (6464), 630–634 DOI: 10.1038/367630a0.

(37) Wagschal, K. C.; Pyun, H. J.; Coates, R. M.; Croteau, R. Monoterpene biosynthesis: isotope effects associated with bicyclic olefin formation catalyzed by pinene synthases from sage (Salvia officinalis). *Arch. Biochem. Biophys.* **1994**, *308* (2), 477–487 DOI: 10.1006/abbi.1994.1068.

(38) Kim, H. C.; Kang, S. H. Total synthesis of azithromycin. *Angew. Chem. Int. Ed.* **2009**, *48* (10), 1827–1829 DOI: 10.1002/anie.200805334.

(39) Corey, E. J.; Cheng, X. M. *The Logic of Chemical Synthesis*; John Wiley & Sons: New York, NY, 1989.

(40) Bauer, C. F. Beyond “Student Attitudes”: Chemistry Self-Concept Inventory for Assessment of the Affective Component of Student Learning. *J. Chem. Educ.* **2005**, *82* (12), 1864 DOI: 10.1021/ed082p1864.

(41) Jamieson, S. Likert scales: how to (ab)use them. *Med. Educ.* **2004**, *38* (12), 1217–1218 DOI: 10.1111/j.1365-2929.2004.02012.x.

(42) Bhattacharyya, G.; Bodner, G. M. “It Gets Me to the Product”: How Students Propose Organic Mechanisms. *J. Chem. Educ.* **2005**, *82* (9), 1402 DOI: 10.1021/ed082p1402.

(43) Grove, N. P.; Cooper, M. M.; Rush, K. M. Decorating with Arrows: Toward the Development of Representational Competence in Organic Chemistry. *J. Chem. Educ.* **2012**, *89* (7), 844–849 DOI: 10.1021/ed2003934.

(44) Bhattacharyya, G. Trials and tribulations: student approaches and difficulties with proposing mechanisms using the electron-pushing formalism. *Chem. Educ. Res. Pract.* **2014**, *15* (4), 594–609 DOI: 10.1039/C3RP00127J.

(45) Randles, C. A.; Overton, T. L. Expert vs. novice: approaches used by chemists when solving open-ended problems. *Chem. Educ. Res. Pract.* **2015**, *16* (4), 811–823 DOI: 10.1039/C5RP00114E.

(46) Hu, L. T.; Bentler, P. M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Modeling* **1999**, *6* (1), 1–55 DOI: 10.1080/10705519909540118.

(47) Ormrod, J. E. *Educational psychology: Developing learners*, 5 ed.; Pearson, 2006.

(48) Villafañe, S. M.; Xu, X.; Raker, J. R. Self-efficacy and academic performance in first-semester organic chemistry: testing a model of reciprocal causation. *Chem. Educ. Res. Pract.* **2016**, *17* (4), 973–984 DOI: 10.1039/C6RP00119J.

(49) Mangels, J. A.; Butterfield, B.; Lamb, J.; Good, C.; Dweck, C. S. Why do beliefs about intelligence influence learning success? A social cognitive neuroscience model. *Soc. Cogn. Affect. Neurosci.* **2006**, *1* (2), 75–86 DOI: 10.1093/scan/nsl013.

Chapter 4 Using cognitive interviews to investigate student reasoning in synthetic problem–solving

Introduction

Goals of the study

Following the investigation outlined in Chapter 3 of this thesis, our last goal was to obtain a more detailed understanding of how students work through synthetic problems than we could obtain from analyzing their final answers; specifically, we wanted to obtain insight into students' thought process as they solved problems, rather than only being able to see the end product of their thinking. It is important to understand students' reasoning as they work through organic synthesis problems because in a long-term sense, our goal is to help students develop the ability for expert-like thinking about organic synthesis, and if we know where they are, then it facilitates the process of getting them where we want them to be. Specifically, this means that we want students to use their understanding of patterns of reactivity in organic chemistry to engage in retrosynthetic analysis, where reactions are proposed based on strategic bond disconnections. In the context of problems where starting materials are given to students, this process is also facilitated by other strategies discussed in this thesis, such as mapping atoms, and explicitly identifying bonds formed/broken and atoms added/removed.

To learn about students thought processes as they solve synthesis problems, we conducted problem-solving interviews with eight Organic Chemistry 2 students. Broadly, this study investigated the following research question:

> *How do students use the key strategies (taught in class and/or in the workshops) to evaluate and/or solve synthetic problems?*

The "how" of this question concerns the mental models that participants constructed, the pictorial representations of their mental models that they were able to construct. We investigated their mental models through the lens of the heuristic-analytic theory of reasoning, and we evaluated the quality of the representations of those mental models using three criteria: completeness, correctness, and abstractness.

The heuristic–analytic theory of reasoning

Beginning with Kahneman and Tversky's seminal 1974 *Science* paper, *Judgment under Uncertainty: Heuristics and Biases*,[1] there has been a trend in the cognitive psychology literature to describe reasoning in two systems, broadly defined as System 1 and System 2. Generally speaking, System 1 processes are often defined as "impulsive" and "automatic" operations, such as the forms you would use to complete the phrase "bread and…", or to answer "2+2 = ?". Meanwhile, System 2 processes are defined as "effortful" and "concentrated"; they place some degree of cognitive strain on the individual. These systems have a wide variety of names, but we will use Kahneman and Tversky's terms: *heuristic* reasoning (System 1) and *analytic* reasoning (System 2).[2]

Heuristic reasoning was first discussed by Tversky and Kahneman in the context of their research finding that people tend to evaluate the probability of an event occurring "by the ease with which relevant instances come to mind." They defined this specific heuristic as *availability*. [3] In the subsequent 45 years since the availability heuristic was first defined, many studies have obtained similar findings. A general example of how the availability heuristic can influence cognition is a study by Schwarz and co-workers,[4] who asked college freshmen/sophomores at the University of Illinois to rate on an 11–point scale ("not so happy" to "extremely happy"):

> *"How happy are you with life in general?"*
> *"How happy are you with your dating?"*

When asked the questions in the order just given (general happiness then dating happiness), there was a low correlation in participants' responses to the two questions (N = 60, r = .16). However, when the questions were asked to another group of students in the reverse order (dating happiness then general happiness), there was a significantly higher correlation between responses (N = 60, r = .55, p < .007). As one's dating life is only one of many parts of what might be expected to contribute to their overall happiness, the authors suggest that this higher correlation can be attributed to participants being cued by the *specific* area of happiness (their dating life) to respond in terms of that case when asked about their *general* happiness, because it was the most accessible case at the time.

Heuristic reasoning has several advantages; it is quick, allows us to propose solutions to problems we may not fully understand, and may help us focus on identifying the most important parts of a problem.[5] That said, we expect that the novice organic chemistry student is unlikely to

possess a broad enough set of case knowledge to be successful using this mode of reasoning to solve unfamiliar synthesis problems, as it is unlikely they would possess enough useful cues or be able to correctly identify the most important parts of a problem without some conceptual understanding of retrosynthesis or other well-defined strategies. Additionally, while heuristic reasoning can still result in success when a useful case is cued by the problem, this success often occurs without the problem-solver really understanding *why* their answer was successful.[6] Therefore, students' reliance on heuristics can result in a barrier to meaningful learning.

Mental models

While many researchers have explored dual-system theories of reasoning, our research is informed by the work of Jonathan St.–B. Evans and Keith Stanovich. According to Evans, this so-called dual-process theory of reasoning supposes that two interdependent cognitive processes can occur during reasoning: hypotheses about the phenomenon are generated by the heuristic system, and these hypotheses are evaluated by the analytic system.[7] In the process of reasoning, we construct and use what are referred to as mental models.

Broadly, mental models are defined in the cognitive psychology literature as representations of a hypothetical situation; one uses mental models to explore possibilities when engaging in hypothetical thinking.[8] Mental models are defined by Bodner and co-workers as "a representation of an idea, object, event, process, or system, which concentrates attention on certain aspects of the system – therefore facilitating scientific inquiry." [9] When we think, plan, and try to explain events, we manipulate elements of mental models; as such, our knowledge of the world depends on our ability to construct models of it. There are three principles that guide this process of constructing mental models: [8]

1. Singularity principle: mental models used to represent a hypothetical situation are generated one at a time

2. Relevance principle: mental models are generated by preconscious heuristic processes

3. Satisficing principle: mental models are likely to be accepted unless analytic processes give the individual good reason to reject them

Based on these three principles, there are fundamental biases that can arise when the heuristic system generates mental models that fail to include information that is relevant to generating a solution, or that include information that is irrelevant. Biases may also arise in

analytic reasoning due to the satisficing principle. Satisficing is a principle based on the concept of bounded rationality: the idea that it is impossible to optimize the choices we make, because the amount of thought needed to do so would take too long to ever *make* a choice. This principle states that people tend to generate mental models that are "good enough" rather than optimal.[10]

This model provides a good framework for understanding how hypothetical thinking and mental model construction occurs, but we agree with Stanovich's believe that a dual-system model is too narrow. Stanovich argues for a tripartite model: along with the heuristic system (what Stanovich calls the autonomous system) the analytic system should be subdivided further into *reflective* thinking (which requires cognitive decoupling) and *algorithmic* thinking (what is termed serial associative cognition). Serial associative cognition refers to a form of reasoning that goes deeper than the hypotheses that are generated by preconscious cues, but that is still *bounded* to the single mental model that was generated by those preconscious cues. As such, further reasoning is limited to exploring confirming evidence for the hypothesis generated by preconscious cues. Meanwhile, cognitive decoupling refers to the *conscious* decision one makes to engage in deeper analytic reasoning, where one generates and explores *alternative* mental models that can be improved through acknowledging disconfirming evidence for heuristically cued mental models.[11,12]

Broad models for problem-solving exist for general problems[13] as well as problems specific to organic synthesis.[14-16] However, based on Piaget's functioning model,[17] we believe the most effective models for novice problem solvers are likely to be ones that they can construct themselves using well-defined strategies that allow them to incorporate fundamental concepts of organic chemistry in a way that is consistent with their existing conceptual frameworks surrounding the discipline.

These ideas regarding the use of models for solving problems are also related to the findings of previous research in students' strategy use during synthetic problem solving, that students tend to be more successful when they create a "map" between a synthetic target and given starting materials, facilitating the process of identifying where atoms have been added and removed and bonds have been formed and broken, which, in turn, facilitates the process of proposing a series of bond-forming reactions.[18]

Evaluating constructed representations of mental models

Domin and Bodner provide a framework for evaluating students' constructed representations of their mental models;[19] that the representation's quality should be judged on its' accuracy, completeness, and abstractness. They define accuracy as how closely their representation resembles the pertinent concept(s) as presented during instruction, and completeness as the extent to which their representation fully incorporates symbols that were a part of the representations presented during instruction.

For the purposes of this study, consider the synthetic strategy of retrosynthetic analysis as an example; using this framework, we would judge a participant's representation as higher quality if they:

- disconnected bonds using a clear and systematic approach such as a synthon-based approach, as presented in the workshops described in the previous chapter
- executed each step of synthon construction (identify a bond to disconnect based on mapping analysis, construct and evaluate synthon pairs, choose analogous reagents)

Domin and Bodner define the last element, abstractness, as the extent to which the representation incorporates elements from that were not part of the representation used when the concept was originally presented. In the context of constructivist theories of learning, it is presumed that these additional elements are contributions from the student's prior knowledge that they have decided are relevant to the situation at hand. Sevian and co-workers have also noted that more abstract reasoning, where analysis is sufficiently removed from the specific problem at hand to be generalized enough to apply to any similar situation, tends to be associated with the ability to propose more plausible solutions to problems.[20]

Methods

Cognitive interviews

A suggested activity for producing clear conceptualizations of participants' thought processes is the cognitive think-aloud interview. These interviews allow us to address a major limitation of previous research using exam question responses:[18] that only the product of students' analysis can be analyzed, not their process. In the think-aloud cognitive interview

protocol we used, participants were asked to complete open-ended tasks related to synthetic problem solving, and asked to tell the interviewer what they were thinking about as they completed each task as best they could. Analyzing the transcripts of these interviews then allows us to take the participants' thoughts into context with what they chose to write on paper, giving us a more complete picture of how they approach these types of problems.

Cognitive interviews are a common research tool in chemistry education research, especially in recent studies that have investigated the modes of reasoning this study is interested in. Most recently, the Weinrich and Talanquer study discussed earlier used a semi-structured think-aloud interview protocol to investigate students' modes of reasoning about inorganic synthesis problems and single-step organic synthesis problems.[21]

An important consideration when conducting cognitive interviews is that the interviewer should attempt to achieve saturation with any point of interest, but should do so without leading the participant in any way. [22] Saturation is achieved when the participant has nothing left to say about the point of interest. Closed questions or leading questions are important to avoid in this setting because they suggest to the participant that there is a specific answer that the interviewer wants to hear, and their response is less authentic with respect to their actual thought process. To avoid leading participants, we used the most general language possible, or framed questions in a way that allowed the interviewer to use the participant's own words.[6]

Participants were not corrected on their use of chemical nomenclature or on their understanding of concepts, so that they were allowed to define how they conceptualized given problems in their own words.[23] Rather than initially cueing the participant to certain parts of the problem that the interviewer was interested about, participants were asked in a general fashion to "tell me about how you decided to approach the problem", so that the participant could decide what they thought was important before being cued to what the interviewer thought was important. Participants were reminded at the start of the interview that there was not a single correct answer or method for completing the tasks they were asked to solve.

Limitations have been identified with the think-aloud interview. Participants may be unskilled at thinking aloud, and therefore, "activities associated with think-aloud speech may serve to burden or contaminate the cognitive processes" of the participant.[23] With this consideration in mind, we instructed participants that they should only think aloud if they were comfortable doing so and if it would not distract them from thinking about the interview tasks.

The interviewer then asked the participant to explain their thought process once they were ready to do so.

Only students who participated in a workshop and who completed the research study outlined in Chapter 4 were invited to participate in interviews, because the data obtained from this work was required to contextualize the interview outcomes. For instance, if a participant exhibited more complex modes of reasoning, we were interested to know how this associated with their metacognitive skillfulness, attitudes towards chemistry, and in which workshop they had participated.

Interview participants completed two synthesis-type tasks (Figure 4-1 and Figure 4-2). In the first task, participants were given a synthetic problem along with three correct solutions (participants were not told whether or not the pathways were correct), and were asked to evaluate each pathway and decide which pathway they thought was best. This synthesis problem was taken from a midterm exam used in Organic Chemistry II at the authors' institution in a previous year, and the solutions given to participants were the three most common answers given by students in that course. In the second task, participants were asked to propose a synthesis of a given target molecule, from given starting materials. After they were finished or stated they were unable to complete the problem, they were given a solution to the problem and were asked to evaluate this solution.

Participants were invited to think aloud as they completed each task only if they were comfortable doing so, as we did not want the think-aloud protocol to distract participants from their reasoning about the task at hand. After participants completed each task, they were prompted by the interviewer to reflect upon the types of reasoning they used to solve the problem. The interviewer used a predefined set of interview prompts to elicit participants' reasoning about different parts of each task, as well as the specific methods participants used to complete each task. Other prompts used by the interviewer were based on observational notes taken as the participant solved the problem.

Both video and audio recording was used in these interviews; a video camera was used to record their writing so that their process could be reviewed in sync with the audio recording, which was transcribed verbatim. The use of both audio and video recording allowed us to link their thoughts with their actions, which was essential to obtaining a complete picture of the problem-solving process.

The structure of the interviews was piloted prior to the administration of the interview structure in its final format, in order to ensure the tasks elicit a response from participants where they recognize the need to develop mental models. [9] The interview protocol is provided in the following section, which also includes a summary of the reactions taught in the course. The reaction summary sheet, which was the same as the summary sheet given in the workshops, was a resource that participants also had to help them complete the interviews.

Development of the interview protocol

Instructions to participants were based those given to participants in existing studies that used a think–aloud protocol. At the start of each interview, participants were reminded about the purpose of the study and the confidential nature of their participants. They were then informed about the various types of prompts that would be used by the interviewer, and that we were more interested in how they (the participant) were thinking about the tasks rather than the answers they were able to provide.[22] Further, they were informed about the open–ended nature of the tasks; we told each participant that there was no single correct method for completing the tasks, and that not all of the tasks had a single correct answer. [24]

We were more interested in what types of reasoning participants used to solve the interview tasks than their ability to recall course material. Therefore, we provided participants with one of the organic chemistry textbooks that were in the syllabus for their course, and a table that summarizes the reactions learned in Organic Chemistry I and II at the participants' institution. Previous research, conducted in a setting similar to this study, showed that students were often unable to effectively use a textbook or reaction list to look up unfamiliar reactions. [25] As such, our participants were reminded to use the textbook's index to help them do so.

Development of the interview activities

The initial interview guide piloted in Summer 2016 consisted of two full "propose a synthesis" problems, along with a series of scaffolding tasks between the first and second synthetic problem, which gave participants the opportunity to practice mapping and identifying atoms added/removed and bonds formed/broken before proposing a solution to a second synthetic problem. However, analysis of these pilot interviews showed that it was difficult to determine how participants were reasoning about the problem; we were only able to determine what concrete strategies they were using to solve the problem. We also had concerns about

participant fatigue, as this interview guide had seven tasks for participants to complete. Therefore, we decided to reduce the number of tasks from seven to three, two of which are discussed in this study.

The following tasks are discussed herein:

1. *Please evaluate each of the synthetic pathways on the page, then decide which one you think is best.* (Figure 4-1) [It was left to the participant to decide how they wanted to define "best", rather than telling them what we wanted them to look for when deciding.]

2. *Please propose a synthesis of the product below using the provided starting materials and any other reagents you think are necessary. For each step in your proposed pathway, provide as many reaction conditions as you think are important for that step*. (Figure 4-2)

The synthetic problem used in Task 1 was chosen because it had been previously used on midterm exams for Organic Chemistry 2, so we knew its difficulty was appropriate for our participants. Furthermore, the solutions we asked participants to evaluate were authentic responses provided by students on that midterm, so we thought it was likely that participants would be familiar with the reagents used in each synthetic pathway.

The synthetic problem used in Task 2 was designed to place more emphasis on the carbonyl–type chemistry learned in the participants' course in the timespan between the workshops and the interviews. This was done to increase the authenticity of the problems (that is, to increase the likelihood that the situations presented by the problems would be novel to the participant). The target molecule (product) was based on the structure of the anti-inflammatory drug rofecoxib, but simplified to make it appropriate for the Organic Chemistry 2 level.

The sequence of the problems in the interview was designed with the intent to refresh the participant on the conventions behind proposing synthetic pathways and thinking about the outcomes of reactions before having to propose a full synthesis, so that they could reasonably be expected to exhibit evidence of more concrete problem–solving strategies and more structured modes of reasoning when solving a full synthetic problem compared to simply evaluating given answers to a synthetic problem.

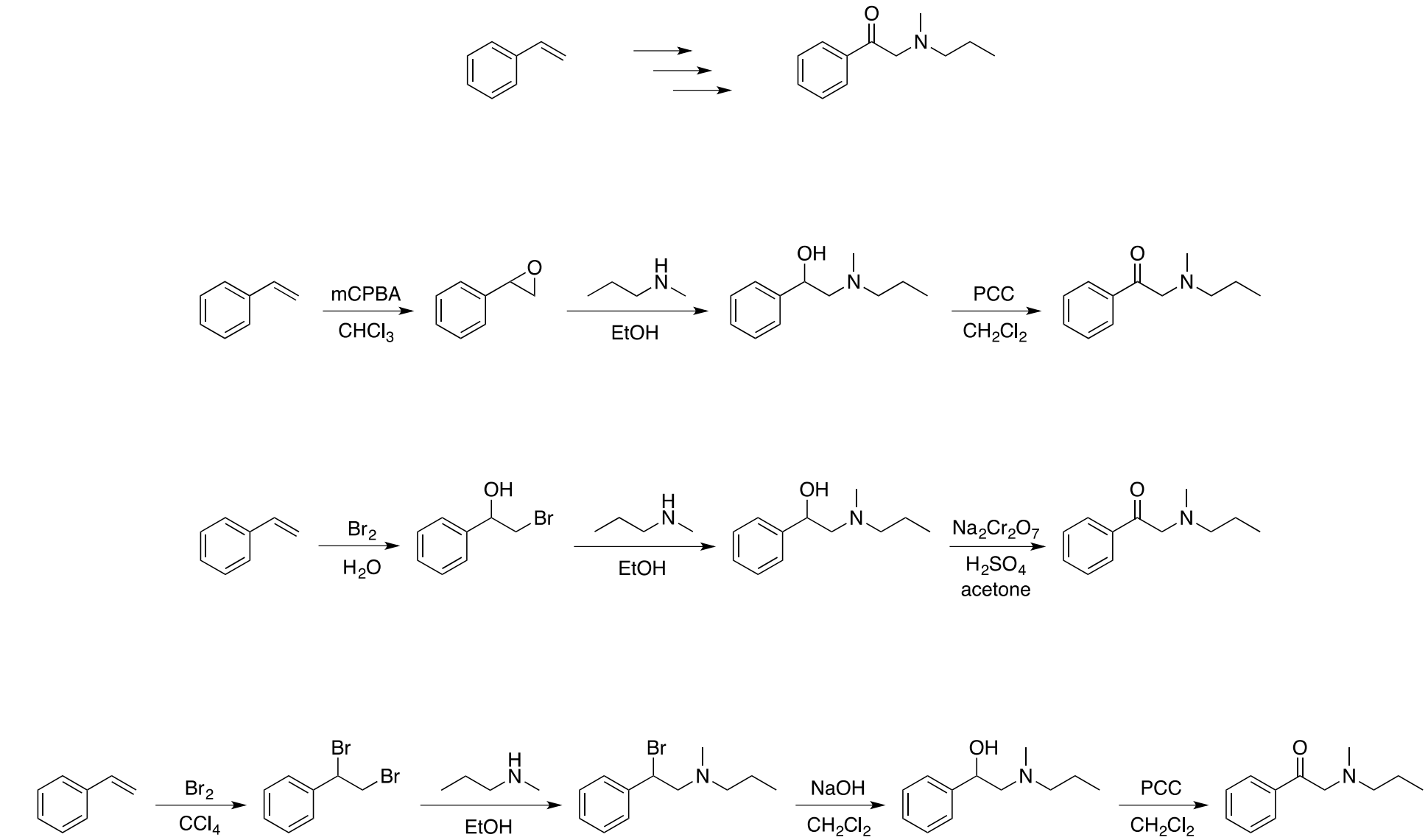


Figure 4-1. Task 1 of the cognitive interview used in this study. mCPBA = 3-chloroperoxybenzoic acid, PCC = pyridinium chlorochromate. 4 iterations of the task used the problem as shown above, while 4 iterations had the structures of mCPBA and PCC instead of the names.

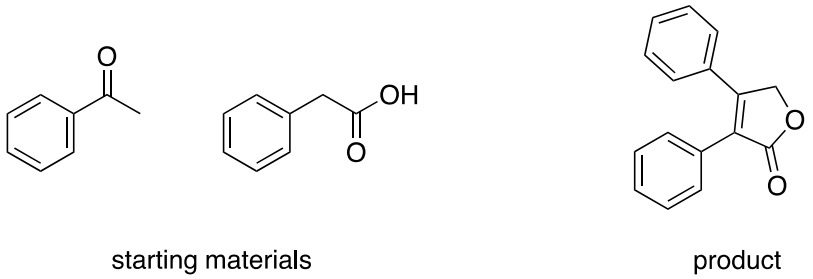


Figure 4-2. The starting materials and product used for the synthetic problem presented in Task 2 of the cognitive interview used in this study.

Interview analysis

After completing all eight interviews, the video recordings were used to transcribe the audio from the interviews verbatim, with annotations added to describe what the participant was doing when the audio lacked this important context. Transcription was carried out by the first author of this research, as well as a research assistant. Additionally, for the analysis of task 2, interview artifacts (participants' work on paper) were scanned, and the first author manually annotated the PDF files using a tablet to highlight common strategies the students used to complete this task. Notes and hypotheses regarding participants' reasoning were also added to contextualize these strategies within the overall sequence of what participants did, so that these annotations could be easily linked to pertinent quotes from the interview transcript. An example of these annotations is illustrated in Figure 4-3.

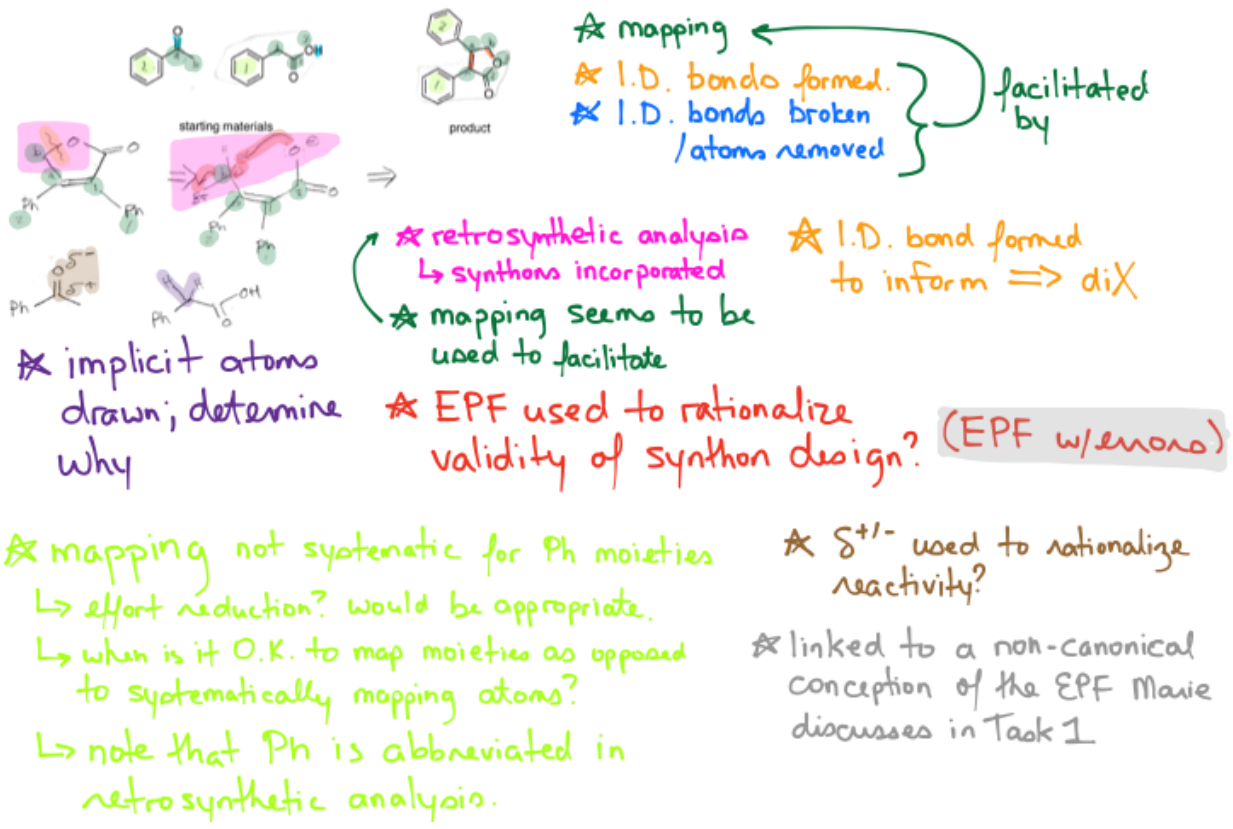


Figure 4-3. An example of how a participant's (Marie) approach to task 2 was annotated to highlight their key strategies, and to summarize questions/hypotheses to consider in the analysis of their interview transcripts. EPF = electron-pushing formalism.

The interview transcripts were analyzed in NVivo using a closed-coding system that was primarily interested in identifying common themes linked to the guiding research questions, and based on the theories of reasoning discussed in the introduction to this chapter. We were primarily interested in determining what strategies participants used to solve the problems, and what types of reasoning they were employing. When participants were asked to explain how they worked through the first task (evaluation of answers to a synthetic problem, the explanations about their reasoning that they provided were coded as either case-based, rule-based, or model-based, as the construction of full representations of their mental models was not generally observed in this task. When analyzing participants' work during the second task, we looked at the models they constructed as part of their problem-solving process, and evaluated these models for their accuracy, completeness, and abstractness. Their follow-up discussion of their reasoning and their problem-solving approach was used to inform this analysis. Our results and discussion of the second task are discussed in the context of these three important qualities of a representation.

Validity

There are multiple conclusions that may be drawn from a set of qualitative data, depending on the lens of the researcher in terms of knowledge, theoretical framework, and paradigm assumptions. In our case, we are analyzing the data through a constructivist paradigm, where the goal is not to define some form of empirical "truth", but to define a context-dependent perspective of reality that can provide an understanding of the group or individuals being studied, and empower action.[26,27]

To establish validity of our findings during the process of data analysis, we engaged in negative case analysis. The author of this thesis met with the research supervisor on a weekly basis to discuss his progress in establishing the themes that were present in the data, and how those themes aligned with the theoretical framework for this study. We then searched for evidence in the data that would disconfirm the presence of these themes. This active searching for disconfirming evidence causes the researcher to actively reflect on the extent to which their conclusions are consistent with the data, and with the guiding theoretical framework for their study.

Additionally, all discussion of our findings is accompanied by detailed quotes from the participants, along with their work on paper when it is relevant. These details are provided in the interest of transparency, so that a reader might review the evidence for our findings themselves, and although they may have other ideas about the data based on their own prior knowledge and experience, the conclusions presented herein should still be logically consistent with the data and theoretical frameworks.

Investigating students' use of heuristics

An important consideration in our analysis of participants' reasoning was that historically, the heuristics and biases literature has been criticized for primarily emphasizing how heuristic reasoning can lead to errors, and how many explanations of common types of heuristics are vague to the point where nearly any experimental result can be attributed to some form of heuristic in a *post-hoc* fashion, while attention would be better focused on providing detailed accounts of reasoning processes.[28]

Shining a negative light on heuristic reasoning in and of itself is inappropriate; Graulich and co–workers argue that heuristics are an essential tool to the expert chemist faced with a problem to solve. However, they point out that this type of thinking depends on being able to

think “conceptually, and not only to memorize,” (p. 1505) and that the chemist must have a high degree of cognitive organization for this type of thinking to be successful.[29]

This is an important point with regards to retrosynthetic analysis. We wish to highlight that although retrosynthetic analysis has been referred to as a heuristic strategy, due to it being an approach with defined rules that can be applied in a somewhat algorithmic fashion,[6] our study is populated with novice organic chemistry students who are familiar with the conventions for constructing synthons via polar bond disconnections, but are not fluent in advanced strategies for retrosynthetic analysis (*e.g.*, topological strategies, transform-based approaches). Therefore, it may be appropriate to classify it as an analytic process when it is carried out by novices, as they may require more meaningful reflection with mechanistic models than with the application of synthetic rule–based reasoning that may be beyond the scope of their knowledge.

For example, a commonly taught association heuristic is the connection between a 1,3 C–C bond disconnection and an aldol reaction, which is a useful association in the synthetic problem participants were given in the interview used as a data collection instrument in this study. However, we would treat this type of reasoning as distinct from the type of reasoning that would allow a participant to make a polar bond disconnection to identify starting materials if this disconnection was based on some form of deeper reflection.

In context with Domin and Bodner’s framework for evaluating representations, we can look at participants’ representations of their mental models, and evaluate the extent to which their reasoning has a high degree of cognitive organization if they are engaging in heuristic reasoning, or serial associative cognition, with the hypothesis that an over-reliance on this form of reasoning may be due to limitations in how synthetic problem solving skills are taught and assessed.

Results and Discussion

Task 1: Evaluate three solutions to a synthetic problem, and choose which one you think is best

Most participants' discussion of their reasoning about reaction outcomes was based on memorized rules

Participants generally approached Task 1 by going through each pathway in a somewhat stepwise fashion and attempting to rationalize each reaction outcome, using rule-based reasoning to do so. Rule–based reasoning describes situations where a problem solver uses "a single chunk of knowledge" (*ref 6,* p. 282) in isolation; the solver might use several rules to propose a solution to a problem, but each rule is only applicable to the situation at hand, and links between the rules are not necessarily established.[6] The steps that participants were most interested in were those involving substitution reactions, where participants invoked memorized rules to assess whether the indicated reagents would give the indicated substitution (or epoxide-opened) product.

This approach was evident as participants worked through the task, but Jane and Marie also emphasized this as a key point of interest for themselves during the follow-up questions we asked once they chose the pathway they thought was best:

Jane

> *Interviewer: Um, so how did you decide where to start in solving the problem?*
>
> *Jane: Um, I guess just looking at... going to the one like that…um, at a quick glance, that I would have recognized and comparing it to the other ones where the other ones might... making sure it's going to work.*

Marie

> *Interviewer: Okay. Um... so, how did you decide where to start solving this problem?*
>
> *Marie: Um, I usually like, go through each answer and see if there's any like, glaring problems [...]*
>
> *Interviewer: Can you tell me what sort of uh, problems, were there certain problems you were looking for, or were you just... yeah, tell me a bit more about what you were looking for.*
>
> *Marie: Well, a problem being like using the wrong reactant, like if you had mCPBA and then it formed some other compound, like you know that - that,*

that [mCPBA] goes with an epoxide. I just, I kind of remember, like, memorized the reactants and then what they're supposed to form, so, I was looking for that, and um... I just, I guess, things that just really didn't make sense, but they all seem logical.

Brandon and Holly

Two participants, Brandon and Holly, gave more detailed accounts of how they reasoned through this task, which made it obvious that they were applying this algorithmic form of reasoning to rationalize the synthetic pathways they were given. Brandon and Holly's approach to rationalizing reaction outcomes was based on a specific table that they were given in their OC2 lab course (Table 4-1), which outlines general rules for what mechanism between $S_N1/S_N2/E1/E2$ would prevail when given a certain nucleophile/base and electrophile/acid. Holly attempted to reproduce this table on her worksheet from memory (Figure 4-4).

Table 4-1. The expected outcome of the reaction of an alkyl halide with nucleophiles/bases. Reproduced from the laboratory manual for the Organic Chemistry 2 Laboratory course from the Fall 2016 semester at the University of Ottawa.

R–X	**Nucleophile/Base Strength**			
	Strong/Strong	Strong/Weak	Weak/Strong	Weak/Weak
Methyl	S_N2	S_N2	E2	-
Primary	S_N2	S_N2	E2	-
Secondary	E2	S_N2	E2	S_N1/E1
Tertiary	E2	-	E2	S_N1/E1

[page break in lab manual]

Strong/Strong	HO^-, RO^- (non-bulky), R_2N^- (non-bulky)
Strong/Weak	$RCOO^-$, I^-, N_3^-, CN^-, R-S^-, R-NH_2, R-SH
Weak/Strong	bulky RO^-, bulky R_2N^-, H^-
Weak/Weak	H_2O, ROH, RCOOH

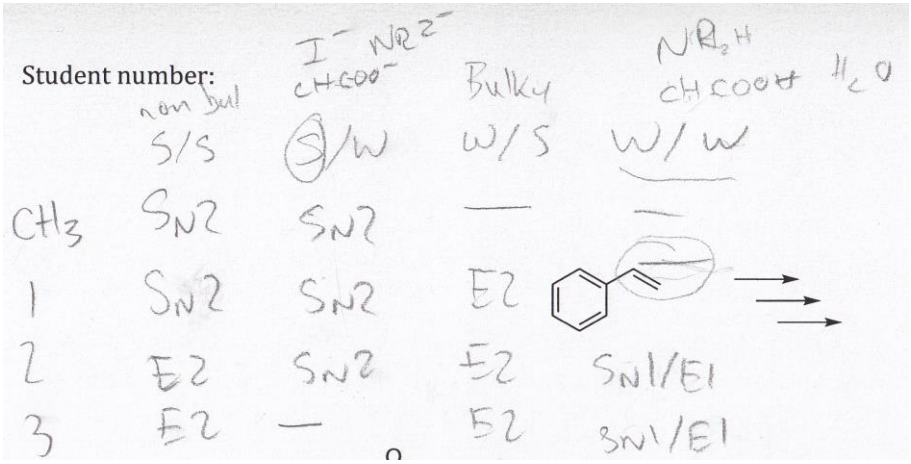


Figure 4-4. Holly's attempt to reproduce Table 4-1 from her OC2 lab manual.

This tool could be useful for students to quickly make decisions about the outcome of a substitution/elimination reaction, but without any further reasoning about the situation at hand, this strategy is highly dependent on having a strong memory of what types of nucleophiles and bases are considered strong or weak.

Brandon also refers specifically to this table but does not try to reproduce it:

> *Brandon: So, looking at the alpha carbon is primary so I'm trying to remember what that means. It's probably going to be substitution, I believe and then looking at the solvent system, I would think SN2.*
>
> *Interviewer: Ok, can you tell me more about how you reasoned through that just now? So, you're saying primary is substitution?*
>
> *Brandon: Well, I remember a chart I got in my lab course outlining and notes in my textbook outlining uh...we would have a methyl, primary, secondary, or tertiary carbon and then it would say describe the solvent system involving the alkalinity and I believe it was nucleophilic strength of the solvent system... and then it would tell you...describe which mechanisms are most likely to take place and which ones would take place at the same time.*

Brandon also discussed the use of a decision-making flow chart that was, in spirit, analogous to this table:

> *Brandon: There was a flow chart that I made a little while ago to prepare for the midterm...making me think about where...if I'm only thinking about those mechanisms. The first step would be to look at the, um – how substituted the alpha carbon is, and then I believe it was the strength of the base for all of them after that, and depending on that you would look at the, um, something else and then if that was...and then if that was wrong or no in the solvent system to see if it was I believe it was aprotic. If it was or wasn't, it would a combination of SN1 and E1 or no, it would be a combination of all of them or none of them.*

While Brandon and Holly's interviews provide more definitive evidence for heavy reliance on rule-based reasoning, four other participants (Skyler, Jane, Marie, and Gretchen) also made statements that suggest they are engaging in a similar form of reasoning.

Skyler

> *Skyler: Adding that to the more substituted carbon instead of where they added it in this. [circles product of second step of pathway 3] Like I would add it, I would have added it here [benzylic carbon] instead of here [where it is actually added].*
>
> *Interviewer: Okay, uh, why would you add it to that carbon?*
>
> *Skyler: Because I remember learning that, at least from - this is probably wrong - but I remember learning that uh, the more substituted carbon is more stable than the least substituted carbon.*

Jane

> *Interviewer: So, you talked a bit about um trying to invoke certain rules for understanding what the product would be. Um, how do you recognize situations where you need to apply those rules, and how do you remember what the outcome of applying that rule would be?*
>
> *Jane: Yeah, um... to be honest I think that's one of the hardest rules about synthesis, because there is so many options you can take and there is so many different rules, I guess. I kind of just think, I pick for example, S_N2 has to have, um, a good leaving group, and it needs to be strong basic conditions, like just remembering the conditions and what kind of exceptions can apply. [...] Stuff like that. Um, yeah, it's basically memorization.*

Marie

> *Marie: I can't remember if water would attack at the most substituted or the least substituted carbon of the epoxide.*
>
> *[...]*
>
> *Interviewer: Do you think one is more likely than the other?*
>
> *Marie: Usually, the um... I get it backwards with the most substituted-least substituted. The - the one that's like, more easily accessible, so like, that's the least substituted carbon is usually faster, but it produces like, a more unstable product. But...*
>
> *Interviewer: Can you tell me why the product is more unstable?*
>
> *Marie: Because... [laughs] good question. Because that's what I learned [laughs].*

Gretchen

> *Gretchen: [Referring to first intermediate of third pathway] Here you would want the SN2 reaction again. But, the fact that this bromine is also a good leaving group then it's on a second- or it's a... secondary um. It forms a secondary carbocation, so a SN1 reaction could possibly happen.*

Gretchen, like Brandon and Holly, also invoked her Organic Chemistry 2 lab course; however, her thinking about the outcome of this reaction was based on the outcome of a similar reaction she had carried out. As she describes it, this initially appears to be an example of case-based reasoning, but her explanation makes it seem that she is only recalling the experiment to facilitate the interviewer's understanding of where she learned the set of rules described by Brandon and Holly:

> *Gretchen: Um, so we did an experiment where...like we were looking at the reaction rates of E2 and SN2.*
>
> *Interviewer: Okay.*

Gretchen: Ok, and that really helped me out actually with understanding it and, um we did this sort of experiment where it was either a secondary carbocation or if the bromine was here [on primary carbon]. Um, so then if the bromine was here [on primary carbon] it would be like more likely to be a SN2, here [bromine on secondary carbon] it would be E2.

Participants often did not use the electron-pushing formalism in an expert-like fashion

Based on participants' explanations of how they rationalized reaction outcomes and/or considered alternative mechanisms for each step of each pathway, the trend we observed in the previous section was a possible over-reliance on the use of commonly-taught rules or algorithms for determining what mechanism will prevail between S_N1/S_N2/E1/E2. This was problematic when participants misremembered these rules, or misremembered how these *general* rules might react fail to consider other factors that dictate mechanism (for instance, the electronics of the phenyl ring in this example).

This approach to determining the likelihood of a reaction occurring is at odds with the intended learning outcome for the participants' organic chemistry course, where students are expected to be able to use the electron-pushing formalism (EPF) to propose a mechanism, and then evaluate the likelihood of that reaction mechanism occurring as they have drawn it. In this setting, five participants did use the EPF, but only Andrea appears to use it in a context where it is adequately rationalizing the regiochemical outcome of a given reaction: the formation of the β-bromo alcohol in the first step of pathway 2 (Figure 4-5). She also uses the EPF in the subsequent step, but she did not discuss this mechanism.

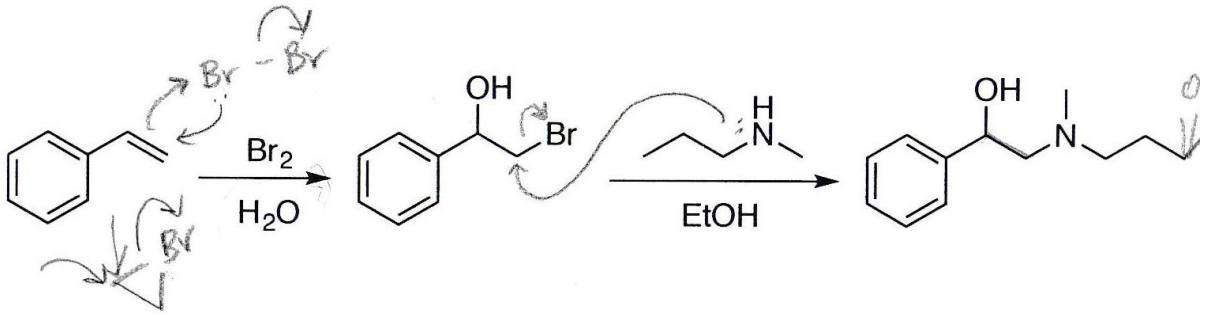


Figure 4-5. Andrea's work when describing how the β-bromo alcohol is formed in the first step of this pathway.

Andrea: …I think that this reaction only forms this product [β-bromo alcohol].

Interviewer: Do you know why?

Andrea: Because, it's um, because I know the mechanism.

Interviewer: Ok. Can you show me the mechanism?

Andrea: So, I'm pretty sure it's like...it adds uh...something like that [mechanism drawn is the bromination of the alkene in water]. If there wasn't...if...and then it would form like this kind of intermediate [drawn below]. If there's wasn't water, the second bromine would come in and add, but because there is water, the water adds to the more substituted position.

Interviewer: Ok. Why does water add to the more substituted position?

Andrea: Uh, because you're forming a temporary positive charge and it's more stabilized by hyperconjugation at the more substituted position.

Interviewer: Ok, where's our positive charge?

Andrea: When this....like it's not...it's like a delta positive, just very temporary in the transition state when this leaves [i.e., indicated C-Br bond breaks] and something comes in here [same C atom] I think.

Andrea's mechanism is abbreviated in some cases and contains errors (for example, the missing positive charge on bromine in the intermediate), but it is supported by a sufficient causal mechanistic explanation of why the indicated product is obtained. Conversely, Gretchen and Skyler are useful examples to illustrate how students may appear to use the EPF correctly, or adequately describe a mechanism in a verbal fashion, but are not using these strategies for their intended predictive power.

Gretchen

Gretchen employs the EPF to demonstrate that the alkyl bromide intermediate in pathway 3 would react with NaOH to form the indicated alkene, and that this reaction would proceed through an E2 mechanism (Figure 4-6). However, her explanation indicates that she is not using the EPF to rationalize this outcome; she has already developed this idea through the rule-based reasoning described in the previous section, and is just using the EPF to illustrate how this would happen.

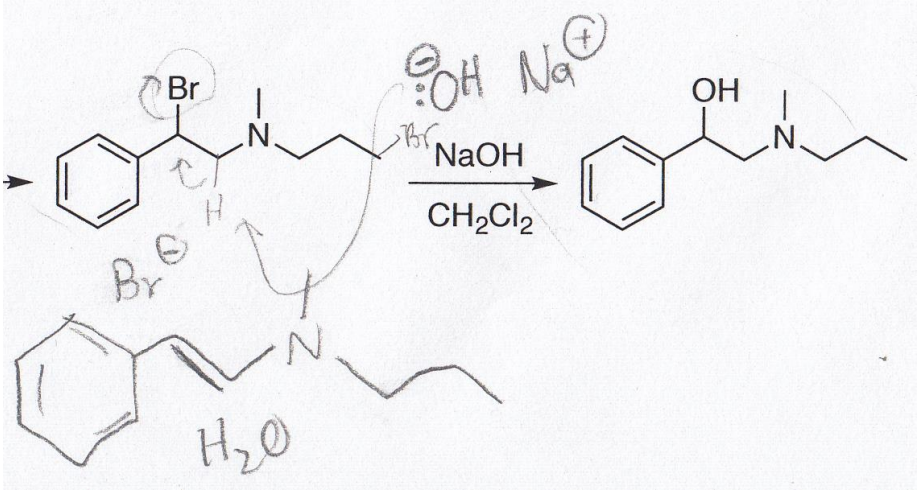


Figure 4-6. Gretchen's proposal for how the third step of pathway 3 might proceed by an elimination pathway rather than substitution.

> *Gretchen: Um… so based on how this is drawn, you can also do a substitution with the bromine, because it's also a good leaving group. But, it probably wouldn't be S_N2. It would have to be, like, a S_N1 type reaction, just because um like I don't... we learned in the lab especially that SN2 reactions don't really happen for uh, when there are secondary carbons, so like uh, I believe a E2 would be more common so this [hydroxide] would grab the hydrogen [beta hydrogen on alkyl bromide] instead and do that. So, it might not happen to do this [indicated substitution pathway].*
>
> *Interviewer: Ok, so what do you think the product would be instead?*
>
> *Gretchen: So, it would be...if that step would be happening with the hydrogen… it would be that [alkene in Figure 4-6]... it would just form a double bond.*

Skyler

Skyler chooses to describe how the epoxide-opening reaction in pathway 1 occurs using a mostly verbal explanation, although the last proton transfer step is described using the EPF (Figure 4-7).

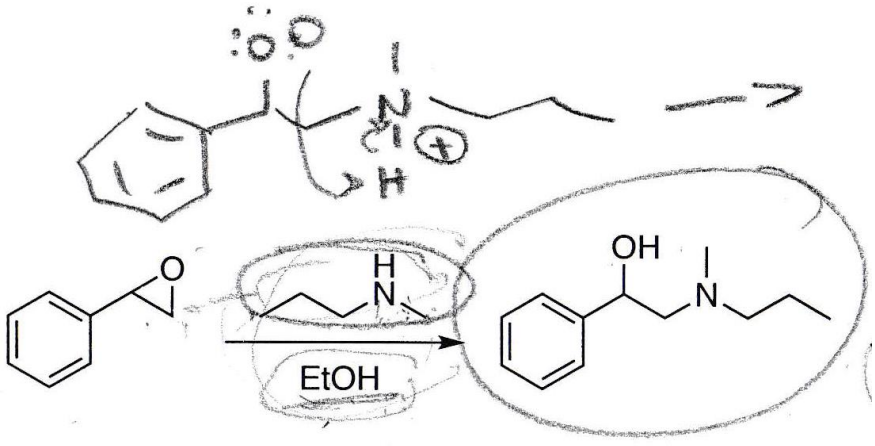


Figure 4-7. Skyler's work when describing the second step of pathway 1.

> *Interviewer: Ok, can you tell me - can you kind of walk me through that pathway and tell me what you like about it?*
>
> *Skyler: This one? [points to pathway 1]*
>
> *Interviewer: Yup.*
>
> *Skyler: I know for a fact that mCPBA makes epoxides.*
>
> *Interviewer: Yup.*
>
> *Skyler: And then, I would... since it's [the amine reagent in step 2 of pathway 1] a, since it's a, like a base, I would add that to the least substituted carbon, and then that would break the epoxide, and then you'd have like a negative oxygen, and then... um, I'm assuming, so you'd have something like this [draws out product of the amine reacting with the epoxide], and then... [continues drawing] right? And I would assume that the oxygen would take the hydrogen off of here [draws arrow from lone pair on oxygen to proton on nitrogen], and then you'd get this product [circles second intermediate in pathway 1], and I know for a fact that uh, PCC oxidizes alcohols to carbonyls.*

Although she has provided a reasonable description of how this mechanism could proceed, the statement "since it's a, like a base, I would add that to the least substituted carbon" indicates that she is still primarily invoking a memorized rule to explain the outcome of this reaction.

Step count was the main variable discussed by participants once they had completed rationalizing each pathway

Once participants chose a pathway, and were asked what features of each pathway they looked at to decide which was best, the most common feature that came up was the number of steps in each pathway; five of the eight participants identified the number of steps in a synthesis as an important determinant of its quality without being prompted to focus on this point. The most common reason participants gave for this was based on the theme of simplicity:

Jane

> *Interviewer: What other types of things did you look at when you were trying to decide which pathway is the best?*
>
> *Jane: Hmmm. I guess a lot of the time like simplicity. Like a lot of the time, the simpler, the better.*
>
> *Interviewer: Can you tell me what you mean by simple?*
>
> *Jane: Like, um... the steps make sense and you get the formation of the product the quickest, like, in the less amount of steps. Usually, that means that you're*

not doing a whole bunch of crazy stuff to try to get something when there is probably a simpler way to do it.

Andrea

Andrea: So, this reaction [third choice] is more complex than these [first two choices] that have four steps or I guess three steps. Um, so a less complicated reaction pathway is better than a more complicated one.

Interviewer: Can you tell me what you mean by complicated?

Andrea: Well, for– it's not practical... if you can do something in less steps, then it's always better to do that, and I mean if you're doing it in a lab it would be a waste of reagents too.

Skyler

Skyler discussed the point of step count without being prompted to do so, but while her reasoning behind the importance of step count was based on simplicity, she was not thinking about actually carrying out the steps in a laboratory; she believed extra steps would make it more likely for her to make a mistake on paper:

Interviewer: Is there anything else you think influenced your choice?

Skyler: Um, generally just like simplicity, and being familiar with the mechanisms [...]

Interviewer: Could you talk a bit more about what you mean by simplicity?

Skyler: Um, simplicity in regards to less steps, like less transitions. I feel like when there's less transitions then there's less room for error. [...]

Interviewer: Ok. Could you talk maybe specifically about what might go wrong if you had additional steps in a synthetic pathway?

Skyler: Um, yeah, you could uh, if you had additional steps you could forget a lone pair, you could forget to make a bond, you could choose the wrong - the wrong uh, carbon to make the bond to, um... you could - you have a higher chance of messing up the solvents, or using the wrong reagent, uh... yeah.

Gretchen

Gretchen's answer discussed the point that it would be "annoying" to have to carry out a synthesis with more steps, but she also pointed out that not all "steps" are equal:

Gretchen: Um, number of steps [...] this [third pathway] has way more. So, that would be like a much harder synthesis.

Interviewer: Ok, can you tell me what you mean by harder?

Gretchen: Uh, just having more steps sucks... knowing that from the lab, it's just really annoying to do more steps. Um, it takes a lot more time and I know that's

not exactly what we look for in the course necessarily. I mean... you should be taught to take less steps. Um, but I don't....if there is a more convenient synthesis and it takes more steps, I think it's better than doing like actually no, maybe not. It depends on the time.

Interviewer: Can you tell me more about that?

Gretchen: Um, because I haven't encountered an example, but it's like if you do four steps and maybe a step takes the full day, but you do like five steps and this takes a full day then might as well do all the five steps in a day.

Marie

Marie was not prompted to discuss step count directly, but was given the prompt below after spending the interview time up to this point evaluating the correctness of each pathway and concluding that "all of [the pathways] look good":

Interviewer: Okay. Do you think that there's any - anything that kind of sets one pathway away from the others? So let's say like, you're working in your lab and your supervisor gives you that starting material and asks you to make that product, um, which one of those pathways would you want to use? Because you can only pick one.

Marie: Hm. Um... not the last one, because it has more steps, so more can go wrong.

Interviewer: Can you tell me more about that?

Marie: Um... there's like more risk for messing up a step, or forming the wrong intermediate, um... [trails off]

Jesse

Jesse did not discuss the number of steps in each pathway independently, but when asked what his thoughts we, he agreed it was important, linking extra steps to the possibility of a lower overall synthetic yield:

Jesse: Uh... I feel like less steps would be easier, there's less chance for uh, side products - if the, if the steps are all uh, if there's, or, if the steps are easy things to control, like if it's easy to control, then only a single product is coming out of it, then less steps, I assume, would be better.

Interviewer: Ok. So less steps you think would be less side products kind of thing?

Jesse: Potentially, depending on what the steps are, but like, if you could choose three steps and - and there's say, at each step there's a ten percent, ten percent side product is made, see if - then you take four steps and a ten percent side product in each one, you're gonna have more of the final product when there's less steps.

Summary and Discussion

During the first interview task, participants were asked to evaluate three solutions to a synthetic problem, and choose which one they thought was best. Analysis of participants' work on this task revealed that these students' primary consideration in judging the quality of a synthesis is in the number of steps from the starting material to the product. When completing the task, students focused most of their efforts on judging the correctness of each synthesis pathway, often using memorized rules to do so.

Rule-based reasoning was most prevalent in steps that involved a nucleophilic substitution reaction, where multiple participants used a memorized decision-making flowchart to determine whether an elimination pathway would compete with the indicated substitution reaction outcome. For this specific problem, one could argue that the use of these rules constitutes a valid effort reduction strategy (heuristic), but if the goal of practitioners is to foster the development of skillful mechanistic reasoning, teaching students these algorithms may be counter-productive; if students can complete assessments using only heuristic reasoning, the instructor who designed that assessment has no indication of whether the student understands why their answer is correct (or incorrect).[30]

An important learning outcome emphasized in participants' organic chemistry course is the ability to use the electron pushing formalism (EPF) to rationalize and/or predict the outcome of a reaction. We found that only five out of eight participants used the EPF on this task. Of those five, the EPF was only employed once or twice, and when it was, it was generally not supported by a causal mechanistic explanation of how the reaction would proceed; rather, participants' explanations still seemed to invoke memorized rules, and the EPF was used simply employed in a way that would be consistent with the outcome indicated by that rule.

Participants discussed an important metric for evaluating a synthesis – the number of steps required to complete it. Some participants focused on this point in the context of synthetic economy, stating that a longer synthesis is likely to decrease the overall yield of the process, and that more steps would generally take more time to carry out (although as Gretchen astutely pointed out, this is not always the case). However, one participant (Skyler) discussed step count as important not because of synthetic economy, but because they felt that more steps would increase their likelihood of making an error on paper (for example, in drawing a mechanism, or forgetting a carbon atom when drawing a product).

While step count is linked to the concept of "synthetic ideality",[16] where one goal is to minimize the number of synthetic steps required to access a synthetic target, another quality of an ideal synthesis is atom and redox economy, where the goal is to avoid making any changes to a substrate (oxidations, reductions, protecting group chemistry) that will not be present in the target molecule. This was evidently not a factor in participants' evaluation of each pathway, as many preferred the second pathway, despite the first pathway having near perfect ideality.

Participants also did not focus on the environmental friendliness or "greenness" of these reactions; when asked during a follow-up prompt, participants had no opinion on the choice of solvents in some of these reactions. Participants were also asked to comment on the choice of a Jones oxidation in step 3 of the second pathway as opposed to the PCC oxidation used in the last step of the other pathways. Their only consideration was that since the substrate is a secondary alcohol, over-oxidation to a carboxylic acid would not be an issue, so either reagent would be fine in their opinions; no discussion of the environmental/health concerns of these reagents was initiated. Green chemistry is not a topic that is taught, practiced, or assessed in the participants' organic chemistry course, but this finding demonstrates that students are unlikely to develop any sense of familiarity with environmental considerations in the evaluation of a synthesis if they are not explicitly taught to make those considerations.

With respect to the factors participants used to decide which pathway was "best, these findings may suggest that students are only thinking about synthetic problems in the context of what they need to do "on paper" to solve the problem, and not making meaningful connections to the laboratory, where they would have to actually carry out the pathway they have proposed. While these considerations may be beyond the scope of lower-level Organic Chemistry courses, it highlights that students are unlikely to consider these factors on their own unless instructors explicitly highlight them as important.

Task 2: Propose a synthesis of a target molecule using given starting materials and any other desired reagents

Completeness and correctness: Participants' reasoning appeared to be bounded by pre-conscious cues, emblematic of serial associative cognition

Andrea

Andrea was the sole participant who was able to propose a complete synthetic pathway from the given starting materials to the target molecule; however, this pathway contained errors. Her approach is outlined herein.

Andrea starts by highlighting a phenyl ring in green, but then draws a mechanism for an E1cb reaction on the target molecule. She then tries the two aldol reactions below. She uses the electron-pushing formalism to represent what is happening, but in both cases, based on the electron-pushing arrows she has drawn, the products of these reactions are incorrect. (Figure 4-8)

> *Interviewer: Ok. Um, can you tell me how you saw this [arrows drawn on the product] here? (Figure 4-8)*
>
> *Andrea: Um, well, we talked in class yesterday about kind of like a system of thirds whenever there is an enolate reaction, there is like 1, 2, 3, but I knew that this double bond [C=C bond in target] could have been made by eliminating um an OH or like leave as a leaving group, so and that would mean that the double bond had to be here [carbonyl] and the enolate had to be here [pointed to area of lactone]. Mmm, so I guess I just recognize from practice problems.*

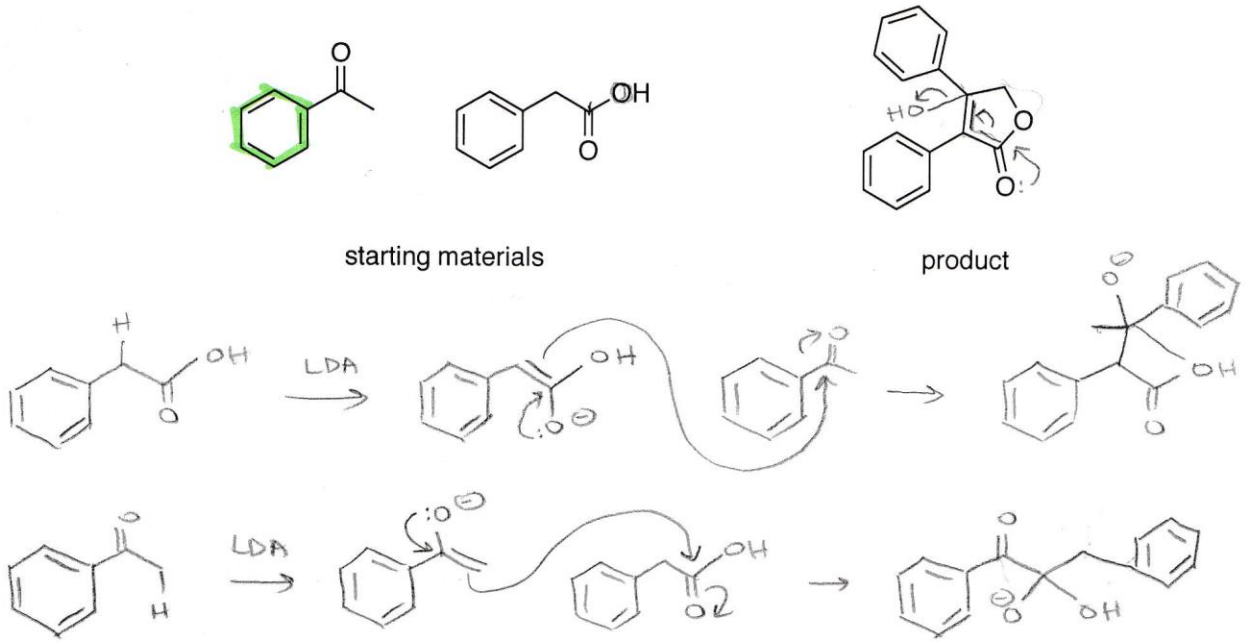


Figure 4-8. Andrea's initial analysis of the problem presented in Task 2.

Here, Andrea is referring to a common heuristic used in retrosynthetic analysis for disconnecting carbon-carbon bonds with a 1,3-difunctionalization pattern; a common strategy for accessing this functionality is through the aldol reaction, where in this case, a subsequent condensation reaction would yield the α,β-unsaturated lactone moiety present in the product. She then explains how she used this information to find the correct aldol reaction through trial and error:

> *Andrea: Ok. So, okay. So, first, um I guess hmmm, I thought it would be a simpler problem so I thought what would happen when you formed an enolate out of both of these [starting materials] and reacted with the other starting material.*
>
> *Interviewer: Ok.*
>
> *Andrea: But, none of them really looked like the exact configurations I needed to make this [lactone] ring […] So, then I decided to do a retrosynthesis problem instead to see how this is going.*

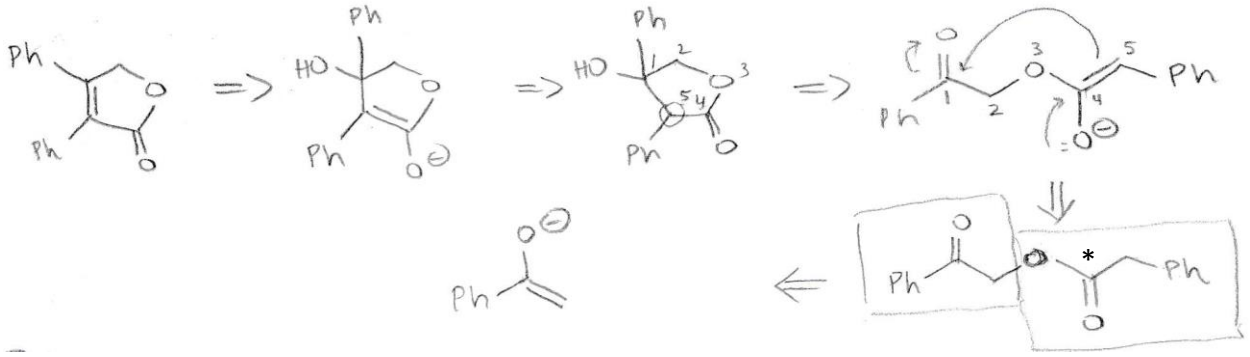


Figure 4-9. Andrea's retrosynthetic analysis of the target molecule in task 2.

While Andrea's retro-aldol disconnection (Figure 4-9) is correct, the subsequent disconnection she makes is incomplete, and the one synthon she draws is unsuitable for the disconnection she has made; it would serve better as an acceptor synthon. However, rather than completing her analysis and considering this alternative model, she considers a Baeyer–Villiger reaction for accessing the ester. This choice appears to be based on a mental model where acetophenone can act only as a nucleophile at the α-carbon (rather than making this carbon an electrophile), because she continues to discuss this given starting material only in the context of converting it to an enolate. She also believes a Baeyer–Villiger reaction would be suitable because the endocyclic oxygen atom in the target molecule does not appear to her to be present in the given starting materials, but she never maps atoms to determine if this is correct.

> *Andrea: Um, yeah so once I got here [step where she drew rectangles around her work], I started to recognize the pieces of the starting materials and except there is an extra O here and now what I'm considering is if it could be a Baeyer-Villiger reaction. Where...well like I haven't really been able to geometrically think through it, but I'm pretty sure that this oxygen is not the same as this oxygen.*
>
> *Interviewer: Can you tell me how you realized that?*
>
> *Andrea: Because an enolate would react here [C_1 of phenylacetic acid], so the bond would be formed with this carbon [w/asterisk above]. So, and it would never form here [pointed to OH portion of phenylacetic acid]... So that oxygen had to come from somewhere else.*
>
> *Interviewer: So, you're thinking about what reactions you would have to use and then seeing if the molecule fits in those reactions?*
>
> *Andrea: Mhm.*

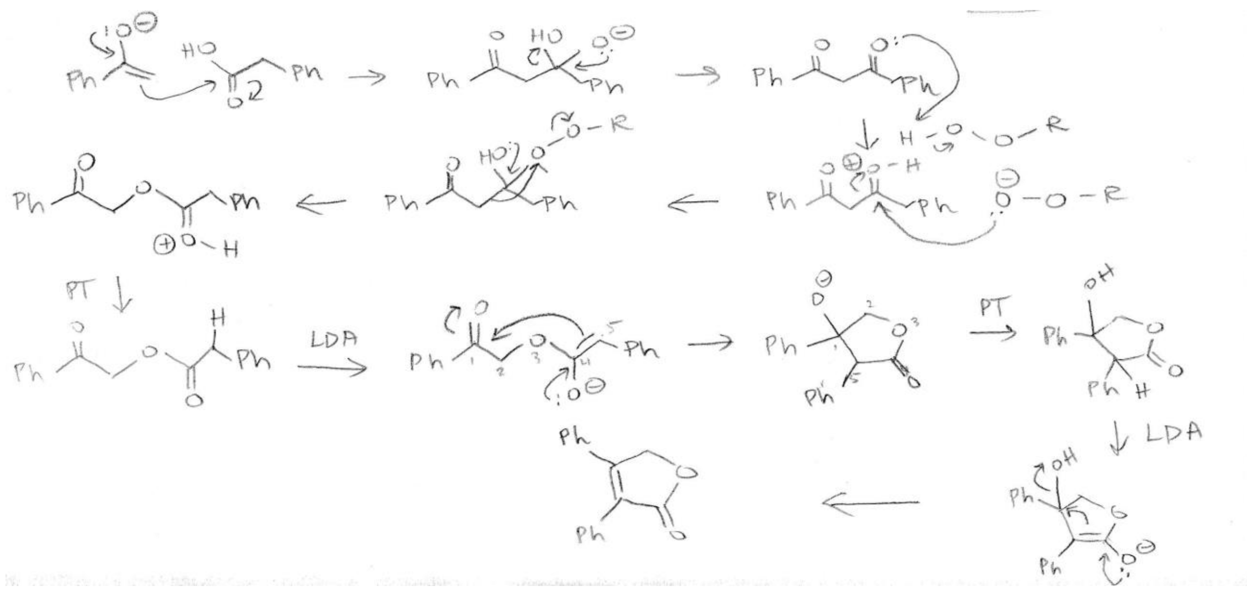


Figure 4-10. Andrea's solution to Task 2.

Andrea's solution is presented in Figure 4-10. It contains two major errors; first, the Claisen condensation she proposes as the first step is unlikely to work well with the acidic proton present on the electrophile. Second, the Baeyer–Villiger reaction she proposes as the second step would likely proceed with the benzyl moiety acting as the migratory group, leading to a different product than the indicated major product. These errors seem to arise from an approach that is influenced too heavily by features of the problem that cued her to the two key reactions she chose (1,3-disconnection = aldol reaction, ester = Baeyer–Villiger reaction), rather than designing a synthesis that is informed by a systematic analysis of bonds formed and broken.

Andrea was chosen as the primary exemplar of how reasoning shortcuts can lead to errors when they are based on an incomplete mental model, because she appeared to be the strongest organic chemistry student of the eight who participated in this study, and likely could have been successful had she engaged in a complete analysis of the problem. However, she was not the only example of a student who thought about the aldol reaction in this manner; three other participants (Jane, Jesse, and Gretchen) indicated that they thought about this disconnection in a similar fashion.

Jane

> *Interviewer: Ok. So, you say that you keep going back to an aldol because there are two carbonyls. Can you tell me why you make that connection?*

Jane: Because, I've seen a lot before... like problems where it's like that, and also... yeah, pretty much this pattern.

Interviewer: Is there a specific pattern you mean or just, like, seeing two carbonyls?

Jane: Just seeing two carbonyls, yeah.

Jesse

Interviewer: Is there a reason why you thought about that [aldol] reaction though?

Jesse: Uh... other than that the molecules, like these are the molecules you see doing [...] when I see these molecules, that's the type of reaction that jumps into my head, that's maybe what's happening.

Interviewer: Ok.

Jesse: Um, why it would happen at this [acetophenone] instead of this one [phenylacetic acid] […] I just feel like this, uh, we've done more, or at least in my head, I think, I think we've seen more with uh... with this molecule [acetophenone]…

Interviewer: Like that specific molecule [acetophenone]?

Jesse: This specific molecule.

Gretchen

Gretchen: I know that a double bond is formed so - and this [carbonyl on acetophenone] had to leave somehow so I was thinking of... what kind of reaction made that so I was thinking the... aldol reaction I think it is. [refers to reaction sheet] Yeah...yeah aldol reaction, and it's on the beta carbon that you form the double bond.

Interviewer: Can you tell me what made you think of the aldol reaction initially?

Gretchen: Initially, okay. Um, so... I - I just know that it's two um, carbonyls reacting with each other once the enolate's formed, so... and I saw the removal of this carbonyl group [in acetophenone] so that signified like, aldol reaction to me.

Many participants appeared to think that using this heuristic (identifying 1,3-difunctionalization patterns) for identifying the aldol reaction as a suitable step in their approach would be sufficient for solving the problem, and appeared to neglect any further analysis of the problem outside the application of this reaction. This is a valid approach to retrosynthetic analysis if used appropriately (other regiochemical patterns are considered, and patterns are re-identified following each subsequent retrosynthetic disconnection). However, in this situation, the approach was not used appropriately, the implication of which is that students are too strongly influenced by their initial impressions when analyzing a synthetic problem, and may

lack the metacognitive skillfulness required to decide to explore alternative models beyond the model that was initially cued by preconscious biases. A limitation in solving this problem of being too strongly influenced by heuristic reasoning is that students are unlikely to explore alternative possibilities when time is a limiting factor (for instance, on an exam). However, it may be beneficial to students if they are reminded in class or in other learning activities to reflect on how they are thinking about these types of problems.

Abstractness: Some participants exhibited familiarity with conventions of retrosynthetic analysis, but seemed unable to apply it successfully in all cases

The previous section outlines a participant (Andrea) who seems to understand how to use retrosynthetic analysis effectively, and was limited in her application of the strategy when her analysis became more strongly influenced by the 1,3-difunctionalization pattern to which she was initially cued. However, Andrea's fluency with this strategy went beyond what was explicitly taught in class and in the workshops, and most other participants seemed unable to extend their understanding of retrosynthetic analysis to actually applying the strategy to a novel problem. Holly and Skyler represent two examples of this phenomenon.

Holly

> *Interviewer: Ok. Um, let's say you're solving a problem and you don't have all the information you think you need. What do you do in those situations where all of the information isn't explicitly given to you?*
>
> *Holly: Um... not sure? Try to separate it out so... like, if I have the final separated and try to see what would need to have a negative charge, what would need to have a positive charge, I suppose?*
>
> *Interviewer: Ok, can you tell me a bit more about that?*
>
> *Holly: Um... split it into the fragments that I think it would need, and then... figure out which one would hold the negative charge or the positive charge, and then... either give it like a leaving group or protonate it?*
>
> *Interviewer: I'll give you the back of this sheet, maybe you can just like, illustrate what you're talking about?*
>
> *Holly: Like, there's just, bond here...then there's like an O, and then something [draws a C bonded to the O], not sure, and then if that would split [draws arrow going to the O], it'd be like an O negative and then... C positive I suppose...*
>
> *Interviewer: Ok. Can you tell me what you would do with that information?*
>
> *Holly: Well then the starting material would probably need to be HO, so it would need to be protonated, and then that [C] probably has a leaving group of some sort attached to it [draws X].*

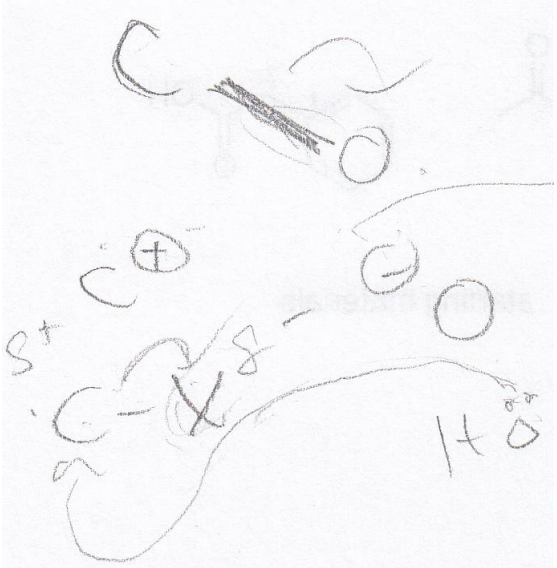


Figure 4-11. Holly's illustration of a retrosynthetic disconnection and the corresponding forward reaction.

Interviewer: Ok. Um... what would you use that information for?

Holly: Well then you know, if that [C-O structure] is the product, then you can go by what starting materials that kind of look like that. And then figure out what reaction, like, what reagent you would need to react them together to actually get... like, this [OH] to come in here and get that guy [X] to leave [draws S_N2-like mechanism].

Interviewer: Ok. Um... would you say as a - that's something you do anytime you solve a synthesis problem?

Holly: I should. I don't always, but I should. It helps think through part of it at least.

Skyler

Interviewer: Okay. Um... how do you decide where to start when you're solving a problem?

Skyler: Uh, in class we talked about like, making strategic cuts to try and find like... I guess that's more retrosynthesis, but like, two pieces that would go together, so I feel like for synthesis, it's trying to find where the two pieces fit. So that's what I try to do, I try to, I guess, going back to like, electrophile-nucleophile, like, see where they connect, and see what happened there.

Interviewer: Ok, um, when you're talking about making strategic cuts, could you tell me a bit more about what that might look like?

Skyler: Um so, you would break like, a carbon-halogen bond, you wouldn't cut a phenyl ring in half, cause that would be really difficult to put back together. Uh, basically like, where you think a possible addition happened. Like, a likely, simple addition would happen. Yeah.

Interviewer: ...so the retrosynthesis approach, like, you said that's like kind of intuitive to you?

Skyler: Mhm, like we talked about synthons in class, so I had a good understanding of what they were, but the problem that I had with solving the actual problem was that I didn't know which reagents to use, cause that wasn't really something they went over.

Interviewer: So like, you kind of know how to work backwards, but you're not really too sure how to work forwards again.

Skyler: Yeah, like I know how to break it, but I didn't know how to piece it back together properly.

Holly and Skyler's discussion of retrosynthetic analysis, which took place during a follow-up question, appears to demonstrate an understanding of how synthon-based retrosynthetic analysis should be carried out, but neither participant seems to know how the strategy could be effectively applied to a problem such as the one they worked on in Task 2. Neither Holly nor Skyler's work on paper contains nothing that would suggest they used any form of retrosynthetic analysis when approaching this problem.

Marie

Marie represents an interesting case because she used retrosynthetic analysis in her approach to both of the strategic bonds she highlighted in the lactone moiety of the target, but was only able to successfully complete the process for a carbon-heteroatom bond, (the example used in the retrosynthetic analysis workshop), and not for the carbon-carbon double bond. Her full approach is outlined herein.

She starts by mapping the atoms between the target molecule and starting materials, and uses that map to identify bonds formed between the starting materials (Figure 4-12). She then goes on to describe in detail how she mapped the atoms, and how that helped her see where bonds were formed:

Marie: this double bond O [lactone carbonyl in target] like, two carbons away, so that [phenylacetic acid] kind of matches up with this [boxed off portion of target molecule], and then this oxygen ["3"; points to it in target then in starting material], it's easy enough to remove a hydrogen, so that seems likely that this group [boxed off portion of target] is this [boxed off portion of phenylacetic acid]. Um, but then that means that there has to be um, a bond formed between this carbon [benzylic C in phenylacetic acid] and this carbon [carbonyl C in acetophenone], and then somehow, the oxygen [in acetophenone] just disappears. Uh... which makes sense, because in the starting materials, there's three oxygen atoms, and just two in the product…

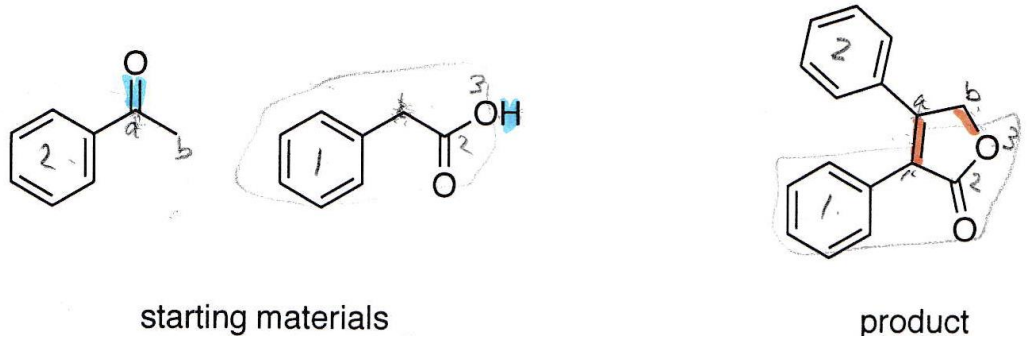


Figure 4-12. Marie's mapping and analysis of bonds formed (orange) and broken (blue).

She then engages in retrosynthetic analysis (Figure 4-13):

> *Interviewer: Can you explain to me what you're doing right now?*
>
> *Marie: Um... so I broke this bond [between oxygen and carbon "b" in redrawn product] um, and... trying to see how this oxygen [negative carbon of carboxylate group in retrosynthetic intermediate drawn] would react with this carbon [carbon b of same molecule; Br and H atoms not drawn yet], um, it's... easier to - to have, like, see oxygen as the nucleophile, and they usually have negative charges, um, and then I'm trying to figure out how that would react with the carbon. There should be something else on this carbon [carbon b]... um... [10 second pause] like a leaving group. Let's just put bromine for a minute, cause that's always a good leaving group [draws in Br atom].*
>
> *…*
>
> *Marie: Okay, now I have to figure out what reaction this would be [draws H atoms on carbon b with Br atom]. Maybe, um... [15 second pause] that seems likely that it would be... [draws in arrows on retrosynthetic intermediate] mm... S_N2 reaction.*

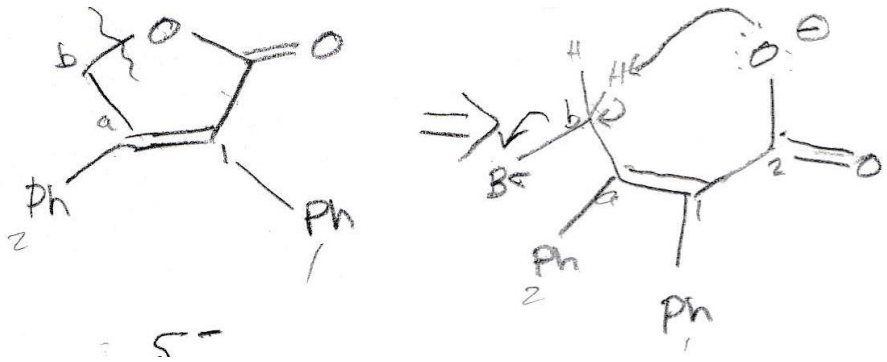


Figure 4-13. Marie's retrosynthetic analysis of the C_b–O bond in the target molecule.

While the mechanism she has drawn to represent an S_N2 reaction is incorrect, in the case of a carbon-oxygen bond, Marie was able to identify the polar bond, construct plausible

synthons, and identify suitable corresponding reagents. However, her approach to disconnecting the second strategic bond she identified was unsuccessful:

> *Interviewer: Can you tell me what you're thinking about right now?*
>
> *Marie: How this carbon [C_1 in phenylacetic acid] could [laughs] be attached to this carbon [C_a in acetophenone].*
>
> *…*
>
> *Marie: Um... this thing is partially negative here, and this carbon is partially positive, cause that's just how... ketones work. And, so that makes this an electrophile, and... this needs to be a nucleophile, which makes no sense, like how, um... that's the only part that I can't figure out, like okay, so that can be an electrophile, but to make this bond... [sighs] maybe I did my mapping wrong.*

Marie has correctly identified the atoms between which a C=C bond in the target is formed, and she correctly reasons that the benzylic carbon atom in acetophenone would be the electrophilic atom while the α-carbon of phenylacetic acid would be the nucleophilic atom. This highlights a limitation of Marie's mental model of retrosynthetic analysis in terms of abstractness; she seems to only be able to apply the strategy to types of bonds she has seen it applied to in the past, and cannot adapt what she knows to other situations. Another key difference in Marie's approach is that she does not explicitly construct synthons this time, making the completeness of her representation another limiting factor. It may be that since Marie has never explicitly been shown how to construct synthons in an aldol condensation-like disconnection, she does not know how to do so, but perhaps Marie could have solved the problem if she could have completed the process of synthon construction.

Summary and Discussion

In our analysis of task 2, we considered three primary factors that should be judged when evaluating a student's representation of their mental model about a given problem: correctness, completeness, and abstraction. Participants' use of mapping and retrosynthetic analysis illustrates all three of these qualities. Participants were often able to see that an aldol reaction could be applied to solving the problem, but often did not do further analysis to see exactly what that reaction would look like (what C atoms need to be connected), and in some cases, this observation was not abstracted enough from the example substrate used in class when the aldol reaction was first presented – it happened to be acetophenone, which was also a given starting material for this problem.

While Andrea was able to provide a solution to this challenging synthetic problem with only minimal errors, her analysis also contained errors, and otherwise, no other participants were able to provide a solution to the problem. However, this chapter explored research questions concerned with extremely high-level learning outcomes for second year students in organic chemistry; it is not uncommon for synthesis problems to be correctly answered by less than one-third of students on a final exam in this course.[18] The fact that most participants were unable to solve the synthetic problem indicate that perhaps it was too challenging, and future work investigating students' reasoning about synthesis problem solving may wish to provide alternative problems to students who are unable to solve a more challenging one. However, it is also possible that this problem was not beyond the scope of what would be reasonable for a second year organic chemistry student to be capable of solving if they did not use inappropriate reasoning shortcuts.

Participants' ability to propose a solution to this problem seemed to be limited both by the completeness of their analysis, where they neglected to fully explore ideas that were cued by heuristic strategies for identifying reasonable bond disconnections, and by their inability to abstract their knowledge of synthon-based retrosynthetic analysis from reactions they were already familiar with. Therefore, the implications of this study for learning is that instruction must foster metacognitive skillfulness, so students can effectively regulate their thoughts when engaging in problem solving, and that instruction must also foster the ability to engage in synthon-based retrosynthetic analysis that is based on a strong understanding of patterns of mechanisms, rather than attempting to fit learned reactions to a disconnection.

The organic chemistry curriculum at the University of Ottawa already has a strong emphasis on using the electron-pushing formalism in the context of unfamiliar reactions. Two key learning outcomes for the course are the ability to draw reaction arrows when given starting materials and products, or to draw reaction products given starting materials and arrows; in both cases, students are expected to be able to do this for reactions they have never seen before. To do so successfully, students are encouraged to use strategies such as mapping, a skill that is strongly emphasized in class, and was also a focus of the synthesis learning workshops described in Chapter 3. However, retrosynthetic analysis introduces an additional layer of complexity, where students must come up with their own retrosynthetic intermediates, and propose a plausible mechanism to account for their disconnection if they are likely to be able to propose a reasonable forward reaction.

Generally, participants discussed retrosynthetic analysis in a way that indicated understanding of the conventions for making polar bond disconnections, but could not apply the strategy fruitfully; often, their disconnections were implausible or contained mapping errors due to an incomplete use of this strategy. A specific example is Marie, who could not apply retrosynthetic analysis in an abstracted fashion; she could only make an S_N2-like disconnection, which was the example taught in the retrosynthetic analysis workshop. Although Marie correctly disconnected the C=C bond in the given target molecule, she did not know how to propose a forward reaction (aldol), likely because she was not yet familiar with how to apply this reaction in a synthetic setting.

Since participants' reasoning appeared to be emblematic of serial associative cognition – where their reasoning was bounded within a model that was pre-consciously cued by one key feature of the problem – a potential barrier for students to overcome with respect to constructing useful representations is that they may believe they have all the information they need without going through the complete process of mapping *all* atoms, or doing a *complete* retrosynthetic analysis.

Situating our findings in other domains of problem-solving

Currently, a gap exists in the literature regarding *novice* organic chemistry students' reasoning when approaching multi-step organic synthesis problems. Weinrich and Talanquer have recently investigated students' reasoning about synthesis pathways,[21] but the prompts used in this study provided students with a greater deal of information that would generally be available to them when solving these types of problems. However, research exists in other domains of problem-solving where parallels can be made to synthesis problem-solving. Two such examples include physics problem-solving, and the game of chess.

Physics

The key determinant of success in chemistry problem solving is the ability to construct representations that are complete, correct, and sufficiently abstracted from related examples of a given phenomenon so that the representation can be successfully applied to a relevant situation. Meanwhile, a critical component of physics problem solving is the ability to translate verbal statements into equations. This process could be viewed as analogous to the situation described for problem solving in chemistry; the process of seeing how mathematical relationships exist between actual physical phenomena requires some degree of abstraction.

Research into novices' reasoning while solving physics problems indicated that their thought processes seem to involve direct syntactic translation of verbal statements into equations; meanwhile, experts seem to generate abstract physical representations that account to identify relevant forces and energies.[31,32] The direct syntactic translation of verbal statements into equations is analogous to the rule-based reasoning we observed in this research; many of our participants' reasoning about reaction outcomes was based on memorized rules of reactivity rather than an electronic account of the situation. Meanwhile, the most successful participants' proposed synthetic pathways were informed by synthon-based retrosynthetic analyses.

Research in physics education has also found that students tend to think about solving problems separately from understanding the underlying concepts.[33] This finding may help to explain the outcomes of our research. Although our participants demonstrated knowledge of most of the reactions they had to reason about during the interviews, along with an understanding of the conventions behind strategies like retrosynthetic analysis, they still appeared to be unable to use problem-solving strategies in a way that allowed them to appropriately apply those reactions to synthesizing a target molecule. The main limitation here seemed to be the completeness, correctness, and abstractness of the representations they constructed when engaging in analysis of the problems, but future work could explore more directly the specific problem of how students think about the underlying concepts needed to solve a problem in the context of the strategies they use to solve that problem.

Chess

The decision-making processes of expert chess players have been described in four phases:[34]

1. Orientation to possibilities; looking at consequences of moves
2. Exploration; considering subsequent moves that would follow a primary move, depending on what the opponent does
3. Investigation; deeper, more serious search for possibilities
4. Final proof; the player uses the results of their search to rationalize their chosen move

Following a great deal of experience playing chess, one will develop an eye for strategic goals, and will have the ability to readily come up with a means for accomplishing them. One also

develops a set of standard procedures for certain positions (i.e., the pieces on the board in a certain layout), as they have likely been in many of these situations before.

This process is reasonably close to what a synthetic chemist must do when planning a synthesis of a complex target molecule; strategic goals are commonly applied in retrosynthetic analysis, as they greatly reduce the challenge of limiting the possible number of retrosynthetic transforms one might choose to make. Any key transform they identify has consequences, with respect to the other disconnections they must make to make that key transform feasible, and then starting materials must be obtained which allow them to set up an advanced intermediate that serves as a precursor to that key synthetic step. If those materials cannot be obtained, or follow-up reactions are not practical, their overall strategy must be revisited (these pitfalls are the “opponent” in synthetic planning!).

In our study, the most skilled participants seemed to orient themselves to possibilities using valid strategies. However, they did not seem to consider alternate possibilities once they had a synthetic step in mind, and their reasoning was often incomplete (it omitted information needed to arrive at a correct solution). Strategic goals in synthetic planning, such as structure-goal analysis and transform-based goal analysis were discussed in the workshop on retrosynthetic analysis. However, it is likely that more experience/practice is needed for students to become fluent in this type of analysis, which goes far beyond the identification of polar bonds in a target molecules and constructing the corresponding synthons; while this is a required skill, it does not necessarily help the problem solver reduce their search for useful disconnections (with respect to simplifying the target molecule).

Future work in synthetic problem solving could consider the parallels between how expert chess players plan their moves in the context of multiple possibilities, and potential issues that may arise following any choice they execute; indeed, the same considerations exist in synthetic planning.

Conclusions

In our analysis of task 1 of the think-aloud interviews, where participants were asked to evaluate three possible solutions to a synthetic problem, our primary findings were as follows:

(1) Most participants' discussion of their reasoning about reaction outcomes was based on memorized rules

(2) Participants often did not use the electron-pushing formalism in an expert-like fashion

(3) Step count was the main variable discussed by participants once they had completed rationalizing each pathway

These first two findings suggest that students' mental models for synthetic problem-solving may not be adequately developed to solve these problems in an expert-like fashion; rather than using the electron pushing formalism to explore reactivity and evaluate the likelihood that a given reaction would occur, many participants applied memorized rules for substitution and elimination reactions, and used mechanisms as a tool to support the validity of that memorized rule. Meanwhile, this third finding suggests that students may not be making meaningful connections between the laboratory and the classroom; while step count is a valid metric for evaluating the quality of a synthesis, many participants discussed this metric in the context of extra steps making them more likely to commit errors on paper, not in the laboratory.

Meanwhile, our analysis of task 2 considered the completeness, correctness, and abstractness of participants' representations of their mental models when they engaged in synthetic problem solving. We found that their representations had shortcomings in all three of these domains. Participants' mental models appeared to be incomplete, as they often started their analysis by centering in on one keying feature of the target molecule (the α,β-unsaturated lactone), and basing the rest of their analysis on the relationship between this functionality and an aldol reaction, without considering alternative possibilities. Participants generally seemed familiar with the conventions of retrosynthetic analysis, but often could not apply it to problem-solving; Marie is the best example of this, as she was able to make an S_N2 disconnection (the example given in the workshop), but she could not disconnect a bond associated with an unfamiliar reaction, indicating an inability to construct a mental model that was sufficiently abstracted from learning activities.

Implications for Teaching and Research

In order to help students become better at synthetic problem solving, we recommend that instructors encourage students to construct complete representations even when they think a partial representation will suffice, to check their use of key strategies for correctness, and to engage in the course material in a way that allows them to integrate their knowledge of the

course as a whole (so that they can construct representations that are abstracted from the few examples that can be covered in class). Future work might investigate specifically how this knowledge integration may be facilitated in the patterns of mechanisms organic chemistry curriculum used at the University of Ottawa. A research question of interest could be to investigate how the ability to organize one's knowledge of organic reactions by their governing mechanism may facilitate the construction of sufficiently abstracted retrosynthetic analyses.

In addition, we recommend that students are encouraged to consider their cognitive biases when engaging in problem solving, and to be systematic in their approach so that they have the opportunity to construct alternative models of the problem. This reflection is a part of being a skilled thinker, and while we did not observe any significant differences in metacognitive skillfulness between participants who provided successful/unsuccessful solutions on the workshop post-test, this specific aspect of metacognitive skillfulness may still be lacking amongst the participants in this study. Future work might investigate more thoroughly the relationship between metacognitive skillfulness and the decision to engage in cognitive decoupling when constructing mental models of a synthetic problem.

References

(1) Tversky, A.; Kahneman, D. Judgment Under Uncertainty: Heuristics and Biases. *Science* **1974**, *185* (4157), 1124–1131.

(2) Kahneman, D. *Thinking, Fast and Slow*; Farrar, Straus, and Giroux: New York, NY, 2011.

(3) Tversky, A.; Kahneman, D. Availability: a Heuristic for Judging Frequency and Probability. *Cogn. Psychol.* **1973**, *5* (2), 207–232.

(4) Strack, F.; Martin, L. L.; Schwarz, N. Priming and Communication: Social Determinants of Information Use in Judgments of Life Satisfaction. *Eur. J. Soc. Psychol.* **1988**, *18* (5), 429–442.

(5) Kolodner, J. L. An Introduction to Case-Based Reasoning. *Artif. Intell. Rev.* **1992**, *6* (1), 3–34.

(6) Kraft, A.; Strickland, A. M.; Bhattacharyya, G. Reasonable Reasoning: Multi-Variate Problem-Solving in Organic Chemistry. *Chem. Educ. Res. Pract.* **2010**, *11* (4), 281–292.

(7) Evans, J. S. B. T. Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annu. Rev. Psychol.* **2008**, *59* (1), 255–278.

(8) Evans, J. S. B. T. The Heuristic-Analytic Theory of Reasoning: Extension and Evaluation. *Psychon. Bull. Rev.* **2006**, *13* (3), 378–395.

(9) Bodner, G. M.; Gardner, D. E.; Briggs, M. W. Models and Modeling. In *Chemists Guide to Effective Teaching*; Pienta, N. J., Cooper, M. M., Greenbowe, T., Eds.; Upper Saddle River, NY, 2005; pp 67–76.

(10) Simon, H. A. *Models of Bounded Rationality*; MIT Press: Cambridge, MA, 1982.

(11) Evans, J. S. B. T.; Stanovich, K. E. Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspect. Psychol. Sci.* **2013**, *8* (3), 223–241.

(12) Stanovich, K. *Rationality and the Reflective Mind*; Oxford University Press, 2011.

(13) Polya, G. *How to Solve It*; Princeton University Press, 1945.

(14) Corey, E. J.; Cheng, X. M. *The Logic of Chemical Synthesis*; John Wiley & Sons: New York, NY, 1989.

(15) Nicolaou, K. C.; Vourloumis, D.; Winssinger, N.; Baran, P. S. The Art and Science of Total Synthesis at the Dawn of the Twenty-First Century. *Angew. Chem. Int. Ed.* **2000**, *39* (1), 44–122.

(16) Gaich, T.; Baran, P. S. Aiming for the Ideal Synthesis. *J. Org. Chem.* **2010**, *75* (14), 4657–4673.

(17) Novak, J. D. Human Constructivism: a Unification of Psychological and Epistemological Phenomena in Meaning Making. *International Journal of Personal Construct Psychology* **1993**, *6* (2), 167–193.

(18) Bodé, N. E.; Flynn, A. B. Strategies of Successful Synthesis Solutions: Mapping, Mechanisms, and More. *J. Chem. Educ.* **2016**, *93* (4), 593–604.

(19) Domin, D.; Bodner, G. Using Students' Representations Constructed During Problem Solving to Infer Conceptual Understanding. *J. Chem. Educ.* **2012**, *89* (7), 837–843.

(20) Sevian, H.; Bernholt, S.; Szteinberg, G. A.; Auguste, S.; Pérez, L. C. Use of Representation Mapping to Capture Abstraction in Problem Solving in Different Courses in Chemistry. *Chem. Educ. Res. Pract.* **2015**, *16* (3), 429–446.

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

(22) Bowen, C. W. Think-Aloud Methods in Chemistry Education: Understanding Student Thinking. *J. Chem. Educ.* **1994**, *71* (3), 184.

(23) Willis, G. B.; National Center for Health Statistics (U.S.). Cognitive Methods Staff. *Cognitive Interviewing and Questionnaire Design*; 1994.

(24) Randles, C. A.; Overton, T. L. Expert vs. Novice: Approaches Used by Chemists When Solving Open-Ended Problems. *Chem. Educ. Res. Pract.* **2015**, *16* (4), 811–823.

(25) Flynn, A. B. How Do Students Work Through Organic Synthesis Learning Activities? *Chem. Educ. Res. Pract.* **2014**, *15* (4), 747–762.

(26) Creswell, J. W.; Miller, D. L. Determining Validity in Qualitative Inquiry. *Theory Pract.* **2000**, *39* (3), 124–130.

(27) Lincoln, Y. S.; Guba, E. G. *Naturalistic Inquiry*; SAGE Publications: Newbury Park, CA, 1985.

(28) Gigerenzer, G. On Narrow Norms and Vague Heuristics: a Reply to Kahneman and Tversky. *Psych. Rev.* **1996**, *103* (3), 592–596.

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

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

(31) Larkin, J.; McDermott, J.; Simon, D. P.; Simon, H. A. Expert and novice performance in solving physics problems. *Science* **1980**, *208 (4450),* 1335–1342.

(32) Chi, M. T. H.; Feltovich, P. J.; Glaser, R. Categorization and representation of physics problems by experts and novices. *Cogn. Sci.* **1981**, *5 (*2), 121–152.

(33) Sabella, M. S.; Redish, E. F. Knowledge organization and activation in physics problem solving. *Am. J. Phys.* **2007**, *75 (11),* 1017–1029.

(34) de Groot, A. D. *Thought and Choice in Chess.* Mouton; The Hague, NED, 1978.

Chapter 5 Conclusions

This thesis has outlined three separate studies in chemical education research. First, the research outlined in Chapter 2 served as a useful evaluation of the learning modules on Nomenclature101.com. This study was able to demonstrate that Nomenclature101.com is effective in helping students achieve significant gains in organic nomenclature skill after just 90 minutes of use, with no significant difference between participants who used the website and participants who completed a guided tutorial on organic nomenclature. We were also able to obtain evidence that students find this website useful and easy to use; the data that supports this finding were validated using exploratory factor analysis, which provided statistical evidence for each of these latent factors. This study also demonstrated that a workshop-style intervention could help students make significant learning gains on an individual topic (albeit, a relatively simple one) in the organic chemistry curriculum, and the findings encouraged us to investigate if similar findings could be obtained from a workshop on a more challenging topic that we wished to know more about (students' ability to solve synthetic problems).

The research described in Chapters 3 and 4 of this thesis was based on evidence supporting that successful solutions to synthetic problems tend to contain strategies that are used either incorrectly or not at all in unsuccessful solutions to those problems. Having also obtained evidence that a one-time workshop-style intervention could lead to significant learning gains in organic nomenclature skill, we wanted to determine if the same format would be amenable to this more advanced set of skills, some of which were already addressed in class for solving mechanistic problems.

The primary findings of the research outlined in Chapter 3 of this thesis were that a single intervention such as the one we used for helping students to learn organic nomenclature is not likely to be sufficient in leading to learning gains with respect to synthetic problem solving. Analysis of participants' post-test responses generally showed little evidence of key strategies being used correctly, which may have also influenced the lack of observed learning gains – participants' representations are incomplete or incorrect.

The results of this study also indicated that metacognitive skillfulness (as measured by the Metacognitive Activities Inventory) was not a factor in predicting whether participants were successful or not in solving synthetic problems. Participants' general attitudes toward the subject of chemistry also did not appear to have an influence on successful problem solving, but

we found that successful problem solvers tended to rate chemistry as a challenging subject despite their apparent skill in the subject. Future work might investigate this phenomenon in the context of a self-efficacy framework.

Chapter 4 investigated in more detail how Organic Chemistry 2 students think about synthesis problems as they are solving them. The participants generally used memorized rules to evaluate solutions to synthesis problems, and although they used the electron-pushing formalism as a tool, they did not use it in the same way experts do; rather than using it to make predictions about reaction outcomes, they essentially used it to illustrate their thoughts about the reactions, often doing so with errors.

Meanwhile, most participants (all except Andrea) were unable to provide a solution to the synthesis problem they were given. In some cases, participants did attempt to use expert-like strategies (retrosynthetic analysis), but appeared to be unable to apply this strategy to bond disconnections for which they were not already familiar with the forward reaction. Furthermore, participants often appeared to make decisions in their analysis that were based on an initial feature of the problem that they noticed (*e.g.*, the α,β-unsaturated lactone). Many participants were cued by this functional group pattern to consider an aldol reaction in their synthetic approach, but seemed unable to consider one of the given starting materials (acetophenone) in any other context than this reaction.

Based on our findings, we suggest that if instructors wish to include the ability to solve multistep organic synthesis problems as a learning outcome for their courses, they do so in a way that is aligned with learning outcomes related to mechanistic reasoning, so that mechanisms and synthesis are not compartmentalized as two separate learning outcomes by students. Additionally, we suggest that future research on synthetic problem solving at the undergraduate level in chemistry education research should investigate how this approach influences students' overall understanding of the course material.

Lastly, it is worth noting that most students taking Organic Chemistry 1 and 2 will not go on to become practicing synthetic organic chemists, and one might argue that these courses should best serve these students by providing them with a mechanistic understanding of organic reactivity that is transferrable to their other courses in chemistry and biomolecular sciences. However, we believe that the assessment of synthetic problem solving can play an important role in introductory organic chemistry courses, as long as associated learning activities foster

mechanistic reasoning, and teach students to make retrosynthetic transforms that are based in patterns of reactivity rather than memorized reactions.