This article can be cited before page numbers have been issued, to do this please use: A. B. Flynn and R. Featherstone, Chem. Educ. Res. Pract., 2016, DOI: 10.1039/C6RP00126B.

This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal’s standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.
Language of mechanisms: exam analysis reveals students' strengths, strategies, and errors when using the electron-pushing formalism (curved arrows) in new reactions

Alison B. Flynn* and Ryan B. Featherstone

Department of Chemistry and Biomolecular Sciences, University of Ottawa, 10 Marie Curie Private, Ottawa, ON, Canada, K1N 6N5

*alison.flynn@uOttawa.ca

Abstract

This study investigated students' successes, strategies, and common errors in their answers to questions that involved the electron-pushing (curved arrow) formalism (EPF), part of organic chemistry's language. We analyzed students' answers to two question types on midterms and final exams: (1) Draw the electron-pushing arrows of a reaction step, given the starting materials and products; and (2) Draw the products of a reaction step, given the starting materials and electron-pushing arrows. For both question types, students were given unfamiliar reactions. The goal was for students to gain proficiency—or fluency—using and interpreting the EPF. By first becoming fluent, students should have lower cognitive load demands when learning subsequent concepts and reactions, positioning them to learn more deeply. Students did not typically draw reversed or illogical arrows, but there were many other error types. Scores on Arrows questions were significantly higher than on Products questions. Four factors correlated with lower question scores, including: compounds bearing implicit atoms, intramolecular reactions, assessment year, and the conformation of reactants drawn on the page. We found little evidence of analysis strategies such as expanding or mapping structures. We also found a new error type that we describe as picking up electrons and setting them down on a different atom. These errors revealed the difficulties that arose even before the students had to consider the chemical meaning and implications of the reactions. Herein, we describe our complete findings and suggestions for instruction, including videos that we created to teach the EPF.

Introduction

Reaction mechanisms are a central aspect of study in chemistry that are commonly explored and explained in organic chemistry using multiple short hands and an electron-pushing (curved arrow) formalism (EPF)—a highly meaningful symbolism (Bhattacharyya, 2013). Although the EPF is a core part of organic chemistry's culture (Bhattacharyya & Bodner, 2014), students struggle to interpret reaction mechanisms and use the EPF. Graduate student participants in one study drew arrows that got them to the products, despite not being representative of the most likely mechanism (Bhattacharyya & Bodner, 2005). In another study, "students provided the mechanism and curved arrows only after having predicted the product" (p. 848), indicating that the EPF was not useful to them in
their problem-solving process as it would be to an expert (Grove, Cooper, & Rush, 2012). To analyze reactions (e.g., identify reagents or predict products), students seem to use a combination of heuristics and product analysis (e.g., stability) rather than the feasibility of a mechanism (Graulich, 2015a; Rushton, Hardy, Gwaltney, & Lewis, 2008). Further, students struggle to connect structure and function (Anzovino & Bretz, 2015; Cartrette & Mayo, 2011; Cruz & Towns, 2014; DeFever, Bruce, & Bhattacharyya, 2015). Despite all these difficulties, there seems to be benefit for students to use mechanistic thinking, especially for more difficult questions for which the answer is not immediately obvious to students (Grove, Cooper, & Cox, 2012), suggesting that instruction should be improved to help students develop expertise earlier in their studies. The goal of our study was to isolate and investigate students’ success and errors when using and interpreting the EPF, independently of their ability to explain and predict reaction mechanisms and outcomes. We believe that being able to use and interpret the EPF is a key step toward understanding the molecular level of reactions, that is, the invisible. We explain our reasoning in the following paragraphs.

Making sense of the invisible is a major challenge that students face in learning chemistry (Kozma & Russell, 1997). The objects and changes we see at the macroscopic level reflect the submicroscopic level (Johnstone, 1982, 2000); these levels are communicated through and represented with the third, symbolic level (Taber, 2013). This third level forms chemistry’s language—its many symbols, representations, and tools (Taber, 2013; Talanquer, 2011) that serve as chemistry’s words, grammar, and syntax (Taskin & Bernholt, 2014) (Figure 1). This language is deeply meaningful to experts but little more than a collection of lines and dots to many students (Bodner & Domin, 2000). Experts readily interpret, construct, and switch between representations and levels to best represent a desired molecule or process (Gilbert, Reiner, & Nakhleh, 2008). However, this switching is often difficult and problematic for learners, who do not always know which aspects of symbolic representations are significant, much less how to translate between them (Cheng & Gilbert, 2009; Gilbert & Treagust, 2009; Johnstone, 2000; Kozma & Russell, 1997; Strickland, Kraft, & Bhattacharyya, 2010). This difficulty developing representational competence transcends STEM fields (LaDue, Libarkin, & Thomas, 2015), including physics (Pape & Tchoshanov, 2001), math (Pollack, 2012), biology (Dees, Momsen, Niemi, & Montplaisir, 2014; LaDue et al., 2015; Wright, Fisk, & Newman, 2014), and biochemistry (Linenberger & Bretz, 2014). Not understanding the language makes it difficult to learn the concepts and easy to misinterpret the message; and yet, most curricula expect students to simultaneously learn and use this new language in complex situations (Taber, 2009).

![Figure 1. Chemistry's triplet (Johnstone, 2000). Note: here the symbolic domain for Tylenol includes symbols (e.g., line structures), and representations of macroscopic (e.g., image of a bottle) and submicroscopic (e.g., electrostatic potential map) levels.](image-url)
The complex language of organic chemistry contributes to its reputation as a difficult and mysterious subject, a gauntlet to be run rather than the pathway to an exciting field (Anderson & Bodner, 2008; Lewington, 2013; Moran, 2013). Taber (2009) stressed that students needed to understand the meaning of curved arrows and argued that “teachers need to do more to induct learners into the intended symbolism we use” (p. 84). Bhattacharyya and Bodner (2005) emphasized that: “First, students need to be more explicitly aware that arrow-pushing serves an explicatory function” (p. 1407).

To help students gain fluency in organic chemistry’s language, the University of Ottawa’s curriculum explicitly teaches students the language before teaching chemical reactions and principles (Flynn & Ogilvie, 2015). This language instruction is anchored in students’ prior knowledge—such as drawing Lewis structures—and has four major learning outcomes (Figure 2): (LO1) Draw the electron-pushing arrows for a step, given the starting materials and products for that step; (LO2) Draw the products of a reaction step, given the starting materials and electron-pushing arrows for that step; (LO3) Draw the transition state structure for a given reaction step (understanding that some of the mechanistic details and conformational information may be absent); and (LO4) Draw a mechanism for the reverse reaction, given the mechanism in the forward direction. These question types always use reactions that students have not learned, to emphasize using and interpreting the EPF and not relying on knowing the mechanism.

![Figure 2. Organic chemistry language learning outcomes (LOs). Reactions are new to the students, such as this one (Clavette et al., 2012).](image-url)

**Research questions**

In this study, we explored Organic Chemistry I and II students’ successes, strategies, and the nature of their errors related to the first two learning outcomes, guided by the following research questions:

RQ1) What are students’ success rates in solving organic chemistry’s language questions that involve using and interpreting the electron-pushing formalism?
RQ2) What types of strategies and errors are students using and making in those question types?

**Conceptual context: information processing theory and meaningful learning**

Information processing theory (IPT) and meaningful learning form the conceptual context of this project (Figure 3) (Mayer, 2012; Schunk, 2016). According to IPT, information enters through the sensory registers. Inputs perceived as useful are attended to and transferred to working memory (WM), which is responsible for holding, rehearsing, and processing new and already stored information. That information may be encoded for storage in long-term memory (LTM), which is facilitated through organization, elaboration, and links with networks—hallmarks of meaningful learning. For learning to be meaningful, the learner: (1) needs relevant prior knowledge upon which to anchor the new knowledge, (2) the new information must be meaningful in and of itself, and (3) the learner has to choose to incorporate this new information with existing knowledge (Ausubel 1968; Bretz, 2001; Novak, 2007).

Also integral throughout IPT are social interactions and the learner’s characteristics, including decisions, self-regulation, culture, and affect (Mayer, 2012). As the learner transfers information from WM to LTM, they are actively constructing their own knowledge (Vygotsky, 1978). Students’ perception and use of symbols and representations will depend on their associations to existing knowledge, the meaning they attribute to the language, the associated cognitive load, and other factors (Matlin, 2009). In principle, if students are “fluent” in chemistry’s language, they should have lower cognitive load demands and will be positioned to more deeply analyze subsequent reactions.

By analyzing students’ success rates and errors, we will begin to learn what features of structures they attend to (and do not attend to) and how they interpret organic chemistry representations.

![Figure 3. Information Processing Theory (IPT) and meaningful learning form the project’s conceptual context.](image-url)
Methods

Setting and Course

This research was conducted in Organic Chemistry I and II courses taught at a large, research-intensive Canadian university; each course is one semester in length. Organic Chemistry I is offered in the winter semester of students’ first year of their studies and Organic Chemistry II is offered in the fall and summer semesters. Students can take these courses in English or French.

Both courses are twelve weeks long and consist of two weekly classes (1.5 hours each, mandatory, flipped format) (Flynn, 2015a) and a tutorial session (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 (‘Top Hat’, 2016). 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).

The course is 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 that has been implemented in all Organic Chemistry sections (six sections of Organic I, 3 sections of Organic II) (Flynn & Ogilvie, 2015).

The organic chemistry language instruction was given in Organic Chemistry I, immediately after a module on structure and before the first reactivity module on acid–base chemistry. Two of the six course sections have piloted the chemistry language instruction. Every assessment in Organic I and II contained questions aligned with the aforementioned language learning outcomes, always using a reaction that students’ had never learned at that point in their studies, so that the focus was on interpreting the symbolism rather than recalling a reaction type. The timing of the questions is summarized in Table 1. The question (e.g., P16_2_399) gives the question type (P = draw the products), question label (16), the total points possible on that question (2) and the number of answers that we analyzed for that question (N = 399).
Table 1. Summary of questions asked in Organic Chemistry I or II.

<table>
<thead>
<tr>
<th>Question</th>
<th>Question type</th>
<th>Course (Organic Chemistry I or II)</th>
<th>Assessment type</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1_8_221</td>
<td>Arrows</td>
<td>II</td>
<td>Exam</td>
<td>2014</td>
</tr>
<tr>
<td>A2_8_388</td>
<td>Arrows</td>
<td>II</td>
<td>Exam</td>
<td>2015</td>
</tr>
<tr>
<td>A4_5_42</td>
<td>Arrows</td>
<td>I</td>
<td>M1</td>
<td>2015</td>
</tr>
<tr>
<td>A5_11_121</td>
<td>Arrows</td>
<td>II</td>
<td>M2</td>
<td>2014</td>
</tr>
<tr>
<td>A6_6_344</td>
<td>Arrows</td>
<td>I</td>
<td>Exam</td>
<td>2015</td>
</tr>
<tr>
<td>A8_8_372</td>
<td>Arrows</td>
<td>I</td>
<td>M1</td>
<td>2015</td>
</tr>
<tr>
<td>A11_8_290</td>
<td>Arrows</td>
<td>I</td>
<td>M2</td>
<td>2015</td>
</tr>
<tr>
<td>A12_4_20</td>
<td>Arrows</td>
<td>I</td>
<td>Exam</td>
<td>2014</td>
</tr>
<tr>
<td>A13_4_344</td>
<td>Arrows</td>
<td>I</td>
<td>Exam</td>
<td>2015</td>
</tr>
<tr>
<td>A14_7_95</td>
<td>Arrows</td>
<td>II</td>
<td>M1</td>
<td>2015</td>
</tr>
<tr>
<td>A15_7_127</td>
<td>Arrows</td>
<td>II</td>
<td>M2</td>
<td>2015</td>
</tr>
<tr>
<td>P1_2_361</td>
<td>Products</td>
<td>I</td>
<td>M1</td>
<td>2015</td>
</tr>
<tr>
<td>P2_2_126</td>
<td>Products</td>
<td>II</td>
<td>M2</td>
<td>2014</td>
</tr>
<tr>
<td>P3_2_200</td>
<td>Products</td>
<td>I</td>
<td>Exam</td>
<td>2014</td>
</tr>
<tr>
<td>P4_2_344</td>
<td>Products</td>
<td>I</td>
<td>Exam</td>
<td>2015</td>
</tr>
<tr>
<td>P5_2_200</td>
<td>Products</td>
<td>I</td>
<td>Exam</td>
<td>2014</td>
</tr>
<tr>
<td>P6_2_345</td>
<td>Products</td>
<td>I</td>
<td>Exam</td>
<td>2015</td>
</tr>
<tr>
<td>P7_2_98</td>
<td>Products</td>
<td>II</td>
<td>M1</td>
<td>2015</td>
</tr>
<tr>
<td>P8_2_200</td>
<td>Products</td>
<td>I</td>
<td>Exam</td>
<td>2014</td>
</tr>
<tr>
<td>P9_2_344</td>
<td>Products</td>
<td>I</td>
<td>Exam</td>
<td>2015</td>
</tr>
<tr>
<td>P10_2_344</td>
<td>Products</td>
<td>I</td>
<td>M1</td>
<td>2015</td>
</tr>
<tr>
<td>P11_2_126</td>
<td>Products</td>
<td>II</td>
<td>M2</td>
<td>2014</td>
</tr>
<tr>
<td>P12_2_290</td>
<td>Products</td>
<td>I</td>
<td>M2</td>
<td>2015</td>
</tr>
<tr>
<td>P13_2_290</td>
<td>Products</td>
<td>I</td>
<td>M2</td>
<td>2015</td>
</tr>
<tr>
<td>P14_2_290</td>
<td>Products</td>
<td>I</td>
<td>M2</td>
<td>2015</td>
</tr>
<tr>
<td>P16_2_399</td>
<td>Products</td>
<td>II</td>
<td>M1</td>
<td>2015</td>
</tr>
<tr>
<td>P17_3_16</td>
<td>Products</td>
<td>II</td>
<td>M2</td>
<td>2015</td>
</tr>
</tbody>
</table>

1 The question (e.g., P16_2_399) gives the question type (P = draw the products), label (16), the total points possible on that question (2) and the number of answers that we analyzed for that question (N = 399).

This study focused only on the first two learning outcomes, abbreviated (1) “Arrows” or “A” series questions (Figure 4) and (2) “Products” or “P” series questions (Figure 5). The compounds were all drawn as if they were flat, to minimize complexity in the drawings and focus on the formalism; orbitals were not included in these question types (stereochemistry and molecular orbitals are included in other question types with other reactions and mechanisms in the course). In total, 15 Arrows and 17 Products questions were asked, giving a total of 7411 answers to analyze (3114 Arrows, 4297 Products). The University of Ottawa’s Office of Research Ethics and Integrity approved this secondary use of data for this study, in accordance with policies for the ethical treatment of human....
research participants (Canadian Institutes of Health Research and Social Sciences and Humanities Research Council of Canada, 2010; Taber, 2014).

Figure 4. “Draw the arrows” questions used in this study (labeled “A”). Students were given the starting materials and products of each reaction step. They had to expand the structure to reveal any non-bonding electrons and bonds involved then draw the electron-pushing arrows. The anticipated answers are drawn in red. Some questions were given on more than one midterm/exam; each instance was given its own label. For example, A4 – A7 refer to the same question given on different examinations (see Table 1).
Figure 5. Draw the products questions used in this study (labeled “P”). Students were given the starting material(s) and electron-pushing arrows and had to draw the product of that step. The anticipated answers are drawn in green.

**Question analysis**

First, the score on each step of the questions (graded by a course teaching assistant using a marking scheme) was recorded on paper; the researcher who entered the scores into the database file verified that the scores were assigned according to the marking scheme (checking for grading errors and consistency) by comparing the assigned score to the same marking scheme.

Next, each question was coded to identify the common strategies and the common errors demonstrated in the answer, similar to a process used when analyzing students’ answers to synthesis questions (Bodé & Flynn, 2016). For the language questions, four different researchers analyzed eight different questions using an open coding method. The coding schemes were compared and a systematic coding template was developed (Appendix 1). All questions were analyzed or re-analyzed using the coding template by one of the researchers. A second rater coded a subset of answers (at least 10%) for a number of questions and inter-rater reliability scores were calculated for each question, resulting in a Krippendorf’s $\alpha \geq .92$ in all cases (Krippendorf, 2011).

The distribution of data was significantly non-normal in all cases. For example, a histogram of answer scores for question A1 (Figure 6) shows negatively skewed data (skewness = -1.5), a Kolmogorov-Smirnov test was significant ($p < .001$), and Q-Q plot showed deviations from normality. As such, non-parametric statistical tests were used to conduct all the analyses.
Results and discussion

Students scored higher on Arrows questions than on Products questions
As we investigated RQ1, we expected scores on both question types (which had a single correct answer) to be higher than the corresponding assessment score, because all the necessary information was given to students to solve the problem and they only had to use or interpret the electron-pushing formalism. Because students were given all the information needed to solve the questions, we think the questions are relatively easy—i.e., at the “understand” level in Bloom’s taxonomy (Krathwohl, 2002)—and they had the prior knowledge and ability to get to the correct answer (e.g., when explicitly asked to do so, students are very successful translating between line and Lewis structures or assigning a formal charge). In other words, we expected these scores to be nearly 100% in all cases. Instead, we found that students’ scores were lower than expected for both question types, on average.

The average and median scores on Arrows questions were higher than on Products questions: 72% and 86% for Arrows questions (N = 2926, SD = 33%), respectively, versus 55% and 50% for Products questions (N = 3993, SD = 40%), \( U = 4,628,100, z = -15.4, p < .001, r = -.18 \) (Figure 7). Each point represents one student’s averaged score within one course; a student who took two courses (e.g., Organic I and II or repeated one of the courses) would appear as a separate point for each course. There is less variance between Products scores because these questions are typically scored out of 2 or 3 whereas Arrows questions are scored out of larger values (e.g., 8). There was low correlation between students’ average question scores and their corresponding midterm or exam scores, \( R^2 = 0.38 \) for Arrows questions and \( R^2 = 0.14 \) for Products questions.

Figure 6. Histogram of scores for question A1 (%) shows negatively skewed data.
Figure 7. Average score on Arrows (left, red triangles) and Products (right, green squares) questions versus the corresponding assessment score (midterm or exam). Each point represents one student’s averaged scores.

Students scored higher on some questions than on others

We analyzed the distribution of responses to each question using boxplots (Figure 8), histograms (not shown), and descriptive statistics (not shown). Specific question attributes were correlated with lower scores on both question types: implicit atoms, intramolecular reactions, year in which the question was asked, and conformation of the reactants drawn on the page. These features in the questions reveal barriers that exist before students even get to thinking about reactive sites in the molecules or other aspects of chemical reactions (e.g., stereochemistry, orbital overlap).

Figure 8. Distribution of scores for each question analyzed. x-axis label (e.g., A1_8_221) represents the question type (A = Arrows, P = products), question number (1), possible score on that question (8 = 8 possible points), and the number of answers analyzed for that question (221). The labels represent Question Type and Number Possible Score _N.

Students’ scores were higher on questions with explicit atoms (Mdn = 83%) compared to implicit atoms (Mdn = 50%). U = 3,786,072, z = -22.6, p < .001, r = -.27. Explicit atoms refer to atoms whose symbols are always explicitly written, such as N for nitrogen and O.
for oxygen, as well as carbon or hydrogen atoms that have been drawn as a fully or partially expanded structure. Implicit atoms refer to the carbon and hydrogen atoms that are embedded in a line structure. For example, students scored higher on A6 ($Mdn = 100\%$) than on A8 ($Mdn = 63\%$), Figure 9.

**Explicit**

A6 ($Mdn = 100\%$): Atoms such as N and O are explicitly drawn.

![Diagram of explicit atoms](image)

**Implicit**

A8 ($Mdn = 63\%$): The methyl ketone was presented as line structure. Students had to expand the $\alpha$ carbon to draw the mechanism.

![Diagram of implicit atoms](image)

Figure 9. Comparison between questions with explicit (top) and implicit (bottom) atoms involved in the mechanism.

Students’ scores were also higher on questions that involved *intramolecular* reaction steps ($Mdn = 75\%$) compared to *intermolecular* steps ($Mdn = 50\%$), $U = 4,850,614, z = -14.2, p < .001, r = -.17$. For example, a number of questions involving intramolecular reactions had median scores of about 50%: P5, P6, P7, and P14.

Some questions had different scores in **different years**. For example, P8 ($Mdn = 50\%$, asked in 2014) and P9 ($Mdn = 100\%$, asked in 2015) represent the same question, which were given on different final exams in Organic I; exam papers were never returned to students, $U = 38,774.5, z = 2.73, p = .006, r = .12$. A similar trend was seen between P2 (Organic II, 2014, $Mdn = 0\%$) and P1 (Organic I, 2015, $Mdn = 50\%$), $U = 17,695, z = -4.26, p < .001, r = -.19$. The effect sizes were small for this trend. These trends could be due to an increased emphasis in this question type in 2015, with associated class videos (Flynn, 2015b), practice, and having a systematic way to solve them.

P3 and P4 were identical questions, except that the reactant was drawn in a **different conformation** in each one (Figure 5). P3’s scores ($Mdn = 100\%$) were significantly higher than P4’s scores ($Mdn = 0\%$), $U = 25,103.0, z = -6.06, p < .001, r = -.26$. The different conformations (on the page) clearly had a marked effect on students’ ability to connect atoms. In another example, P16’s low scores ($Mdn = 0\%$) are likely a reflection of the nature of the question: intramolecular, implicit atoms, and an unusual substructure in the product—a cyclopropanone.
Students’ strategies and errors on Arrows questions

We analyzed students’ answers to find out what kinds of strategies they were using (Figure 10) and what errors they were making (Figure 11); most students (>95%) attempted every question. We analyzed five types of Arrows questions, breaking down our analysis by reaction step, giving a total of 18 reaction steps analyzed (Figure 10).

Figure 10. Scores and strategies demonstrated on answers to Arrows questions. Each reaction step is represented with a different lowercase letter (see legend). LP = lone pairs or non-bonding electrons. The given action (e.g., drew required LPs) was required for steps with a dot (·).

Figure 11. Scores and errors demonstrated on answers to Arrows questions. Each reaction step is represented with a different lowercase letter (see legend).
Most students expanded the structures when required (Figure 10), with exceptions in steps A8a and A14d. In both questions, few students expanded the reactive carbon to reveal the reactive proton and bond; implicit atoms were involved in both cases. Mapping and expanding are two strategies that could have been helpful in these cases, which seemed more difficult than other questions.

Some low scores may reflect symbolism use issues or conceptual understanding issues. For example, scores on all steps of A4, A8, A11, and A14 might have been higher if the non-bonding electrons (lone pairs) had been drawn in (Figure 10); students may not have understood the bonding changes in the reaction or may simply have neglected to draw in the electrons. Students’ conceptual understanding of bonding could also have impacted their analysis, although we did not analyze their conceptual understanding in this study.

Very few arrows were drawn in reverse, from atom to electrons (Figure 11). In line with our requirements for these questions, most arrows started from electrons (non-bonding or bonding) and pointed to an atom or bond. In only a few cases, the arrows started from an atom (instead of drawing in the non-bonding electrons) or from the negative charge; these ways of drawing arrows are commonly accepted in the organic chemistry community and so other instructors might have accepted these arrows as being correct (which would correspondingly lead to higher scores in these questions). There were only a few cases of extra, missing, or illogical arrows (e.g., that did not have meaning in the context of the reaction), suggesting that students are tracking electron movement (in most cases) and understand how to account for that movement within each reaction step.

Not expanding or mapping atoms and electrons seems to be a major source of errors, although students might have been doing this internally or by drawing out electrons, atoms, or bonds instead of numbering or colouring the reaction components. For example, we analyzed steps A14a and A14d to identify the answers students were giving. In A14a, only about half of students provided the correct arrows (Figure 12). Another 35% showed one of the correct arrows (answers 2–4), which depicted how the negative charge on the sulfur atom was generated. Still, half of students could not correctly identify the electron source of the new N=N π bond, with 35% of students drawing an arrow from the negative charge. For step A4d (Figure 13), only 18% of students provided the accepted arrows, either the anticipated ones (answer 1) or an accepted alternative (answer 2). 76% of students drew the π bond forming using the carbene electrons, which had been explicitly drawn for the students, but did not show how the hydride migrated to the carbene atom (answer 3). 68% of students never drew in a hydrogen atom β to the carbene (not shown as a separate answer), presumably because they did not realize it was involved in the reaction.
Strategies and errors on Products questions

Products questions had more errors. Most students attempted every question (Figure 14). A variable number of students drew non-bonding electrons (i.e., lone pairs). Few students expanded bonds or atoms, redrew the structure or showed evidence of mapping. Students had been taught to use these strategies, initially as part of in-class instruction (2014), then
as a series of videos and in-class activities (2015) (Flynn, 2015b). Further, we have seen that these strategies are associated with successful solutions to synthesis problems (Bodé & Flynn, 2016; Flynn, 2015b).

![Graph showing strategy demonstration strategy demonstrated on Products questions.](image-url)

**Figure 14. Strategies that students demonstrated on Products questions.**

The majority of students (>90%) correctly drew in non-bonding electrons, expanded bonds and atoms, and mapped, when they chose to do so. When these questions were first given to students, only the non-bonding electrons explicitly involved in the reaction mechanism were shown.

For Arrows questions in the first phase of this project (2012–2014), only non-bonding electrons directly involved in a mechanism were drawn out. Based on students’ errors, showing all the electrons does not seem to be necessary or create confusion except in more complicated examples. In those cases, not showing students the electrons seemed to lead to confusion. In one example that involved a rearrangement, a large proportion of students drew non-bonding electrons in P2 (56%) and the majority did so incorrectly (58%), unlike in easier questions. Most of these errors related to non-bonding electrons drawn on the nitrogen atom of the acetylated hydroxamic acid. Now during instruction, all the non-bonding electrons are drawn on atoms directly involved in a reaction step. Students will also be required to do so on assessments.

We found differences between the answers to P3 and P4 in addition to the higher scores in P3. All the students who expanded structures in P3 did so correctly, compared to only 90% in P4. In P3, 88% of students drew in non-bonding electrons correctly, compared to only 67% of students in P4.

Many students (16%, N = 120) gave an unanticipated answer to the third step of question A15 (Figure 15). Another 28% gave the anticipated response, 24% gave a partial response (i.e., were missing arrows), and 32% made a fundamental error in their response (e.g., reversed arrow, etc.). There are a couple of different ways to interpret students’
unanticipated answers. Although they had learned much about carbonyl chemistry by this point in the course (e.g., acylation, saponification), and could have used their chemistry knowledge when answering this question, perhaps they were simply conducting a bookkeeping exercise. That is, perhaps they were simply adding arrows that get them to the product (Bhattacharyya & Bodner, 2005). Alternatively, perhaps they believed their answer was chemically plausible, in the same way that a chemist would propose and explore many possible mechanisms and discount certain ones based on experimental evidence. Students’ alternative answers provide excellent opportunities for further class discussion and analysis.

Anticipated answer (28% of responses)

![Anticipated answer](image)

Common answer (16% of responses, accepted also)

![Common answer](image)

Figure 15. Two common answers to A15c (step 3). Students did not have to label the specific oxygen atoms. N = 120.

When we analyzed students’ errors in the Products questions, we found many formal charge errors throughout the majority of questions (Figure 16). We contrasted these many errors with situations in which students choose to draw out the non-bonding electrons or expand structures; in the latter situations, the majority did so correctly (>90%). The only questions that did not have a high incidence of formal charge errors in the answer were P3 and P11. In P3, the structure was drawn in the pseudo-reactive conformation and the median score was 100%. We believe that the process of explicitly drawing electrons, bond, and atoms would help students more clearly interpret a reaction process and avoid formal charge errors.
Figure 16. Errors found in Products questions.

There was also a type of error that we had not observed before, which we describe as the following arrow interpretation: picked up electrons and moved them to a completely different atom. For example, two different answers were accepted for P1 and P2 that show how students interpret electron-pushing arrows differently (Figure 17). In many answers, the nitrogen–carbon electrons were “placed” on the nitrogen atom instead of migrating the bond and aryl group to the nitrogen atom. We had never seen this type of error on past exams, but we had also not asked this type of question where the arrows were provided. We do not know if instruction has influenced the way students interpret this question type (i.e., by asking students to interpret reactions they have not seen before) or if students instinctively interpret these arrows in this way. The latter has important considerations for instruction, as students seem to be seeing alternative interpretations of the electron-pushing arrows that we present in classes and would need opportunities to learn the chemistry (plausibility) behind these alternate reaction pathways.
We found many error types on answers to P10. This question was given in the first midterm of Organic Chemistry I, only a week after students had learned the electron-pushing formalism (through videos, in-class activities, and a problem set). These errors suggest how students initially interpret the electron-pushing formalism and support our opinion that students should explicitly be taught the EPF, so that the formalism itself does not become a barrier to learning the chemistry that follows. Time, practice, and feedback are needed to help support students’ learning.

In 2015, the instructor explained to students that the electrons from a reacting bond stay with one of the atoms from that bond (Flynn, 2015b). Even so, more than half of students in the final exam of the 2015 Organic II course interpreted an electron-pushing arrow by picking up and moving electrons (Figure 18, answers 2–4). The high proportion of students who gave this answer might have interpreted the cyclopropanone substructure as being implausible, which would suggest that they are considering the nature of the product formed and not simply conducting a bookkeeping exercise for the sake of answering a question. In other words, they might be linking the electron-pushing formalism to chemistry meaning, a possibility that we are investigating in a related project. Some students wrote notes beside their answers, to support this idea, such as one student who drew the correct answer and wrote: “This is what it [the question] seems to be indicating.
but tried it with a molecular model and didn’t work too well.” Note: although there are alternative reaction mechanisms for the Favorski reaction, this was the version presented to students (as starting material and electron-pushing arrows). More concerning than an alternate interpretation of the arrow formalism, a full 70% of students made an implicit atom error in this question (answers 3–6).

![Figure 18](image_url)

Figure 18. Distribution of common responses to P16 revealed many implicit atom errors, formal charge errors, and “pick up electrons” interpretations. N = 104.

Very few students drew pentavalent carbon atoms (<3%), except for P10 and for P16. Question P10 was given in the first midterm of Organic I, when students had just learned the electron-pushing formalism, which could explain that error. For question P16, students may have realized that answer 5 (Figure 18) was not a plausible answer but could not think of a better alternative, or they might not have realized their error.

Our investigation of students’ strategies and errors suggested further barriers, in that students demonstrated few of the strategies that we had taught them and made a number of errors when interpreting the formalism to draw products. Why were there so few demonstrated strategies even though students were taught to use them? Perhaps these strategies are not being consistently modeled in class. That is, perhaps students were being told to use strategies but the instructor was not doing so consistently, and so the students did not get sufficient practice to see the importance of the strategies. Perhaps the students did not see the purpose of the strategies because they did not recognize they were making mistakes. If the strategies are beneficial, then students who begin to use them should have higher success on these questions and presumably on other mechanism type questions.

Some errors are particularly surprising, like formal charge errors, especially when the students who expand structures do not typically make these errors. There are a few
possibilities for this: these errors could be caused because the novice students perceive the structures differently than do more experienced students and experts. That is, they may not “see” the implicit atoms. This difficulty with implicit atoms could be leading to a string of subsequent errors. If this were true, then students’ scores should increase if instructors consistently modeled the strategies of expanding and mapping structures in the course (i.e., going beyond recommending it and modeling this strategy at the beginning then for only hard questions) or if they required students to expand structures (at least around the reacting centres).

Limitations

We do not yet know if students are gaining greater fluency in the electron-pushing formalism by teaching the formalism before reactions, giving students opportunities to practice, and assessing their skills on the intended learning outcomes on all midterms and exams. Future studies will continue to investigate this question. We do not know from this study how students conceptualize or construct their understanding of the reactions or the questions of this type and are following up on their meaning-making in another study.

Conclusions

In this study, students’ scores, strategies, and errors were investigated on two organic chemistry language/EPF question types: (1) drawing the electron-pushing arrows, given the starting materials and products of a reaction step (called “Arrows”), and (2) drawing the products of a reaction step, given the starting materials and electron-pushing arrows of that step (called “Products”). We explicitly teach students to interpret and use organic chemistry’s language as a step or scaffold as they develop greater competency in chemistry. Any reaction can be used to create the types of questions drawn herein; we choose reactions that students have not yet learned, and we show each step of the reaction explicitly. The goal for these question types is for students to interpret and use the electron-pushing formalism, not predict or interpret the chemistry principles involved in the reaction.

The vast majority of students attempted every question and very few reversed or illogical errors were found. Students’ scores were higher on the Arrows questions than the Products questions. Questions with implicit atoms or intramolecular steps were significantly correlated with lower scores, which signify barriers to later chemistry learning. Beyond identifying errors with the conventions of the EPF, our study found more fundamental issues of understanding, such as students interpreting some electron-pushing arrows as saying that electrons should be relocated to a different atom (especially in rearrangements).

Few students demonstrated using the strategies they had been taught, including: expanding structures by drawing out non-bonding electrons, bonds, or implicit atoms (carbon or hydrogen), and mapping the atoms and electrons. These strategies could be beneficial, especially in more challenging questions (e.g., with many implicit atoms, intramolecular reactions, unfamiliar structures or atoms).
There were some questions in which students provided unexpected answers that could enrich classroom discussions. P1 and P2 are a prime example of this (Figure 17). Alternative mechanism interpretations can be linked to a research setting in which a reaction’s mechanism is being investigated. In class, we can discuss how to probe various alternatives and what to look for in the data. While we would not expect students early in an Organic I course to be able to propose mechanisms, we would expect them to know that alternatives are possible; by the end of an Organic II course, students could be expected to propose or compare mechanistic pathways for a reaction.

Research has shown the many difficulties students have with using this formalism and with reaction mechanisms. These difficulties include connecting Brønsted-Lowry theory, Lewis theory, nucleophile/electrophile concepts and reactivity, and structure/property relationships (Anzovino & Bretz, 2015; Cartrette & Mayo, 2011; Cruz & Towns, 2014; DeFever et al., 2015; Strickland et al., 2010), connecting structure and property relationships. Students often seem to rely on intuitive judgments when making decisions (Graulich, 2015b; Weinrich & Talanquer, 2015, 2016), focus on product stability rather than feasibility of the reaction mechanism (Rushton et al., 2008), and struggle with multivariate problems (Kraft, Strickland, & Bhattacharyya, 2010).

Experts largely agree that the principal use for mechanistic reasoning using the EPF is to explain and predict outcomes of chemical processes (Bhattacharyya, 2013) and research has shown a benefit to using mechanistic thinking, despite students’ difficulties (Grove, Cooper, & Cox, 2012). However, students often propose mechanisms that are not meaningful—ones that “get them to the product” (Bhattacharyya & Bodner, 2005), they “decorate with arrows” after providing a product (Grove, Cooper, & Rush, 2012), or interpret representations based on surface features (Kozma & Russell, 1997). With so many aspects of language and tools unique to organic chemistry (Bhattacharyya & Bodner, 2014), students understanding and interpretation of this language and tools will certainly impact their ability to learn chemistry concepts and make connections.

Our study revealed that many of the difficulties arise before students even have to consider the chemical meaning and implications of the reactions—when “all” they have to do is interpret the symbolism. Some of our findings are consistent with previous research; for example, they demonstrate that the symbolism to them is simply a collection of “letters and lines and numbers that cannot correctly be called symbols because they do not represent or symbolize anything that has physical reality” (Bodner & Domin, 2000). This seems to especially be the case when implicit atoms are involved in a reaction. However, we found few reversed arrows or illogical errors, suggesting that students are attributing the meaning of electron-movement and bond formation/breakage to the curved arrows. In other words, the EPF seems meaningful to students in many contexts.

As our next research steps, we are investigating how students connect the organic chemistry symbolism with their prior learning, and how they describe their interpretation of EPF as it related to the sub-microscopic level (i.e., how they understand and translate between these levels).

If students are to understand the sub-microscopic level, they need an understanding and fluency in the symbolic level and an ability to connect the sub-micro and macroscopic
levels (Gilbert & Treagust, 2009; Johnstone, 2000). For instruction, we recommend modeling such strategies, embedding these strategies and question types in formative and summative assessment opportunities, and giving rapid or immediate feedback. To help students practice the skills involved, we developed a free, bilingual, online learning module that involves explanatory videos (Flynn, 2015b; Flynn et al., 2016), questions, feedback, and a metacognitive skill building layer. Instruction (explanations, modeling, activities, and assessment) could also help students connect the symbolic level with the macro- and sub-microscopic level (Cheng & Gilbert, 2009), along with explorations of the strengths and limitations of the representations we use.

**Acknowledgements**

We thank the University of Ottawa for funding this work and the following research assistants for their contributions: Declan Webber, Delphine Amellal, Jade Choo Foo, John Giorgi, and Sarang Gupta.
# Appendix 1

## Appendix 1. Coding template for Arrows and Products questions

<table>
<thead>
<tr>
<th>Arrows questions (analysis repeated for each step)</th>
<th>Products questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>File #, Step #</td>
<td>File #</td>
</tr>
<tr>
<td>Assessment score</td>
<td>Assessment score</td>
</tr>
<tr>
<td>Step a/b/c/d/e (possible score)</td>
<td>Score (possible score)</td>
</tr>
<tr>
<td>No attempt</td>
<td>No attempt</td>
</tr>
<tr>
<td>Required LPs drawn/not drawn (R/W)</td>
<td></td>
</tr>
<tr>
<td>Some LPs drawn (R/W)</td>
<td>Some LPs drawn (R/W)</td>
</tr>
<tr>
<td>All LPs drawn/not drawn (R/W)</td>
<td>All LPs drawn/not drawn (R/W)</td>
</tr>
<tr>
<td>Expanded bonds/atoms -- required ones (R/W)</td>
<td></td>
</tr>
<tr>
<td>Expanded bonds/atoms -- some more than required (R/W)</td>
<td>Expanded bonds/atoms -- some (R/W)</td>
</tr>
<tr>
<td>Expanded bonds/atoms -- all (R/W)</td>
<td>Expanded bonds/atoms -- all (R/W)</td>
</tr>
<tr>
<td>Redrew starting material(s) - part (R/W)</td>
<td>Redrew structure - part (R/W)</td>
</tr>
<tr>
<td>Redrew starting material(s) - All (R/W)</td>
<td>Redrew structure - All (R/W)</td>
</tr>
<tr>
<td>Counted Cs in SM</td>
<td></td>
</tr>
<tr>
<td>Mapping evidence (R/W)</td>
<td>Mapping evidence (R/W)</td>
</tr>
<tr>
<td>Formal charge error</td>
<td>Formal charge error</td>
</tr>
<tr>
<td>Wrong reacting partners</td>
<td>Picked electrons up and moved them to a completely different atom/bond</td>
</tr>
<tr>
<td>Wrong LP used</td>
<td>Added extra step/intermediate</td>
</tr>
<tr>
<td>Arrow reversal (atom to e-s)</td>
<td></td>
</tr>
<tr>
<td>Arrows: Extra (E) or Missing (M)</td>
<td>Ignored/missed an arrow (didn't do what the arrows says to do) - which one?</td>
</tr>
<tr>
<td>Arrows: didn't start from e-s</td>
<td>Wrong configuration (R/S, E/Z)</td>
</tr>
<tr>
<td>Implicit atom error (H or C as part of a line structure) (E/M)</td>
<td>Stereochemical discrepancy (e.g., 3D on sp² C in plane)</td>
</tr>
<tr>
<td>Comments</td>
<td></td>
</tr>
<tr>
<td>Carbons: Extra/Missing (E/M)</td>
<td></td>
</tr>
<tr>
<td>Functional group-Extra/Missing (E/M)</td>
<td></td>
</tr>
<tr>
<td>Bond: Extra/Missing (E/M)</td>
<td></td>
</tr>
<tr>
<td>Implicit atom error (H or C as part of a line structure) (E/M)</td>
<td></td>
</tr>
<tr>
<td>Comments</td>
<td></td>
</tr>
</tbody>
</table>
References


Weinrich, M., & Talanquer, V. (2015). Mapping students’ conceptual modes when thinking about chemical reactions used to make a desired product. *Education in Chemistry*. JOUR. Retrieved from [http://pubs.rsc.org/en/content/articlehtml/2015/rp/c5rp00024f](http://pubs.rsc.org/en/content/articlehtml/2015/rp/c5rp00024f)
